Quantum Bayesian networks with application to games displaying Parrondo's paradox

by

Michael Pejic

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Mathematics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor F. Alberto Grünbaum, Chair
Professor Nicolai Reshetikhin
Umesh Vazirani

Fall 2014

Quantum Bayesian networks with application to games displaying Parrondo's paradox

Abstract

Quantum Bayesian networks with application to games displaying Parrondo's paradox

by

Michael Pejic

Doctor of Philosophy in Mathematics

University of California, Berkeley

Professor F. Alberto Grünbaum, Chair

Bayesian networks and their accompanying graphical models are widely used for prediction and analysis across many disciplines. We will reformulate these in terms of linear maps. This reformulation will suggest a natural extension, which we will show is equivalent to standard textbook quantum mechanics. Therefore, this extension will be termed *quantum*. However, the term *quantum* should not be taken to imply this extension is necessarily only of utility in situations traditionally thought of as in the domain of quantum mechanics. In principle, it may be employed in any modelling situation, say forecasting the weather or the stock market– it is up to experiment to determine if this extension is useful in practice. Even restricting to the domain of quantum mechanics, with this new formulation the advantages of Bayesian networks can be maintained for models incorporating quantum and mixed classical-quantum behavior. The use of these will be illustrated by various basic examples.

Parrondo's paradox refers to the situation where two, multi-round games with a fixed winning criteria, both with probability greater than one-half for one player to win, are combined. Using a possibly biased coin to determine the rule to employ for each round, paradoxically, the previously losing player now wins the combined game with probability greater than one-half. Using the extended Bayesian networks, we will formulate and analyze classical observed, classical hidden, and quantum versions of a game that displays this paradox, finding bounds for the discrepancy from naive expectations for the occurrence of the paradox. A quantum paradox inspired by Parrondo's paradox will also be analyzed. We will prove a bound for the discrepancy from naive expectations for this paradox as well. Games involving quantum walks that achieve this bound will be presented.

To my parents.

Contents

Contents **ii**

List of Figures **iv**

1 Introduction **1**

I Quantum Bayesian networks **4**

2 Bayesian networks-graphical models **5**
2.1 Graphical models as the form in which information is presented 5
2.2 Transition probability functions 8
2.3 Graphical models as constraints 10
2.4 Directed graphical models as tensor networks 11
2.5 The Copy map and restriction maps 13

3 Hidden classic and quantum nodes **14**
3.1 Principles of linearity and potential universality 14
3.2 Options I and II 16
3.3 Quantum nodes 20
3.4 No quantum copying 22
3.5 Embedding quantum models into classical ones 26
3.6 Embedding classical models into quantum ones 32
3.7 Classical physics 33
3.8 Additional structures for Bayesian networks? 34

4 Constructing the networks **36**
4.1 Graphical constructs and rules 36
4.2 Data for the structures using Option I' 37
4.3 Data for the structures using Option II 38
4.4 Example-double-slit experiment 39
4.5 Quantum model 40
4.6 The Bayesian network approach versus the standard textbook approach 42

4.7 Classical hidden model . . . 43
4.8 What is a quantum system? . . . 45

5 Relation to textbook quantum mechanics 46
5.1 Textbook rules for quantum mechanics . . . 46
5.2 Weak measurements . . . 52
5.3 Comparison of Bayesian networks to quantum circuits and tensor networks . . . 54

6 Further examples 56
6.1 No-cloning–classical and quantum . . . 56
6.2 Teleportation-classical and quantum . . . 57
6.3 Bell's inequality for Bayesian networks without metaphysical limitations . . . 60

II Parrondo's paradox and a Parrondo-like paradox 64

7 Parrondo's paradox 65
7.1 Defining the paradox . . . 65
7.2 Classical observed Markov chain game . . . 66
7.3 Classical hidden-Markov chain game . . . 70
7.4 Defining a quantum analogue of the Markov model . . . 75
7.5 Parrondo's paradox for the quantum game . . . 80
7.6 Summary of results . . . 83

8 A Parrondo-like paradox for an one-round game 86
8.1 Defining the game . . . 86
8.2 Defining geodesics on the space of wavefunctions . . . 87
8.3 Bounds on the extent of the Parrondo-like paradox . . . 89
8.4 Conditions for the occurrence of the Parrondo-like paradox . . . 92

9 Quantum walks and the Parrondo-like paradox 95
9.1 Classical random and classical hidden walks . . . 95
9.2 Quantum walks . . . 99
9.3 The Parrondo-like paradox for quantum walks . . . 104

Bibliography 109

A General propositions 114
A.1 Banach space propositions . . . 114
A.2 Hilbert space propositions . . . 115
A.3 Transition function propositions . . . 116

B Propositions for option I 119

B.1 Measures 119
B.2 Maps on subspaces of the space of measures 121
B.3 L^1-spaces 128
B.4 Density-matrix-valued L^1-spaces 128
B.5 Maps on density-matrix-valued L^1-spaces 129
B.6 Vector measures 139

C Propositions for option II **141**
C.1 Real-valued, continuous functions 141
C.2 Maps on real-valued, continuous functions 142
C.3 Compact-operator-valued, continuous functions 146
C.4 Operator inequalities 146
C.5 Maps on compact-operator-valued, continuous functions 150

List of Figures

7.1 Minimum achievable values of P_A^{comb} for $p = 0.5$ for the two-state, classical hidden game. 84
7.2 Minimum achievable values of P_A^{comb} for $p = 0.1$ for the two-state, classical hidden game. 84
7.3 Minimum achievable values of P_A^{comb} for $p = 0.5$ for the two-dimensional, quantum game. 85
7.4 Minimum achievable values of P_A^{comb} for $p = 0.1$ for the two-dimensional, quantum game. 85

Acknowledgments

Michael Pejic acknowledges support from the Applied Math. Sciences subprogram of the Office of Energy Research, US Department of Energy, under Contract DE-AC03-76SF00098, and from AFOSR grant FA95501210087 through a subcontract to Carnegie Mellon University.

Chapter 1

Introduction

Outline of the work

Bayesian networks and graphical models are useful for classical systems because they are much more intuitive than a list of conditional dependencies. It is also sometimes useful to introduce additional, hypothesized nodes to break complicated dependencies into simpler, potentially universal modules. Usually these are treated just as observable nodes which are always hidden; however, this imposes constraints that are metaphysical in origin and raises difficulties of interpretation. We will give an alternate approach using linear maps on measures, with additional constructions to those generally utilized in graphical models, that resolves those issues. While, in certain situations, this introduces additional maps, not previously available, these new maps are not only in and of themselves of limited interest, but also introduce undesired complications. However, what is extremely fruitful is simply the conceptual leap. Thinking in terms of linear maps on spaces of measures immediately raises the question of looking at linear maps on other spaces. This approach leads to a natural extension, which we will prove is equivalent to standard, textbook quantum mechanics (which is infamous for its apparently unmotivated and incomprehensible formulation). Therefore, this extension will be termed *quantum*. To avoid being swept away by a flood of details, propositions of a more general nature, together with their proofs, needed to show the sensibility and consistency of the extension and its equivalence to quantum mechanics are placed in appendices.

However, the term *quantum* should not be taken to imply this extension is necessarily only of utility in situations traditionally thought of as in the domain of quantum mechanics. In principle, it may be employed in any modelling situation, say forecasting weather or stock prices–it is up to experiment to determine if this extension is useful in practice. In particular, there is no reason for $\hbar$ to necessarily enter into these models if they are outside the realm of physics. Even restricting to the traditional domain of quantum mechanics, with this new formulation the advantages of Bayesian networks can be maintained for models incorporating quantum and mixed classical-quantum behavior. The use of these will be illustrated by various examples. In particular, we will show that some of the supposed hallmarks of quantum mechanics, no-cloning and teleportation, apply for classical hidden systems as well.

In the second part, we will utilize these extended Bayesian networks in the study of various games displaying Parrondo's paradox–the phenomenon of two games each winning for one player with probability greater than one-half, yet their convex combination (in a sense to be specified) paradoxically winning for the previously losing player with probability greater than one-half. We will prove bounds for the discrepancy from naive expectations for classical versions of a game; those bounds will then be shown to be broken by a quantum analogue of the game. [26]

A quantum paradox inspired by Parrondo's paradox will also be analyzed. We will prove a bound for the discrepancy from naive expectations for this paradox. Games involving quantum walks that achieve this bound will be presented.

Philosophical interlude–rejection of metaphysics

> For a man will attain unto nothing more perfect than to be found to be most learned in the ignorance which is distinctly his. The more he knows that he is unknowing, the more learned he will be.–Nicholas of Cusa [27]

The philosophy we employ in this work is one with a long-standing pedigree: we know nothing about underlying reality and, therefore, any claims about it or appeals to it are invalid. There is no reason to believe reality is doing calculations at all similar to the ones we employ in our hypothesized models or even that it is doing calculations at all. In particular there is the issue of contextuality–we want to employ potentially universal modules in our models since only they have predictive ability in novel situations, but perhaps reality is fundamentally contextual.

Constraints imposed on our models

In keeping with the expressed philosophy, no metaphysical constraints will be placed upon the mathematical operations and constructs that can be employed in the models. Rather, there are only three rules that will be enforced. Firstly, the quantities calculated by the mathematical models must be interpretable as probabilities; in particular, they must be positive[1]. Secondly, the mathematical models must be composed of linear maps (which we will show is the weakening of a principle already in wide use, if not always acknowledged). Thirdly, the mathematical models must be composed of potentially universal modules (in a manner that we will precisely define). Of course, the particular model employed in a particular situation may fail to be universal when actually employed in a different context; the point is that this failure should be as a result of experiment and not be preordained as result of our choice of mathematics employed in modelling.

[1]To aid readability, *positive* is used instead of *nonnegative* throughout. Wherever strict positivity is required, the word *strict* will be added.

These constraints are extremely restrictive. For maps to be linear, they must clearly live on linear spaces. Furthermore, they also impose strong restrictions on the linear spaces these maps live on. Thus far, we are only aware of three classes of linear spaces that meet the imposed restrictions: *(i)* certain subspaces of measures; *(ii)* density matrices on complex Hilbert spaces ; *(ii)* and their tensor products. The former gives what is traditionally thought of as classical behavior and we will prove the latter two give behavior that has traditionally been taken the domain of quantum mechanics. However, since we do not start from quantum mechanics, but instead only from the above principles, models involving maps on density matrices may be found to be of utility in situations not traditionally thought of as related to quantum mechanics.

Part I

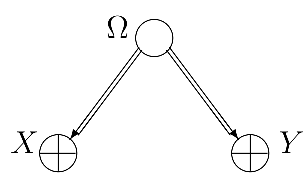
Quantum Bayesian networks

Chapter 2

Bayesian networks-graphical models

2.1 Graphical models as the form in which information is presented

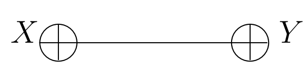
There are many ways to present joint probability in terms of other quantities and mathematical constructs. Graphical models [28] are a useful way to sort through what otherwise can seem hopelessly complicated in the usual notation. For instance, given a probability space $(\Omega, \mathcal{E}, \pi)$ with two generalized random variables, $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$ for sets $\mathcal{X}$ and $\mathcal{Y}$, then for any $A \in X(\mathcal{E})$ and $B \in Y(\mathcal{E})$, the joint probability $\text{Prob}(X \in A \text{ and } Y \in B)$ is $\pi\left(X^{-1}(A) \cap Y^{-1}(B)\right)$. The graphical model corresponding to presenting the joint probability in this manner–namely by giving $(\Omega,\ \mathcal{E},\ \pi,\ \mathcal{X},\ X,\ \mathcal{Y},\ Y)$–is



(2.1)

where the double arrows stand for deterministic causation.

The resulting joint probability determines a probability space $(\mathcal{X} \times \mathcal{Y}, \mathcal{F}, \rho)$ with the probability measure ρ given on rectangular subsets by $\rho(A \times B) = \text{Prob}\left(X^{-1}(A) \cup Y^{-1}(B)\right)$, which can then be extended to a probability measure for the σ-algebra[1] $\mathcal{F}$ generated by the rectangular subsets on $\mathcal{X} \times \mathcal{Y}$. The graphical model corresponding to presenting the joint probability in this manner–namely by giving $(\mathcal{X} \times \mathcal{Y},\ \mathcal{F},\ \rho)$–is



(2.2)

[1] A *σ-algebra* is a collection of subsets of a set $\mathcal{X}$, including both $\varnothing$ and $\mathcal{X}$, that is closed under relative complementation and countable unions.

For each $B \in Y(\mathcal{E})$, the probability for X determines a probability space $(\mathcal{X}, \mathcal{G}, \mu_B)$ with σ-algebra $\mathcal{G} = X(\mathcal{E})$ and with probability measure μ_B given by $\mu_B(A) = \rho(A \times B) = \pi(X^{-1}(A) \cup Y^{-1}(B)) = \text{Prob}\,(X \in A \text{ and } Y \in B)$. The marginal probability for X is then given by $\mu = \mu_{\mathcal{Y}}$. Let $\mathcal{G} \times \mathcal{Y}$ signify the σ-algebra of all rectangular sets of the form $A \times \mathcal{Y}$ for $A \in \mathcal{G}$. Since μ is a finite measure (hence, σ-finite[2]) and μ_B is absolutely continuous[3] with respect to μ, by the Radon-Nikodým theorem [53], the measures $\{\mu_B\}$ determine the conditional probability (shown in two common notations) $\tau(B|\mathcal{G} \times \mathcal{Y})(x)$ or $\tau(B|x)$ as the Radon-Nikodým derivative $\frac{d\mu_B}{d\mu}(x)$. The function $\frac{d\mu_B}{d\mu}$ is in $L^1(\mathcal{X};\mu)$, which is clearly isometrically isomorphic to the subspace of $\mathcal{G} \times \mathcal{Y}$-measurable functions within $L^1(\mathcal{X} \times \mathcal{Y}; \rho|_{\mathcal{G} \times \mathcal{Y}})$.

Let $\mathcal{H}$ be the σ-algebra $Y((E)$. For any disjoint, countable collection $\{B_j\} \subset \mathcal{H}$,

$$\mu_{\bigcup_j B_j} = \rho(\cdot \times \bigcup_j B_j) = \sum_j \rho(\cdot \times B_j) = \sum_j \mu_{B_j} \tag{2.3}$$

with convergence in norm. Since $\left\| \frac{d\mu_B}{d\mu}(x) \right\| = \|\mu_B\|$,

$$\tau\left(\bigcup_j B_j \middle| \cdot\right) = \sum_j \tau\left(B_j|\cdot\right) \tag{2.4}$$

with convergence in norm. Hence, the conditional probability $\tau(\cdot|\cdot)$ is a $L^1(\mathcal{X};\mu)$-valued vector measure[4] on $\mathcal{Y}$. The other interpretation which may at times exist, as a function on $\mathcal{X}$ with values in the measures on $\mathcal{Y}$, is less useful because the function is not generally Bochner integrable[5] unless the σ-algebra $\mathcal{H}$ on $\mathcal{Y}$ is generated by a countable collection of atoms[6]. For example, if $\mathcal{X} = \mathcal{Y}$ and the random variables X and Y are the same, then $\tau(B|x) = 1_B(x) = \delta_x(B)$, where the second interpretation as $\tau(\cdot|x) = \delta_x$ is not Bochner integrable unless the σ-algebra on $\mathcal{X}$ is generated by a countable collection of atoms.

Then the joint probability Prob$(X \in A$ and $Y \in B)$ is given by

$$\int_{x \in A} \tau(B|x)\, d\mu(x) \tag{2.5}$$

[2]A measure μ on set $\mathcal{X}$ is *σ-finite* if there are a countable collection of μ-measurable subsets $\{B_j\}$ such that $\bigcup_j B_j = \mathcal{X}$ with each $\mu(B_j)$ finite.

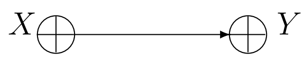
[3]A measure ν on a set $\mathcal{X}$ is *absolutely continuous* with respect to a measure μ on $\mathcal{X}$ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all μ-measurable subsets A.

[4]A *vector measure* is a countably-additive set function with values in a Banach space where the convergence for the countably-additivity is in norm.

[5]For any Banach space B, a B-valued function on a set $\mathcal{X}$ with a measure μ is *Bochner integrable* if there is a sequence of simple functions (functions taking only finitely many values with each value achieved on a set with finite μ-measure) converging to it, both pointwise in B-norm almost everywhere and in $L^1(\mathcal{X};\mu;\mathsf{B})$-norm.

[6]A set in a set algebra is an *atom* if it is indivisible in the set algebra.

The directed graphical model corresponding to presenting the joint probability in this manner–namely by giving $(\mathcal{X}, \mathcal{G}, \mathcal{Y}, \mathcal{H}, \mu, \tau(\cdot|\cdot))$–is



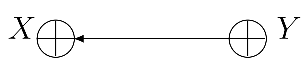
(2.6)

Let ν be the marginal probability for Y, $\nu(B) = \rho(\mathcal{X} \times B)$. Then $\tau(\cdot|\cdot)$ is absolutely continuous with respect to ν in the sense that $\tau(B|\cdot)$ is the zero function for every B such that $\nu(B) = 0$. However, unless we are in the common case where the σ-algebra on $\mathcal{X}$ is generated by a countable collection of atoms, $L^1(\mathcal{X};\mu)$ does not have the Radon-Nikodým property[7] [77] (the example given above demonstrates this); hence, there is in general no $f \in L^1(\mathcal{Y};\nu;L^1(\mathcal{X};\mu))$ (which by Fubini's theorem [54] is the same as $L^1(\mathcal{X}\times\mathcal{Y};\mu\times\nu)$ for the product measure[8] $\mu\times\nu$) such that the joint probability $\text{Prob}(X \in A \text{ and } Y \in B)$ is given by $\int_{(x,y)\in A\times B} f(x,y)\, d(\mu\times\nu)(x,y)$.

Lastly, defining $\zeta(\cdot|\cdot)$ symmetrically to $\tau(\cdot|\cdot)$, the directed graphical model corresponding to presenting the joint probability

$$\text{Prob}(X \in A \text{ and } Y \in B) = \int_{y\in B} \zeta(A|y)\, d\nu(y) \tag{2.7}$$

by giving the marginal probability ν and the conditional probability $\zeta(\cdot|\cdot)$ is

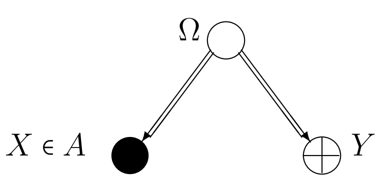


(2.8)

As an example, consider calculating the conditional probability

$$\text{Prob}\,(Y \in B|X \in A) = \frac{\text{Prob}(Y \in B \text{ and } X \in A)}{\text{Prob}(X \in A)} \tag{2.9}$$

(for $\text{Prob}(X \in A) \neq 0$–otherwise the joint probability does not determine the conditional probability) using the information presented in the manner corresponding to each of the four graphical models (the filled circle indicates which node is being conditioned on):

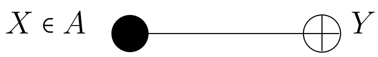


$$\frac{\pi\,(X^{-1}(A)\cap Y^{-1}(B))}{\pi\,(X^{-1}(A))}$$

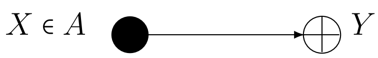
(2.10)

[7]A Banach space B has the *Radon-Nikodým property* if, for any B-valued vector measure ν on a set $\mathcal{X}$ which is absolutely continuous with respect to some σ-finite measure μ on $\mathcal{X}$, there is a Bochner integrable, B-valued function, $\frac{d\nu}{d\mu}$, such that $\nu(A) = \int_A \frac{d\nu}{d\mu}\, d\mu$ for any ν-measurable A.

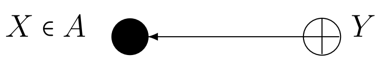
[8]Unfortunately, by convention the tensor product of measures is called the *product* measure and written using × instead of the more appropriate ⊗ (however, see [81] for a use of the latter notation).



$$\frac{\rho\left(A \times B\right)}{\rho\left(A \times \mathcal{Y}\right)} \tag{2.11}$$



$$\frac{\int_{x \in A} \tau(B|x)\, d\mu(x)}{\mu(A)} \tag{2.12}$$



$$\frac{\int_{y \in B} \zeta(A|y)\, d\nu(y)}{\int_{y \in \mathcal{Y}} \zeta(A|y)\, d\nu(y)} \tag{2.13}$$

For instance, Bayes' theorem is simply the calculation corresponding to the presentation of information by the last graphical model.

A similar situation holds for any finite number of random variables [52], with a graphical model with deterministic causation emanating from a fundamental, hidden probability space, a graphical model with a clique of all the nodes for the random variables, and various directed models with probabilistic causation. For the common case where the various σ-algebras are generated by finitely many atoms, the various measures become vectors, the conditional probabilities become stochastic matrices[9] or tensors, and the integrations become sums.

2.2 Transition probability functions

Depending on how the information for the calculation of the joint probability is presented, we may imagine different ways of varying it. For (2.1), it is most natural to imagine independently varying the maps X and Y. For (2.2), there is nothing to independently vary other than the joint probability itself. For (2.6), we would like to imagine varying the marginal probability μ and the conditional probability $\tau(\cdot|\cdot)$ independently. This is a problem because the space $\tau(\cdot|\cdot)$ lives in–the $L^1(\mathcal{X};\mu)$-valued vector measures–depends on μ. This problem will exist even in the commonly occurring case where the σ-algebra $\mathcal{G}$ is generated by a countable collection of atoms if μ is zero on some atom (other than the empty set), since then the conditional probability when conditioning on that atom is not well-defined.

One solution to this problem, following [52], is provided by introducing the following notion:

[9]A matrix is *stochastic* if all entries are either positive or zero and all column sums are one.

Definition 2.2.1 For σ-algebras $\mathcal{G}$ on $\mathcal{X}$ and $\mathcal{H}$ on $\mathcal{Y}$, a function $\tau(\cdot|\cdot) : \mathcal{H} \times \mathcal{X} \to \mathbb{R}$ is a *transition probability function* if: *(i)* for each $x \in \mathcal{X}$, $\tau(\cdot|x)$ is a probability measure on $\mathcal{Y}$ with event σ-algebra $\mathcal{H}$; and *(ii)* for each $B \in \mathcal{H}$, $\tau(B|\cdot)$ is a bounded, $\mathcal{G}$-measurable function on $\mathcal{X}$.

By taking $\tau(\cdot|\cdot)$ to be a transition probability function rather than a conditional probability, it is now possible to vary μ and $\tau(\cdot|\cdot)$ independently. For a fixed choice of $\tau(\cdot|\cdot)$, there is a convex linear map L from probability measures on $\mathcal{X}$ to probability measures on $\mathcal{Y}$ given by

$$(L\mu)(B) = \int_{x \in \mathcal{X}} \tau(B|x)\, d\mu(x) \tag{2.14}$$

The map L extends to a linear map on more general measures and signed measures.

Removing metaphysical constraints

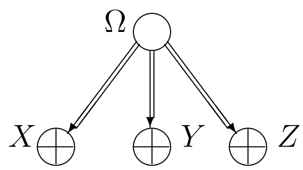
Imagining one can vary μ and keep the transition probability function $\tau(\cdot|\cdot)$ fixed, there is an intuitive interpretation of $\tau(\cdot|\cdot)$ as an idealized conditional probability[10], with $\tau(B|x)$ being the probability to observe B given the event $\{x\}$, even if the latter has probability zero or is not even in the event σ-algebra (although in this latter case it is constant within any atom). This leads to a metaphysical notion of the actual existence of a variable taking an actual value with probability reflecting our ignorance of its value. This may be of value for nodes in a graphical model which are observable; however, it is common to add hypothesized, hidden nodes to a directed graphical model in order to (hopefully) break it up into smaller, more manageable pieces [29]. For these, there is no justification to necessarily limit oneself to transition probability functions. Furthermore, the probabilities and conditional probabilities involving the hidden nodes lack any meaning in either the Bayesian or frequentist interpretations–the word *probability* then only means positive and norm-one.

Another solution (among many others) to the above posed problem is therefore to take the independently varied objects to be the marginal probability μ and the linear map L. Not all linear maps that take probability measures to probability measures are necessarily induced by some transition probability function as in (2.14) (however, they are all induced by some *pseudo-transition function*–see **B2.7**, **B2.8**, and **B2.9**), so this generally introduces additional maps. These additional maps, involving operations such as Lebesgue decomposition [55], are not in and of themselves of great interest; in fact the raise rather undesired complications (see §3.2). Furthermore, in the common case where the σ-algebras are generated by countably many atoms, any linear map is induced by some transition probability function, so there are no additional maps. However, what is fruitful is the conceptual shift; as will be explored in the following chapter, once we are thinking in terms of linear maps, we are immediately drawn to consider the question of looking at linear maps between spaces other than spaces of measures.

[10] As noted in [52].

2.3 Graphical models as constraints

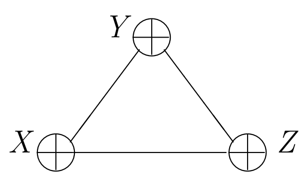
In addition to showing the form in which information is presented, graphical models can also show constraints on the information in a far simpler form than the usual notation. For instance, consider the Markov chain with three random variables X, Y, Z. For the graphical model



(2.15)

with the joint probability $\mathrm{Prob}(X \in A \text{ and } Y \in B \text{ and } Z \in C)$ given by $\pi\left(X^{-1}(A) \cap Y^{-1}(B) \cap Z^{-1}(C)\right)$, it is necessary to explicitly add the constraint that for every $C \in Z(\mathcal{E})$, the conditional probability $\tau(C|x,y) = \frac{d\mu_C}{d\mu}$, for $\mu_C(A \times B) = \pi\left(X^{-1}(A) \cap Y^{-1}(B) \cap Z^{-1}(C)\right)$ and $\mu = \mu_{\mathcal{Z}}$, is independent of x (in the almost-everywhere, probabilistic sense). If, as above, we imagine varying the maps X and Y, it is not at all clear how to do this while maintaining the constraint.

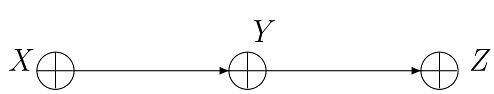
Similarly, for the graphical model



(2.16)

with the joint probability $\mathrm{Prob}(X \in A \text{ and } Y \in B \text{ and } Z \in C)$ given by $\rho(A \times B \times C)$, it is necessary to explicitly add the constraint that for every $C \in \mathcal{I} = Z(\mathcal{E})$, the conditional probability $\tau(C|x,y) = \frac{d\mu_C}{d\mu}$, for $\mu_C(A \times B) = \rho(A \times B \times C)$ and $\mu = \mu_{\mathcal{Z}}$, is independent of x (in the almost-everywhere, probabilistic sense). If, as above, we imagine varying the joint probability ρ, it is not at all clear how to do this while maintaining the constraint.

However, consider the directed graphical model:



(2.17)

which corresponds to presenting the information to calculate the joint probability as ϕ, $\eta(\cdot|\cdot)$, and $\tau(\cdot|\cdot)$ where

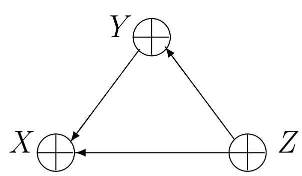
$$\mathrm{Prob}(X \in A, Y \in B, \text{ and } Z \in C) = \int_{(x,y) \in A \times B} \tau(C|y)\, d\mu(x,y) = \int_{y \in B} \tau(C|y)\, d\xi(y) \qquad (2.18)$$

for the marginal probabilities μ on $\mathcal{X} \times \mathcal{Y}$ and ξ on $\mathcal{Y}$ given by

$$\mu(A \times B) = \int_{x \in A} \eta(B|x)\, d\phi(x), \xi(B) = \mu(\mathcal{X} \times B) = \int_{x \in \mathcal{X}} \eta(B|x)\, d\phi(x) \tag{2.19}$$

The restriction is that the conditional probability $\tau(\cdot|\cdot)$ is independent of x (in the almost-everywhere, probabilistic sense). It is not necessary to give this constraint explicitly since it is indicated by the graphical model through the lack of an arrow directly from X to Z. If, as above, we take $\eta(\cdot|\cdot)$ and $\tau(\cdot|\cdot)$ as transition probability functions instead of conditional probabilities, it is now easy to see how to vary ϕ, $\eta(\cdot|\cdot)$, and $\tau(\cdot|\cdot)$ while maintaining the constraint–namely, by only allowing $\tau(\cdot|\cdot)$ that are independent of x.

Note if the wrong directed model is chosen, the constraint can be masked. For instance, for the graphical model



(2.20)

which corresponds to presenting the information to calculate the joint probability as ν, $\theta(\cdot|\cdot)$, and $\zeta(\cdot|\cdot)$ where

$$\text{Prob}(X \in A, Y \in B, \text{ and } Z \in C) = \int_{(y,z) \in B \times C} \zeta(A|y,z)\, d\kappa(y,z) \tag{2.21}$$

for the marginal probability κ on $\mathcal{Y} \times \mathcal{Z}$ given by

$$\kappa(B \times C) = \int_{z \in C \times} \theta(B|z)\, d\nu(z) \tag{2.22}$$

Once again, it is necessary to explicitly add the constraint that for every $C \in \mathcal{I}$, the conditional probability $\tau(C|x,y) = \frac{d\mu_C}{d\mu}$, for $\mu_C(A \times B) = \int_{(y,z) \in B \times C} \zeta(A|y,z)\, d\kappa(y,z)$ and $\mu = \mu_{\mathcal{Z}}$, is independent of x (in the almost-everywhere, probabilistic sense).

This can be readily generalized to more complicated graphical models–any graph that is not simply a clique[11] of all the nodes implies constraints on the allowed joint probabilities. In this manner the various dependencies are displayed in a far more intuitive manner than through a long list of opaque constraints. Of course, it is always possible to impose additional constraints explicitly.

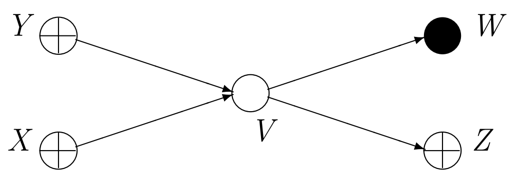
2.4 Directed graphical models as tensor networks

Tensor networks are a commonly employed, diagrammatic device for contracting tensors and vectors. For the common case where the various σ-algebras are generated by finitely many

[11] A *clique* is a group of nodes that are all connected to one another.

atoms, a directed graphical model together with all its information corresponds to a tensor network or, if conditioning is present, the ratio of tensor networks. If the conditioning is only on nodes without parents, the denominator is necessarily one, so these can also be considered tensor networks. Each node in the graphical model with either no children or only one child becomes one node in the tensor network. For nodes in the graphical model with multiple children, it is best to replace them with two nodes, one of which takes in all the inputs and has a single connection to the other, which is a *copying* or *diagonal* node that is zero unless all its connections are the same, when it has the value one, which branches out to all the outputs.

For example, consider the graphical model:

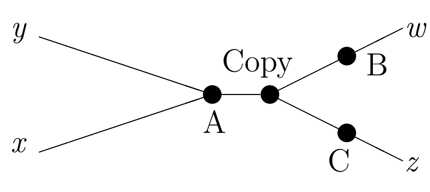


(2.23)

with corresponding information $\text{Prob}(X = x)$, $\text{Prob}(Y = y)$, $\text{Prob}(V = v|X = x, Y = y)$, $\text{Prob}(Z = z|V = v)$, and $\text{Prob}(W = w|V = v)$, so the conditional probability $\text{Prob}(X = x, Y = y, Z = z|W = w)$ is

$$\frac{\sum_{v\in\mathcal{V}} \left(\text{Prob}(X = x)\,\text{Prob}(Y = y)\,\text{Prob}(V = v|X = x, Y = y)\;\text{Prob}(Z = z|V = v)\,\text{Prob}(W = w|V = v)\right)}{\sum_{x'\in\mathcal{X}, y'\in\mathcal{Y}, v'\in\mathcal{V}} \left(\text{Prob}(X = x')\,\text{Prob}(Y = y')\,\text{Prob}(V = v'|X = x', Y = y')\;\text{Prob}(W = w|V = v')\right)} \tag{2.24}$$

The corresponding tensor network for the numerator is:

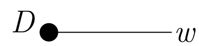


(2.25)

where $A_{xyv} = \text{Prob}(X = x)\,\text{Prob}(Y = y)\,\text{Prob}(V = v|X = x, Y = y)$, which is equal to $\text{Prob}(V = v, X = x, Y = y)$, the Copy tensor is zero unless all its subscript are equal, in which case it has value one, $B_{vw} = \text{Prob}(W = w|V = v)$ and $C_{vz} = \text{Prob}(Z = z|V = v)$. The value for the tensor network is then

$$\sum_{v,v',v''\in\mathcal{V}} A_{xyv}\text{Copy}_{vv'v''}B_{v'w}C_{v''z} = \sum_{v\in\mathcal{V}} A_{xyv}B_{vw}C_{vz} \tag{2.26}$$

which equals the numerator. For the denominator, the tensor network in this case is simply



(2.27)

where

$$D_w = \sum_{x\in\mathcal{X},y\in\mathcal{Y},v\in\mathcal{V}} \text{Prob}(X = x)\,\text{Prob}(Y = y)\,\text{Prob}(V = v|X = x, Y = y)\text{Prob}(W = w|V = v) \tag{2.28}$$

which is equal to $\text{Prob}(W = w)$.

As this example illustrates, the advantages of the Bayesian network over the tensor network are that: *(i)* it is possible to show which nodes are being observed, marginalized, or conditioned on; and *(ii)* the nodes in the Bayesian network have a more intuitive interpretation. On the other hand, the tensor network does highlight the importance of copying for there to be multiple child nodes, which will be important later for incorporating quantum nodes (see §3.2, §3.4, and §4.1).

2.5 The Copy map and restriction maps

Going along with the linear maps on measures induced by transition probability function (see (2.14)), we have the following additional useful linear maps for the evaluation of the joint probability for a directed graphical model. For a set $\mathcal{X}$ with σ-algebra $\mathcal{E}$, there is a Copy map from $\mathcal{E}$-measures on $\mathcal{X}$ to $\mathcal{F}$-measures on $\mathcal{X}\times\mathcal{X}$, where $\mathcal{F}$ is the σ-algebra generated by the rectangular sets $\mathcal{E}\times\mathcal{E}$. It is given by, for any set $A\in\mathcal{F}$ and $\mathcal{E}$-measure μ, $\text{Copy}(\mu)(A) = \mu(\{x\in\mathcal{X}|(x,x)\in A\})$. It is induced by the transition probability function $\tau(\cdot|\cdot)$ given by

$$\tau(A|x) = \begin{cases} 1 & \text{if } (x,x)\in A \\ 0 & \text{otherwise} \end{cases} \tag{2.29}$$

This can clearly be generalized for creating any finite number of copies.

For each $A\in\mathcal{E}$, there is a restriction map, which is an idempotent, sending $\mathcal{E}$-measures on $\mathcal{X}$ to $\mathcal{E}$-measures[12] on $\mathcal{X}$, $\mu\to 1_A\mu = \mu(A\cap\cdot)$. It is induced by the transition probability function $\tau(\cdot|\cdot)$ given by $\tau(A|x) = 1_A(x)$.

[12]We adopt the convention that a function before a measure, $f\mu$, is the signed measure $f\mu(A) = \int_A f\,d\mu$.

Chapter 3

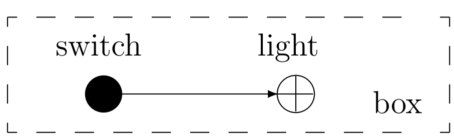
Hidden classic and quantum nodes

3.1 Principles of linearity and potential universality

The traditional approach to using Bayesian networks with directed graphical models gives a marginal probability measure for each input[1] node and a transition probability function for each of the remaining nodes. Then the following principle [36] is utilized, which is so reasonable it almost always goes unmentioned, but is just implicitly assumed:

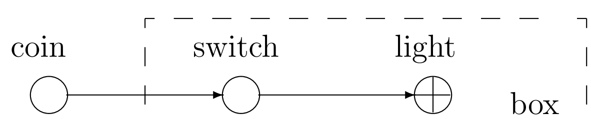
Principle 3.1.1–Measurement independence The input marginal probabilities can be varied independently of each other and the transition probability functions.[2]

As an example of both the utility and reasonableness of this principle, consider a box with a light and a switch. Suppose we know the light will turn on with 0.9 probability if the switch is in position one, and will turn on with 0.3 probability if the switch is in position two. The graphical model for this is:



(3.1)

Now suppose a fair coin is used to determine the position of the switch, so the graphical model is:



(3.2)

[1] All nodes without parents.

[2] Also known as the *free choice* or *free will* principle.

Using the preceding principle, then there is a $0.5 \cdot 0.9 + 0.5 \cdot 0.3 = 0.6$ probability the light will turn on. Without this principle or something similar, it would be impossible to predict the behavior of the light using the coin to determine the switch position just using our knowledge about what would happen with each position of the switch.

From the metaphysical viewpoint this principle makes perfect sense. If there are actually existing variables that take actual values (such as the position of the switch), observed outcomes depend only these and not on probability measures, which are only a reflection of our ignorance. As has already been commented on, this justification fails for hidden nodes, which are hypothetical constructs we introduce. Without this metaphysical backing, the principle is actually far stronger than what is required in that it assumes the existence of transition probability functions.

If the principle is assumed true, then, from (2.14), the calculation of the joint probability reduces to some combination of composition and tensor product of of linear maps, involving both those induced by the given transition probability functions, restriction maps, and possibly the Copy map (the last two of which are also induced by certain transition probability functions–see §2.5). By **A3.2** and **A3.3**, this calculation is well-defined. For instance, the calculation of the joint probability for (2.20), with $\theta(\cdot|\cdot)$ and $\zeta(\cdot|\cdot)$ as transition probability functions, can be given as[3,4]

$$(R_A \circ K \circ ((R_B \circ L) \otimes I) \circ \text{Copy } \circ R_C \nu)(\mathcal{X}) \tag{3.3}$$

for L the map induced by $\theta(\cdot|\cdot)$, K the map induced by $\zeta(\cdot|\cdot)$, I the identity map on measures, and R_A, R_B, R_C restriction maps. By introducing an initializing map, L_i, on the trivial measure space[5], which is isomorphic to $\mathbb{R}$, with constant value μ and a terminal map, L_t, which evaluates the measure on $\mathcal{X}$ (hence, is a map to the trivial measure space), this can be written purely in terms of maps:

$$L_t \circ R_A \circ K \circ ((R_B \circ L) \otimes I) \circ \text{Copy } \circ R_C \circ L_i \tag{3.4}$$

Later, we will introduce hidden nodes of a special form to account for initializing or terminating maps (see §4.1).

Therefore, the joint probability is a multilinear functional on the input marginal probabilities. Furthermore, by a convergence theorem for sequences of measures [56], it is not just linear for finite linear combinations in each marginal measure, but absolutely convergent countable linear combinations as well. Hence, for any one input node, if Φ is the functional

[3] We adopt the usual mathematical convention of maps acting on the left. The opposite convention of maps acting on the right is also commonly employed in the literature for the classical observed case [47].

[4] We will follow the convention that the tensor product of vector spaces consists of all finite linear combinations (the algebraic tensor product) except if both are Hilbert spaces, in which case it is the Hilbert space given by the completion using the standard induced inner-product. The tensor product $K \otimes L$ of linear maps $L : \mathsf{A} \to \mathsf{C}$ and $K : \mathsf{B} \to \mathsf{D}$ is, for closed, linear spaces $\mathsf{E} \subset \mathsf{A} \otimes \mathsf{B}$ and $\mathsf{F} \subset \mathsf{C} \otimes \mathsf{F}$, the set of all linear maps $M : \mathsf{E} \to \mathsf{F}$ which, when restricted to $\mathsf{A} \otimes \mathsf{B}$, agree with $K \otimes L$. If this set consists of a single map, $K \otimes L$ is termed *well-defined*.

[5] A measure on a set $\mathcal{X}$ is *trivial* if the only subsets in its σ-algebra are $\{\varnothing, \mathcal{X}\}$.

(with all the marginal probability measures for other nodes fixed) and $\langle \mu_j \rangle$ is a sequence of marginal probability measures for that node with $\sum_j \|\mu_j\|$ finite, then $\Phi\left(\sum_j \mu_j\right) = \sum_j \Phi\mu_j$. It is obvious that it would still be possible to make the calculation for the above example with the box based simply on this linearity property, which is weaker than the above property.

As we have already mentioned, we will consider hidden nodes associated to maps on spaces other than the space of measures. Therefore, we introduce the following weaker and more general version of the above principle:

Principle 3.1.2–Linearity The maps for a Bayesian network are linear and bounded (hence, continuous).

Moreover, we want the maps for graph fragments to be universal in the sense that the modules (such as the box in the preceding example) can be used to make predictions in novel situations. While this may fail in practice, this should be as a result of experiment and not be preordained by the mathematical models employed. We insist, therefore, on potential universality in the sense that any possible tensor product (not just those for a particular network) of the linear maps employed should always be well-defined.

Principle 3.1.3–Potential universality The space of linear maps employed must be such that any tensor product of maps in the space is well-defined.

The repercussions of the latter two principles will be studied in the following.

3.2 Options I and II

A problem arising from potential universality

Generalizing from maps induced by transition probability functions to more general linear maps, composition is not an issue. However, the principle of potential universality does not hold in general for maps on measures on specified sets for specified σ-algebras of events. One problem is the lack of uniqueness. Suppose one has the identity map I on Borel[6] measures on the interval $[0,1]$ with the usual topology. Define $I \otimes I$ to be the set of all bounded, linear maps L on the Borel measures on $[0,1] \times [0,1]$, with the usual product topology, such that, restricted to product measures, $\mu \times \nu$, $L(\mu \times \nu) = (I\mu) \times (I\nu) = \mu \times \nu$. One obvious member of $I \otimes I$ is simply the identity map on the Borel measures on $[0,1] \times [0,1]$. This is the only weak*-continuous[7] map in the set. Another map in the set is given by $K : \rho \to \bigvee(\rho \parallel (\mu \times \nu))$, where

[6] The *Borel* σ-algebra on a topological space is that generated by the open subsets.

[7] Using the Riesz theorem [57] which states that Radon (defined in the following) measures on a compact set are dual to the continuous functions on that set. A measure on a topological space is *inner regular* if the measure of any set is approximated by the measure of compact sets it contains. It is *outer regular* if the measure of any set is approximated by the open sets that contain it. A measure is *Radon* if it is Borel and

the supremum is taken over all product measures and $\rho \parallel (\mu \times \nu)$ is the part of ρ absolutely continuous with respect to $\mu \times \nu$ using Lebesgue decomposition [55]. This is a well-defined map by **B2.1** that differs from the identity map. For instance, let ρ be the diagonal Lebesgue measure,

$$\rho(A) = \lambda\left(\{a \in [0,1] | (a,a) \in A\}\right) \tag{3.5}$$

Then the identity map sends ρ to itself, whereas $K\rho = 0$.

Another possible problem is the lack of existence. Not every linear map can be extended. For example, the space of sequences with limit zero, c_0, is a norm-closed[8], weak*-dense[9] subset of the bounded sequences, ℓ_∞. However, there is no extension of the identity map $c_0 \to c_0$ to a projection $\ell_\infty \to c_0$. [3] [90] However, it is not clear whether there is a similar problem with extending the tensor product of linear maps.

Resolving the problem–two options

There are various ways to resolve this dilemma. One is to revert to only considering linear maps induced by transition probability functions, which works by **A3.3**. We will not pursue this approach since it has no ready generalization to linear maps on spaces other than those of measures. Instead, we will consider two approaches which do. The first, which will be termed option **I**, is to limit the space of measures so as to eliminate measures such as ρ above. The alternative, which will be termed option **II**, is to impose additional structure on the sets and then to limit the space of maps, so as to eliminate maps such as K above.

To implement option **I**, a basic property that we want for our subsets of measures is defined by:

Definition 3.2.1 A subset A of measures is *absolutely-continuous-complete* if, for any μ in A, all measures absolutely continuous with respect to μ are also in A.

By the Radon-Nikodým theorem and the density of simple functions[10] in L^1-spaces, a subset of finite measures A will have this property if, given any measure $\mu \in A$, all the restrictions $\mu|_E$ over μ-measurable subsets E are also in A. Hence, this property is the minimal requirement to show that a map is indeed positive. Then from **A1.2**, **A1.3**, **B1.3**, **B1.4**, and **B2.5**, the necessary and sufficient condition to implement option **I** is that the tensor product of the subsets of measures for each set is norm-dense[11] in the subset of measures for the product set.

---

inner regular. If the space is compact and Hausdorff, and the measure is finite, then it is also necessarily outer regular. If the space is compact, Hausdorff, metric, and separable, then Borel measures are necessarily Radon [58].

[8] Using the supremum norm.

[9] Using the duality of ℓ_1 and ℓ_∞.

[10] A *simple function* takes on only finitely many values.

[11] Using the total-variation norm.

A sufficient method, termed option **I'**, to insure this is met is for each set to have an associated *base measure*. The base measure need not be finite, but it must be σ-finite. Then the associated subset of measures is the set of all finite measures absolutely continuous with respect to the base measure, which by the Radon-Nikodým theorem is equivalent to the space of L^1-functions with respect to the base measure. The base measure for the direct product of sets must be the product of the base measures for the individual sets. The sufficiency of this prescription is given by **B3.1**.

Another sufficient method is to instead use the space of atomic measures[12]. If there are uncountably many atoms in the σ-algebra, this is distinct from option **I'**; otherwise, the counting measure that assigns one to each atom is a base measure. The sufficiency of this prescription is given by **B6.1** and **B6.2**. It is also possible to combine these two sufficient approaches, say by using the atomic, $L^1(\mathcal{X};\mu)$-valued vector $\mathcal{E}$-measures on $\mathcal{Y}$. This is sufficient by **B3.1**, **B6.1**, and **B6.2**.

The implementation of option **II** is more straightforward. Each set has a topological structure that makes it a compact, Hausdorff space. By Tychonoff's theorem [42] [59], the direct product of compact spaces with the product topology is necessarily compact. All the σ-algebras are required to be the Borel σ-algebra. The linear maps are restricted to those that are weak* continuous; in other words, those maps that are the adjoints to linear maps on continuous functions going in the opposite direction. In practice, rather that working with the adjoint maps, one works with the linear maps in the opposite direction. The propositions **A1.3**, **C1.1**, and **C2.4** gives the sufficiency of this prescription.

Comments on the two options

For option **I'**, the need for base measures is not generally a troublesome issue. For σ-algebras generated by a countable number of atomic subsets[13], the counting measure that assigns one to each atom is a base measure for any finite measure. For classical physics, with configuration space $\mathcal{X}$ and phase-space given by the cotangent bundle $T^*\mathcal{X}$, the symplectic phase-space volume-form[14] Ω provides a natural base measure, since by Heisenberg's uncertainty principle, not even all measures absolutely-continuous with respect to Ω are accessible, let alone more singular measures.

Note that allowing the base measure to be a σ-finite measure is only for purposes of convenience in allowing the commonly employed Lebesgue measure on unbounded subsets of $\mathbb{R}^n$, as the following theorem shows:

Theorem 3.2.2 Given any σ-finite measure μ on a set $\mathcal{X}$, there is a finite measure ν such that $L^1(\mathcal{X};\mu)$ is isometric to $L^1(\mathcal{X};\nu)$, where the isomorphism is a pointwise scaling.

[12] A measure is *atomic* if there is a union of countably many atoms in the σ-algebra such that the complement of the union has measure zero.

[13] A subset in a σ-algebra is *atomic* if it is indivisible in the σ-algebra.

[14] As a measure, locally Ω is simply Lebesgue measure with respect to any canonical choice of local position and momentum coordinates.

Proof If μ is finite, there is nothing to show, so assume it is not. Since μ is σ-finite, there is a countable collection of disjoint subsets $\{B_j\}$ of $\mathcal{X}$ such that $\mu(B_j)$ is finite and nonzero for each $j \in \{1, 2, \ldots\}$ and $\bigcup_{j=1}^{\infty} B_j = \mathcal{X}$. Then let the finite measure ν be given by $\nu|_{B_j} = \dfrac{\mu|_{B_j}}{2^j \mu(B_j)}$.
□

However, there are two complaints with option **I'**. One is that for σ-algebras not generated by a countable number of atomic subsets, there is no Copy map; hence, except for this (effectively discrete) case, hidden nodes either have only one child node or are terminated. The second complaint is that passing a continuously variable parameter to a hidden node as a simple number is not permitted (unless that particular value corresponds to an atom in the base measure); instead, one must use a sharply peaked measure. This adds significant complexity for little gain in cases where one is not especially interested in modelling uncertainty in the inputs (see §4.5 and §8.2 for instances).

For option **II**, the restriction to topological spaces and Borel σ-algebras is also not troublesome, since these are typically used in any case. The limitation of using compact spaces appears severe, but locally compact spaces[15] can also be used with the restriction that the maps take continuous functions vanishing at infinity to continuous functions vanishing at infinity, so the adjoint maps on measures do not "leak away" measure at infinity.

In addition, for option **II**, the first complaint above does not occur since the Copy map is adjoint to the map Copy_* that takes continuous functions on $\mathcal{X} \times \mathcal{X}$ to continuous functions on $\mathcal{X}$ by $(\text{Copy}_* f)(x) = f(x, x)$ (which can obviously be generalized to make any finite number of copies). The second complaint does not occur either since the evaluation map is well-defined for continuous functions. However, there is now the opposite problem in that we wish to calculate probabilities on sets, so we need maps on characteristic functions, not just continuous ones. One solution is to extend each map to one from bounded, Borel measurable functions to bounded, Borel measurable functions; by **C2.9** this can always be done in a unique manner. Another solution is to use the results on continuous functions to get the result for characteristic functions of open sets as in the proof of the Riesz theorem [57]; then outer regularity gives the result on any characteristic function of a Borel set.

Note that for option **I'**, the considered linear map $L : L^1(\mathcal{X}; \mu) \to L^1(\mathcal{Y}; \nu)$ is always induced by an object $\tau(\cdot|\cdot)$ which is given by, for ν-measurable sets B, $\tau(B|\cdot) = L^* 1_B$ (which is of course actually an equivalence class of functions that agree almost everywhere with respect to μ) with the adjoint map $L^* : L^\infty(\mathcal{Y}; \nu) \to L^\infty(\mathcal{X}; \mu)$:

$$\begin{aligned} \int_{y \in B} (Lf)(y)\, d\nu(y) &= \int_{y \in \mathcal{Y}} 1_B (Lf)(y)\, d\nu(y) = \int_{x \in \mathcal{X}} f(x)(L^* 1_B)(x)\, d\mu(x) \\ &= \int_{x \in \mathcal{X}} f(x) \tau(B|x)\, d\mu(x) \end{aligned} \tag{3.6}$$

[15] A space is *locally compact* if it can be compactified by the addition of one point, the *point-at-infinity*. [43]. [60]

for any $f \in L^1(\mathcal{X};\mu)$. It is not clear whether it is possible to select a particular function from each equivalence class consistently to get a transition probability function. Similarly, if we use option **I** with the measures limited to the atomic measures, the considered linear map $L : \mathcal{A}(\mathcal{X};\mathcal{E}) \to \mathcal{A}(\mathcal{Y};\mathcal{F})$ is always induced by an object $\tau(\cdot|\cdot)$ which is given by, for sets $B \in \mathcal{F}$, $\tau(B|x) = (L\delta_A)(B)$ for x in the atomic set $A \in \mathcal{E}$:

$$(L\mu)(B) = \int_{x\in\mathcal{X}} \tau(B|x)\, d\mu(x) \tag{3.7}$$

for any atomic measure $\mu \in \mathcal{A}(\mathcal{X};\mathcal{E})$. However, in general $\tau(\cdot|\cdot)$ will not be a transition probability function since there is no reason for $\tau(B|\cdot)$ to necessarily be $\mathcal{E}$-measurable. Also, for option **II**, by **C2.8**, for the considered linear map $L : \mathcal{C}(\mathcal{Y}) \to \mathcal{C}(\mathcal{X})$, the adjoint map $L^* : \mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{Y})$ is induced by the transition probability function $\tau(\cdot|\cdot)$ given by $\tau(B|x) = (L\delta_x)(B)$ for any $x \in \mathcal{X}$ and Borel subset $B \subset \mathcal{Y}$:

$$(L^*\mu)(B) = \int_{x\in\mathcal{X}} \tau(B|x)\, d\mu(x) \tag{3.8}$$

for any Radon measure $\mu \in \mathcal{M}(\mathcal{X})$. However, it is more fruitful to consider the maps themselves as the primary objects of interest rather than the transition probability functions or the similar objects. As is shown in the following section, the maps can be generalized to be maps on structures other than measures, whereas the transition probability functions or similar objects do not.

3.3 Quantum nodes

Expanding the space of considered maps

As an alternative to linear maps on measures, consider linear maps on density matrices[16,17] or, more generally, density matrix-valued measures. We will show below that, with some restrictions on the maps, this can be made to work consistently with the propositions given in §3.1. We will show in §5.1 that this gives rise to models that are consistent with the usual, textbook quantum mechanics; therefore, nodes whose linear maps involve density matrices will be termed *quantum*. However, the term *quantum* should not be taken to imply these maps are necessarily only of utility in situations traditionally thought of as in the

[16]A *Hilbert space* will be taken to be any complete, sesquilinear inner-product space, without regard to cardinality of dimension or separability. The *density matrices* $\mathcal{D}(\mathsf{H})^+$ will be taken to be the self-adjoint, positive operators on a given Hilbert space H.

[17]A linear map from $\mathcal{D}(\mathsf{H})$ to $\mathcal{D}(\mathsf{J})$ is commonly referred to as a *superoperator* in the literature. We choose not to employ this terminology for the following reasons: *(i) linear maps* is already standard mathematical terminology and is in common use in the analogous classical situation, for instance *Markov maps*; *(ii) superoperator* seems to imply a map on all bounded operators, $\mathcal{B}(\mathsf{H})$, when in general it is not possible to extend the domain of the map beyond the trace-class operators, $\mathcal{S}_1(\mathsf{H})$; and *(iii)* the use of *super-* risks confusion with the unrelated *supersymmetry* and *superstrings*.

domain of quantum mechanics–in principle, any hidden node in any forecasting situation, say forecasting weather or stock prices, could be a quantum node. It is up to experiment to determine if these are of utility. Thus far, we are unaware of any structures besides measures and density matrices that have the requisite properties to be employed in modelling. The question of whether or not there are such additional structures will be further explored in §3.8 below.

Following our work in the preceding section, for option **I**, instead of linear maps on subsets of real-valued measures, we have linear maps on subsets of the $\mathcal{D}(\mathsf{H})^+$-valued vector measures. For option **I'**, we will require the vector measures to be absolutely continuous with respect to a base measure. Since $\mathcal{D}(\mathsf{H})$ has the Radon-Nikodým property [14], this is equivalent to having $\mathcal{D}(\mathsf{H})^+$-valued, Bochner-integrable functions. The constraint of potential universality mandates having well-defined tensor product of maps; this is maintained by **A1.3** and **B4.1** if the tensor product is bounded.

Similarly to the classical case, another sufficient approach to implementing option **I** is to take the atomic, $\mathcal{D}(\mathsf{H})^+$-valued vector measures, as is shown by **B6.1** and **B6.2**. It is also possible to combine these two sufficient approaches, say by using the atomic, $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$-valued vector $\mathcal{E}$-measures on $\mathcal{Y}$. This is sufficient by **B4.1**, **B6.1**, and **B6.2**.

For option **II**, instead of maps on real-valued, positive, continuous functions, one has maps on continuous functions that take values in the self-adjoint, positive, compact[18] operators on a Hilbert space H, $\mathcal{K}(\mathsf{H})^+$. Self-adjoint, compact operators are used since they are the predual to the self-adjoint trace-class operators. [50] Once again, potential universality mandates having well-defined tensor products of maps; as before, this is maintained if the tensor product is bounded, now by **A1.3** and **C3.1**.

Problem arising from positivity and potential universality

Of course, everything is not really that simple. The tensor product of bounded maps may not be bounded. Also, even if both maps are positive, their tensor product need not be. To illustrate these problems, take a separable, infinite-dimensional Hilbert space H, fix some orthonormal basis $\{\mathbf{e}_j\}$, and consider the *transpose map* T relative to that basis:

$$T\left(\sum_{j,k} a_{jk}\mathbf{e}_j \otimes \mathbf{e}_k^*\right) = \sum_{j,k} a_{kj}\mathbf{e}_j \otimes \mathbf{e}_k^* \tag{3.9}$$

where $\mathbf{e}_k^*$ is the functional $\langle\cdot,\mathbf{e}_k\rangle$. This is well-defined on the space of density-matrices on H, $\mathcal{D}(\mathsf{H})^+$, since, by the spectral theorem for compact operators [17], any such operator can be written in the form of an infinite matrix with finite rank operators corresponding to truncated matrices converging in trace norm. (Of course, since density-matrices are self-adjoint, T could also be termed the *conjugate map* relative to the basis). This map is clearly positive and has operator norm one.

[18] An operator is *compact* if the image of any bounded sequence has a convergent subsequence.

However, consider the tensor product map $T \otimes I_{\mathcal{M}_n}$ acting on density matrices $\mathcal{D}(\mathsf{H} \otimes \mathbb{C}^n)^+$, where $I_{\mathcal{M}_n}$ is the identity map acting on $n \times n$-matrices. For the rank-one operator $\psi \otimes \psi^*$ with

$$\psi = \sum_{j=0}^{n-1} \mathbf{e}_{(n+1)j+1} \tag{3.10}$$

we have $(T \otimes I_{\mathcal{M}_n})(\psi \otimes \psi^*) = S$. Truncating S to the span of $\{\mathbf{e}_1, \ldots, \mathbf{e}_{n^2}\}$ (it is zero elsewhere), it is the matrix form of the transpose map acting on $\mathcal{M}_n$ written in vector form using the Vec operation[19]. Therefore, S clearly has eigenvalue one with multiplicity $\frac{n(n+1)}{2}$ and eigenvalue minus one with multiplicity $\frac{n(n-1)}{2}$. Hence, $T \otimes I_{\mathcal{M}_n}$ is not positive, and

$$\|T \otimes I_{\mathcal{M}_n}\|_{\text{op}} \geq \frac{1}{\|\psi\|^2} \left(\frac{n(n+1)}{2} |1| + \frac{n(n-1)}{2} |-1| \right) = \frac{n^2}{n} = n \tag{3.11}$$

By [19], $\|T \otimes I_{\mathcal{M}_n}\|_{\text{op}} = n$ and this example is maximal. This is clearly unbounded as $n \to \infty$.

Solution to the problem

The solution to both the positivity and the boundedness problem is to require complete-positivity for the maps. This has several definitions (see **B2.6**, **B5.6**, **C2.1**) that are equivalent (see **B5.8**, **C5.6**); the basic notion is that all tensor products with various identity maps should be positive. Since the composition of positive maps is positive, this immediately implies that complete-positivity is preserved under both composition and tensor products. Furthermore, from **B2.5**, **B2.5**, **C2.4**, and **C5.3**, the operator norm is a cross-norm[20] for completely-positive maps, so this resolves the boundedness problem as well. The completely-positive maps clearly form a convex cone within the space of all maps; this cone is closed in the norm topology and in various weaker topologies by **B5.12**, **B5.14**, and **C5.10**; however, unlike the cone of positive maps, in infinite dimensions it has no interior in any of these topologies, which raises issues for approximation in numerical calculation.

3.4 No quantum copying

It is commonly stated that cloning is something that is possible classically, but is impossible in quantum mechanics. This is based on false analogy. The correct situation is that there are two different notions, that of copying and that of cloning, that are being confused. Once

[19] Vec takes a $n \times n$-matrix to a column vector of height n^2 by stacking columns.

[20] A norm is a *cross-norm* if $\|\mathsf{a} \otimes \mathsf{b}\| \leq \|\mathsf{a}\| \|\mathsf{b}\|$. Note the property of being a cross-norm depends on the choice of norms for the individual spaces as well as for the larger space containing the tensor products.

these two have been separated, we have the following situation:

	classical	quantum
copying	Exists and implementable since linear.	Does **not** exist.
cloning	Exists but **not** implementable since neither linear nor the ratio of linear maps.	Exists but **not** implementable since neither linear nor the ratio of linear maps.

The confusion is comparing the upper left and lower right entries instead of correctly going across. Cloning is possible for neither classical nor quantum Bayesian networks (as will be shown below in §6.1) for exactly the same reason, so it does not differentiate the two. On the other hand, copying is possible classically (except for issues arising from potential universality considered above in §3.2), but cannot even be defined as a mathematical operation on density matrices.

Classical copying

As has already been mentioned (see §2.5), there is a Copy map from measures on a set $\mathcal{X}$ to measures on $\mathcal{X} \times \mathcal{X}$. When $\mathcal{X}$ is a compact set, this map is weak*-continuous, being the adjoint of the previously discussed map $(\text{Copy}_* f)(x) = f(x, x)$. The Clone map is given by $\mu \to \mu \times \mu$. For the single atom measure for atom C, where for any measurable subset $A \subset \mathcal{X}$,

$$\delta_C(A) = \begin{cases} 1 & \text{if } A \supset C \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

we have

$$\text{Copy } \delta_C = \delta_C \times \delta_C = \text{Clone } \delta_C \tag{3.13}$$

This is likely the source of confusion between the Copy and Clone maps for the classical case.

Instead of using this explicit form for Copy, an approach that will prove useful in the quantum case is to start with some basic properties, then find the implications. One property of what is commonly accepted as the notion of a copy is that the probability for both copies to have a specified property is equal to that for each copy to have it, which is equal to that of the original having it, so for any unit-norm measure μ on $\mathcal{X}$ and any μ-measurable set $A \subset \mathcal{X}$,

Property C $\text{Copy } \mu(A \times A) = \text{Copy } \mu(A \times \mathcal{X}) = \text{Copy } \mu(\mathcal{X} \times A) = \mu(A)$

Note this property implies $\text{Copy } \mu(A \times (\mathcal{X} \smallsetminus A))) = \text{Copy } \mu((\mathcal{X} \smallsetminus A) \times A) = 0$. Now given a σ-algebra $\mathcal{E}$ of subsets of $\mathcal{X}$, let $\mathcal{F}$ be the σ-algebra generated by the rectangular subsets $\mathcal{E} \times \mathcal{E}$. Then we have the following:

Theorem 3.4.1 Any map L from unit-norm $\mathcal{E}$-measures on $\mathcal{X}$ to unit-norm $\mathcal{F}$-measures on $\mathcal{X} \times \mathcal{X}$ obeying property **C** is linear for convex linear combinations.

Comment Since any finite measure can be scaled to have unit-norm, this implies the map can be extended to all finite measures, with the extended map being positively linear. By the generating property of measures among signed measures as a result of Jordan decomposition [61], this implies the map can further be extended to a linear map on signed measures.

Proof Let L be such a map and μ an unit-norm $\mathcal{E}$-measure on $\mathcal{X}$. For any subsets $A, B \in \mathcal{E}$, by the properties of measures,

$$\begin{aligned}(L\mu)(A \times B) =&(L\mu)((A \cap B) \times (A \cap B)) + (L\mu)((A \cap B) \times (B \setminus (A \cap B))) \\ &+ (L\mu)((A \setminus B) \times B)\end{aligned} \tag{3.14}$$

However, $(A \cap B) \times (B \setminus (A \cap B)) \subset (A \cap B) \times (\mathcal{X} \setminus (A \cap B))$ and $(A \setminus B) \times B \subset (\mathcal{X} \setminus B) \times B$, so, by property **C** and its implication,

$$(L\mu)(A \times B) = (L\mu)((A \cap B) \times (A \cap B)) = \mu(A \cap B) = (L\mu)(B \times A) \tag{3.15}$$

Let ρ be another unit-norm $\mathcal{E}$-measure on $\mathcal{X}$. Then for any $t \in [0,1]$, $(1-t)\rho + t\mu$ will be a unit-norm $\mathcal{E}$-measure on $\mathcal{X}$. Consider the signed measure on $\mathcal{X} \times \mathcal{X}$ given by

$$\nu_t = L((1-t)\rho + t\mu) - (1-t)L\rho - tL\mu \tag{3.16}$$

Take any $A \in \mathcal{E}$. Then, by property **C**, $\nu_t(A \times A) = 0$. However, by the above symmetry property of L,

$$\nu_t(A \times B) = \frac{1}{2}\left(\nu_t(A \times B) + \nu_t(B \times A)\right) \tag{3.17}$$

which is equal to

$$\begin{aligned}&\tfrac{1}{2}\left(\nu_t((A \cup B) \times (A \cup B)) - \nu_t((A \setminus B) \times (A \setminus B))\right. \\ &\left.-\nu_t((B \setminus A) \times (B \setminus A)) + \nu_t((A \cap B) \times (A \cap B))\right)\end{aligned} \tag{3.18}$$

which is zero by the preceding property of ν_t. Since A, B were arbitrary, ν_t must be the zero measure. □

Quantum Copying

The Clone map taking a density matrix on the Hilbert space H to one on $\mathsf{H} \otimes \mathsf{H}$ is defined as $\rho \to \rho \otimes \rho$. How to define a Copy map is not obvious. By analogy to the classical case, it should have the following properties for any unit-trace density matrix ρ and any projector E:

Property Q tr $(E \otimes E)$ Copy ρ = tr $(E \otimes I_{\mathsf{H}})$ Copy ρ = tr $(I_{\mathsf{H}} \otimes E)$ Copy ρ = tr $E\rho$

Note this implies tr $(E \otimes (I_{\mathsf{H}} - E))$ Copy ρ = tr $((I_{\mathsf{H}} - E) \otimes E)$ Copy $\rho = 0$. Then we have the following:

Theorem 3.4.2 Any map L from unit-trace density matrices on H to unit-trace density matrices on $\mathsf{H} \otimes \mathsf{H}$ obeying property **Q** is linear for convex linear combinations.

Comment Since any density matrix can be scaled to have trace one, this implies the map can be extended to all density matrices with the extended map being positively linear. By the generating property of density matrices among signed density matrices as a result of the spectral theorem for compact operators, this implies the map can further be extended to a linear map.

Proof Let L be such a map and ρ an unit-trace density matrix on H. For any commuting projectors E, F,

$$\begin{aligned}\text{tr } (E \otimes F)(L\rho) =& \text{tr } (EF \otimes EF)(L\rho) \\ &+ \text{tr } (EF \otimes (F - EF))(L\rho) + \text{tr } ((E - EF) \otimes F)(L\rho)\end{aligned} \tag{3.19}$$

However, by positivity and the implication of **Q**,

$$0 \leq \text{tr } (EF \otimes (F - EF))(L\rho) \leq \text{tr } (EF \otimes (I_{\mathsf{H}} - EF))(L\rho) = 0 \tag{3.20}$$

and

$$0 \leq \text{tr } ((E - EF) \otimes F)(L\rho) \leq \text{tr } ((I_{\mathsf{H}} - F) \otimes F)(L\rho) = 0 \tag{3.21}$$

so, using **Q**,

$$\text{tr } (E \otimes F)(L\rho) = \text{tr } (EF \otimes EF)(L\rho) = \text{tr } EF\rho = \text{tr } (F \otimes E)(L\rho) \tag{3.22}$$

Let τ be another unit-trace density matrix on H. Then for any $t \in [0,1]$, $(1-t)\rho + t\tau$ will be a unit-trace density matrix on H. Consider the signed density matrix on $\mathsf{H} \otimes \mathsf{H}$ given by

$$\nu_t = L((1-t)\rho + t\tau) - (1-t)L\rho - tL\tau \tag{3.23}$$

Take any projector E. Then, by property **Q**, tr $(E \otimes E)\nu_t = 0$. However, by the above symmetry property of L, for any commuting projectors E, F,

$$\text{tr } (E \otimes F)\nu_t = \frac{1}{2}(\text{tr } (E \otimes F)\nu_t + \text{tr } (F \otimes E)\nu_t) \tag{3.24}$$

which is equal to

$$\begin{aligned}&\frac{1}{2}(\text{tr } ((E + F - EF) \otimes (E + F - EF))\nu_t - \text{tr } ((E - EF) \otimes (E - EF))\nu_t \\ &-\text{tr } ((F - EF) \otimes (F - EF))\nu_t + \text{tr } (EF \otimes EF)\nu_t)\end{aligned} \tag{3.25}$$

which is zero by the preceding property of ν_t. Since E, F were arbitrary, ν_t must be the zero operator. □

We then have the following theorem, based on the argument of Wooters and Zurek [92] that density matrices of rank greater than one can be expressed in more than one way (infinitely many ways actually) to create a contradiction.

Theorem 3.4.3 There is no quantum Copy map for non-trivial[21] H.

Proof Take any orthonormal $\{\mathbf{u}, \mathbf{v}\} \subset \mathsf{H}$ with corresponding adjoint operators $\mathbf{u}^* = \langle \cdot, \mathbf{u} \rangle_{\mathsf{H}}$ and $\mathbf{v}^* = \langle \cdot, \mathbf{v} \rangle_{\mathsf{H}}$. Consider $\rho = \frac{1}{2}(\mathbf{u} \otimes \mathbf{u}^* + \mathbf{v} \otimes \mathbf{v}^*)$. By linearity,

$$\text{Copy } \rho = \frac{1}{2}(\text{ Copy } (\mathbf{u} \otimes \mathbf{u}^*) + \text{ Copy } (\mathbf{v} \otimes \mathbf{v}^*)) \tag{3.26}$$

By property **Q**, for rank one density matrices Copy must be the same as Clone, so Copy ρ is uniquely given as

$$\frac{1}{2}(\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u}^* \otimes \mathbf{u}^* + \mathbf{v} \otimes \mathbf{v} \otimes \mathbf{v}^* \otimes \mathbf{v}^*) \tag{3.27}$$

However, it is also possible to write ρ as

$$\frac{1}{4}((\mathbf{u} + \mathbf{v}) \otimes (\mathbf{u} + \mathbf{v})^* + (\mathbf{u} - \mathbf{v}) \otimes (\mathbf{u} - \mathbf{v})^*) \tag{3.28}$$

Then Copy ρ is uniquely given as

$$\begin{aligned} &\frac{1}{8}((\mathbf{u} + \mathbf{v}) \otimes (\mathbf{u} + \mathbf{v}) \otimes (\mathbf{u} + \mathbf{v})^* \otimes (\mathbf{u} + \mathbf{v})^* \\ &\quad +(\mathbf{u} - \mathbf{v}) \otimes (\mathbf{u} - \mathbf{v}) \otimes (\mathbf{u} - \mathbf{v})^* \otimes (\mathbf{u} - \mathbf{v})^*) \end{aligned} \tag{3.29}$$

$$\begin{aligned} = \frac{1}{4}(&\mathbf{u} \otimes \mathbf{u} \otimes \mathbf{u}^* \otimes \mathbf{u}^* + \mathbf{u} \otimes \mathbf{u} \otimes \mathbf{v}^* \otimes \mathbf{v}^* + \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{u}^* \otimes \mathbf{v}^* \\ &+ \mathbf{u} \otimes \mathbf{v} \otimes \mathbf{v}^* \otimes \mathbf{u}^* + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{u}^* \otimes \mathbf{v}^* + \mathbf{v} \otimes \mathbf{u} \otimes \mathbf{v}^* \otimes \mathbf{u}^* + \\ &+\mathbf{v} \otimes \mathbf{v} \otimes \mathbf{u}^* \otimes \mathbf{u}^* + \mathbf{v} \otimes \mathbf{v} \otimes \mathbf{v}^* \otimes \mathbf{v}^*) \end{aligned} \tag{3.30}$$

Clearly, (3.27) and (3.30) are unequal, which is a contradiction. □

3.5 Embedding quantum models into classical ones

A construction for option I' using atomic measures

For option **I'**, there is a way to embed quantum behavior into a purely classical, but contextual, model. For any Hilbert space H, let $\mathbb{S}_{\mathsf{H}}$ be the closed unit ball within H. Let $\mathbb{S}_{\mathsf{H}}/\sim$ be

[21] A *trivial* Hilbert space has dimension one.

the quotient set formed from $\mathbb{S}_{\mathsf{H}}$ by the equivalence relation $\psi \sim \xi$ if there is a phase[22] w such that $\psi = w\xi$. Clearly $\mathbb{S}_{\mathsf{H}}/\sim$ is in one-to-one correspondence to rank-one projectors on H. Let $\mathcal{E}$ be any σ-algebra on $\mathbb{S}_{\mathsf{H}}/\sim$ such that all points are atoms (such as the Borel σ-algebra). For any set $\mathcal{X}$ and base measure μ denote the space of atomic, finite-norm, $L^1(\mathcal{X};\mu)$-valued vector $\mathcal{E}$-measures on $\mathbb{S}_{\mathsf{H}}/\sim$ by $\mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))$. This is a Banach space by **B6.1**. Given $\tau \in \mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))$, for any μ-measurable subset $B \subset \mathcal{X}$, define τ_B to be the atomic, signed $\mathcal{E}$-measure on $\mathbb{S}_{\mathsf{H}}/\sim$ defined by $\tau_B(A) = \int_B \tau(A)\,d\mu$ for any $A \in \mathcal{E}$. Let $\sim'$ be the equivalence relation on $\mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))$ given by $\tau \sim' \chi$ if, for any μ-measurable subset $B \subset \mathcal{X}$,

$$\int_{s\in\mathbb{S}_{\mathsf{H}}/\sim} ss^*\,d\tau_B(s) = \int_{s\in\mathbb{S}_{\mathsf{H}}/\sim} ss^*\,d\chi_B(s) \tag{3.31}$$

Since the equivalence class using $\sim'$ of zero is a closed, linear subspace, the quotient space $\mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))/\sim'$ is a Banach space using the standard norm for quotient spaces, $\|[\tau]\| = \inf_{\chi\in[\tau]}\|\chi\|$. Define the positive cone on the quotient space to be those equivalence classes with a positive member, using the obvious notion of positivity on $\mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))$.

Then we have the following theorem:

Theorem 3.5.1 There is a positive, linear, isometric isomorphism,

$$L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})) \cong \mathcal{A}\left(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu)\right)/\sim'$$

Proof If the measure μ is trivial, then $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})) \cong \mathcal{D}(\mathsf{H})$. Define the map $\Psi : L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})) \to \mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))/\sim'$ by $\Psi(\rho)$ being the equivalence class of $\sum_j a_j\delta_{[\psi_j]}$ for $\rho = \sum_j a_j\psi_j\psi_j^*$ with countable collections $\{\psi_j\} \in \mathbb{S}_{\mathsf{H}}$ and $\{a_j\} \subset \mathbb{R}$, which is always possible by the spectral theorem for compact operators.

For more general measures on $\mathcal{X}$, first start with the observation that, given any $\rho \in L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$, each $\rho(x)$ lives in the same separable subspace of H for almost every x with respect to μ, namely the subspace G that the operator $\int_{x\in\mathcal{X}}|\rho(x)|\,d\mu \in \mathcal{D}(\mathsf{H})$ lives in. Let $\{\mathbf{e}_j\}$ be an orthonormal basis for G. For each $j \in \{1,2,\ldots\}$, let P_j be the orthogonal projector onto the span of $\{\mathbf{e}_1,\mathbf{e}_2,\ldots,\mathbf{e}_j\}$. Since simple functions are norm-dense in $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$, the sequence $\langle P_j\rho P_j\rangle$ is a Cauchy sequence by **C4.3**; hence, it converges in norm by the completeness of the Banach space $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$. It is readily seen that the limit point is ρ. Now define the map $\Psi : L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})) \to \mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))/\sim'$ by first defining $\Psi(\psi\psi^* f)$, for $\psi \in \mathsf{H}$ and $f \in L^1(\mathcal{X};\mu)$ to be the equivalence class of the vector measure $f \otimes \delta_{[\psi]}$. Since the linear space $\mathcal{D}(P_m\mathsf{H})$ is spanned by the m^2 operators

$$\{\mathbf{e}_j\mathbf{e}_j^*\}_{j\in\{1,\ldots,m\}} \cup \{(\mathbf{e}_j+\mathbf{e}_k)(\mathbf{e}_j+\mathbf{e}_k)^*, (\mathbf{e}_j+\imath\mathbf{e}_k)(\mathbf{e}_j+\imath\mathbf{e}_k)^*\}_{j,k\in\{1,\ldots,m\},j<k} \tag{3.32}$$

the map Ψ can be extended by linearity to all ρ with the property that $\rho(x)$ lives on the same finite-dimensional subspace of H for almost every x with respect to μ.

[22] Elements of $\mathbb{C}$ with magnitude one.

Now take any ρ with this property which is also a simple function. Then, applying the (finite-dimensional) spectral theorem to each of the finitely many values ρ takes, Ψ is readily seen to be a positive isometry. Since simple functions are norm-dense, by **A1.2**, Ψ is a positive isometry on all ρ with the property that $\rho(x)$ lives on the same finite-dimensional subspace of H for almost every x with respect to μ. However, by an above argument, such ρ are norm-dense in $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$. Hence, using **A1.3** to extend Ψ to all of $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$, Ψ is a positive isometry. It remains to show it is surjective, but that is easily seen, with

$$\Psi^{-1}\left(\left[\sum_j f_j \otimes \delta_{[\psi_j]}\right]\right) = \sum_j f_j \psi_j \psi_j^* \tag{3.33}$$

for any $\sum_j f_j \otimes \delta_{[\psi_j]} \in \mathcal{A}(\mathbb{S}_{\mathsf{H}}/\sim;\mathcal{E};L^1(\mathcal{X};\mu))$. □

Comment Since $\mathbb{S}_{\mathsf{H}} \times \mathbb{S}_{\mathsf{J}} \ncong \mathbb{S}_{\mathsf{H}\otimes\mathsf{J}}$ if neither H nor J are trivial, this construction is necessarily contextual. Therefore, it does not violate Bells' inequality (see §6.3 below).

The nonexistence of the corresponding construction using a base measure

The corresponding construction using a base measure ν would be for there to be a positive isometry from $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$ to a quotient space of some $L^1(\mathcal{Y};\nu)$. This is impossible, as the following theorem shows:

Theorem 3.5.2 If the Hilbert space H is non-trivial, there are no: **(i)** set $\mathcal{Y}$; **(ii)** σ-finite measure ν; and **(iii)** equivalence relation $\sim$ induced by a closed, linear subspace $\mathsf{B} \subset L^1(\mathcal{Y};\nu)$– such that there is a positive, linear isomorphism, $\Psi : L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})) \to L^1(\mathcal{Y};\nu)/\sim$, which is also an isometry on the positive cone.

Proof Suppose otherwise. Then there is a $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$-valued vector measure τ on $\mathcal{Y}$ provided by $\tau(A) = \Psi^{-1}([1_A])$ for any ν-measurable subset $A \subset \mathcal{Y}$. The spaces $L^1(\mathcal{Y};\nu)^* \cong L^\infty(\mathcal{Y};\nu)$ by Riesz's theorem [62]. The dual to $L^1(\mathcal{Y};\nu)/\sim$ is provided by the *annihilator* $\mathsf{B}^\perp$: the closed, linear subspace of $L^\infty(\mathcal{Y};\nu)$ that annihilates B. (Note, in particular that since Ψ is an isometry on the positive cone, the constant function $1_\mathcal{Y} \in \mathsf{B}^\perp$.) Therefore, there is an adjoint map $\Psi^* : \mathsf{B}^\perp \to L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))^*$ given by

$$\int_\mathcal{Y} f\,\Psi\rho\,d\nu = (\Psi^* f)\,\rho \tag{3.34}$$

for any $f \in \mathsf{B}^\perp$ and $\rho \in L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$. By the basic properties of Banach spaces, $(\Psi^*)^{-1} = (\Psi^{-1})^*$, Ψ^* is positive, and Ψ^* is an isometry on the positive cone.

Then,

$$\int_A \Psi^{*-1}\Phi\,d\nu = \Phi(\tau(A)) \tag{3.35}$$

for any linear functional $\Phi \in L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))^*$ and ν-measurable subset $A \subset \mathcal{Y}$. For I_H the identity operator, $\Psi^{*-1}(I_\mathsf{H} 1_\mathcal{X})$ is the element of $\mathsf{B}^\perp$ that agrees with the norm when integrated with any positive function in $L^1(\mathcal{Y};\nu)/\sim$; hence, it must be $1_\mathcal{Y}$. Therefore,

$$\int_\mathcal{X} \mathrm{tr}\ \tau(A)\, d\mu = \int_A \Psi^{*-1}(I_\mathsf{H} 1_\mathcal{X})\, d\nu = \nu(A) \tag{3.36}$$

However, this gives rise to a contradiction. Fix some subset $B \subset \mathcal{X}$ with $0 < \mu(B) < \infty$. Take any unit norm $\psi \in \mathsf{H}$. Since Ψ is positive, by definition, the equivalence class $\Psi(\psi\psi^* 1_B)$ has a positive member, call it g_ψ. There must be some $\psi \neq \xi$ such that $A = \{g_\psi > 0\} \cap \{g_\xi > 0\}$ has strictly positive ν-measure; otherwise, there would be an uncountable collection $\{g_\psi > 0\}$ indexed by unit norm $\psi \in \mathsf{H}$ of subsets of $\mathcal{Y}$, each with strictly positive ν measure, but whose pairwise intersections all have ν-measure zero. The existence of such a collection would contradict ν being σ-finite by **B1.6**. Since Ψ is an isometry on the positive cone,

$$\int_\mathcal{Y} g_\psi\, d\nu = \mu(B)\,\mathrm{tr}\ I\psi\psi^* = \mu(B) = \mu(B)\,\mathrm{tr}\ \psi\psi^*\psi\psi^* = \int_{y\in\mathcal{Y}} g_\psi(y)\, d\langle\tau\psi,\psi\rangle(y) \tag{3.37}$$

Hence, $\langle\tau\psi,\psi\rangle \leq \mathrm{tr}\ \tau = \nu$ must be equal to ν when restricted to $\{g_\psi > 0\} \supset A$. By a similar argument, $\langle\tau\xi,\xi\rangle$ must must be equal to ν when restricted to $\{g_\xi > 0\} \supset A$. These conditions are impossible to satisfy. □

The special case of two-dimensional Hilbert spaces

It is possible to circumvent the conclusion of the preceding theorem if the positive isomorphism is with a closed, linear subspace of $L^1(\mathcal{Y};\nu)/\sim$. Specialize to $\mathcal{Y} = \mathcal{X}\times\mathcal{Z}$ and $\nu = \mu\times\eta$ and let C be the closed, linear subspace of $L^1(\mathcal{X}\times\mathcal{Z};\mu\times\eta)/\sim$. Using the notation of the preceding proof, let $\sim'$ be the equivalence relation on the annihilator $\mathsf{B}^\perp \subset L^\infty(\mathcal{X}\times\mathcal{Z};\mu\times\eta)$ induced by the annihilator $\mathsf{C}^\perp \subset L^\infty(\mathcal{X}\times\mathcal{Z};\mu\times\eta)$, so $f \sim' g$ if $\int_{\mathcal{X}\times\mathcal{Z}}[h]f\, d(\mu\times\eta) = \int_{\mathcal{X}\times\mathcal{Z}}[h]g\, d(\mu\times\eta)$ for all $[h] \in L^1(\mathcal{X}\times\mathcal{Z};\mu\times\eta)/\sim$. From the proof of the preceding theorem, it is then necessary that $\Psi^{*-1}(I_\mathsf{H} 1_\mathcal{X}) = [1_{\mathcal{X}\times\mathcal{Z}}]$. A sufficient way to accomplish this would be for there to be a positive map (not necessarily linear in $\psi\psi^*$) $\Lambda : \mathbb{S}_\mathsf{H}/\sim_{\text{phase}} \to L^\infty(\mathcal{Z};\eta)$, where $\sim_{\text{phase}}$ is the equivalence relation on $\mathbb{S}_\mathsf{H}$ given above, such that

$$\Psi^{*-1}(\psi\psi^* 1_\mathcal{X}) = [1_\mathcal{X} \otimes \Lambda(\psi)] \text{ for all } \psi \in \mathbb{S}_\mathsf{H}/\sim_{\text{phase}} \tag{3.38}$$

and, for almost every $z \in \mathcal{Z}$ with respect to η, the map $\psi \to \Lambda(\psi)(z)$ is a frame function with weight one[23].

For H of dimension two, we then have the following based on a construction by Kochen and Specker [30]. Let ω be the usual measure on the sphere $\mathbb{S}^2 \subset \mathbb{R}^3$, so, with Cartesian coordinates (x_1,x_2,x_3) for $\mathbb{R}^3$,

$$\omega(A) = \int_{(x_1,x_2,x_3)\in A} = (x_1\, dx_2 \wedge dx_3 - x_2\, dx_1 \wedge dx_3 + x_3\, dx_1 \wedge dx_2) \tag{3.39}$$

[23]A function $f : \mathbb{S}_\mathsf{H}/\sim_{\text{phase}} \to \mathbb{R}$ is a *frame function with weight* w if it is zero except for a separable subspace of H and $\sum_j f(\mathbf{e}_j) = w$ for any orthonormal basis $\{\mathbf{e}_j\}$ of that subspace.

for any Borel subset $A \subset \mathbb{S}^2$. Let $\sim$ be the equivalence relation on $L^1(\mathcal{X} \times \mathbb{S}^2; \mu \times \omega)$ given by $f \sim g$ if $\int_{B\times H} f\, d(\mu \times \omega) = \int_{B\times H} g\, d(\mu \times \omega)$ for all μ-measurable subsets $B \subset \mathcal{X}$ and all hemispheres $H \subset \mathbb{S}^2$ (whether the hemispheres are taken open or closed is irrelevant). Equip the quotient space $L^1(\mathcal{X} \times \mathbb{S}^2; \mu \times \omega)/\sim$ with a norm in the usual way via $\|[f]\| = \inf_{g\in[f]} \|g\|$. Define the positive cone on $L^1(\mathcal{X} \times \mathbb{S}^2; \mu \times \omega)/\sim$ by those equivalence classes that contain a positive element of $L^1(\mathcal{X} \times \mathbb{S}^2; \mu \times \omega)$.

Theorem 3.5.3 There is a positive, linear isometry, $\Psi : L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H})) \to L^1(\mathcal{X} \times \mathbb{S}^2; \mu \times \omega)/\sim$ with an associated map $\Lambda : \mathbb{S}_\mathsf{H}/\sim_{\text{phase}} \to L^\infty(\mathbb{S}^2; \omega)$ with the above properties.

Proof Take any orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ for $\mathsf{H} \cong \mathbb{C}^2$. Let $\Psi : L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H})) \to L^1(\mathcal{X} \times \mathbb{S}^2; \mu \times \omega)/\sim$ be given by first taking Ψ on elements of the form $\psi\psi^* f$ for $\psi \in \mathbb{S}_\mathsf{H}$ and $f \in L^1(\mathcal{X}; \mu)$ to be the equivalence class of the positive function

$$f = \begin{cases} \frac{1}{\pi}\mathbf{y}\mathbf{x} & \text{if } \mathbf{y}\mathbf{x} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.40}$$

$$\text{for } \mathbf{y} = \begin{bmatrix} |\psi_1|^2 - |\psi_2|^2 & \Re\, 2\psi_1\overline{\psi_2} & \Im\, 2\psi_1\overline{\psi_2} \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

where $\psi_j = \langle \psi, \mathbf{e}_j \rangle$. Since H is two-dimensional, $\mathcal{D}(\mathsf{H})$ is four-dimensional, with basis $\{\mathbf{e}_1\mathbf{e}_1^*, \mathbf{e}_2\mathbf{e}_2^*, (\mathbf{e}_1+\mathbf{e}_2)(\mathbf{e}_1+\mathbf{e}_2)^*, (\mathbf{e}_1+\imath\mathbf{e}_2)(\mathbf{e}_1+\imath\mathbf{e}_2)^*\}$. Therefore, Ψ can be extended by linearity to all of $L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H}))$.

For ρ a simple function, by using the spectral theorem for the each of the finite number of values ρ attains, the map Ψ is readily seen to be positive and an isometry. Since simple functions are dense in $L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H}))$, by **A1.2**, Ψ is a positive isometry for all of $L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H}))$.

For this Ψ, an associated Λ does exist. It is given by first defining

$$\mathbf{z}^{\mathrm{T}} = \begin{bmatrix} |\xi_1|^2 - |\xi_2|^2 & \Re 2\xi_1\overline{\xi_2} & \Im 2\xi_1\overline{\xi_2} \end{bmatrix} \Leftrightarrow \xi \propto \sqrt{\frac{1+z_1}{2}}\mathbf{e}_1 + \frac{z_2 + \imath z_3}{\sqrt{2(1+z_1)}}\mathbf{e}_2 \tag{3.41}$$

where the proportionality for ξ is up to an irrelevant phase. Then $\Lambda(\xi) = 1_{H_z}$, where H_z is the hemisphere centered at z. This has the required properties since **(i)** for any orthonormal $\{\xi, \zeta\}$, $\Lambda(\xi) + \Lambda(\zeta) = 1_{H_z} + 1_{H_{-z}} = 1_{\mathbb{S}^2}$, with equality in the $L^\infty(\mathbb{S}^2; \omega)$-sense of almost everywhere with respect to ω and **(ii)** for any $\xi, \psi \in \mathbb{S}_\mathsf{H}$ and $f \in L^1(\mathcal{X}; \mu)$,

$$\int_{\mathcal{X}\times\mathbb{S}^2} (1_\mathcal{X} \otimes \Lambda(\xi))\, \Psi(\psi\psi^* f)\, d(\mu\times\omega) = \frac{1}{2}(1 + \mathbf{y}\mathbf{z})\left(\int_\mathcal{X} f\, d\mu\right) = |\langle \xi, \psi \rangle|^2 \left(\int_\mathcal{X} f\, d\mu\right) \qquad \square$$

However, for Hilbert space H of dimension greater than two, there is no construction of this form because Gleason proved [24] that all frame functions on Hilbert spaces of dimension greater than two are regular[24].

Theorem 3.5.4 There are no: **(i)** set $\mathcal{Z}$; **(ii)** σ-finite measure η; and **(iii)** equivalence relation $\sim$ induced by a closed, linear subspace $\mathsf{B} \subset L^1(\mathcal{X} \times \mathcal{Z}; \mu \times \eta)$–such that there is a positive, linear isometry, $\Psi : L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H})) \to L^1(\mathcal{X} \times \mathcal{Z}; \mu \times \eta)/\sim$, which has an associated map $\Lambda : \mathbb{S}_{\mathsf{H}}/\sim_{\text{phase}} \to L^\infty(\mathcal{Z}; \eta)$ with the properties given above.

Proof Suppose otherwise. By Gleason's result, there is a $T \in L^1(\mathcal{Z}; \eta; \mathcal{D}(\mathsf{H})^+)$ such that $T(z)$ has trace one for almost every $z \in \mathcal{Z}$ with respect to η and

$$\int_{\mathcal{X}\times\mathcal{Z}} (1_{\mathcal{X}} \otimes \langle T\xi, \xi\rangle)\, \Psi(\psi\psi^* 1_B)\, d(\mu \times \eta) = |\langle \xi, \psi\rangle|^2 \mu(B) \tag{3.42}$$

for all $\psi, \xi \in \mathbb{S}_{\mathsf{H}}/\sim_{\text{phase}}$ and μ-measurable $B \subset \mathcal{X}$. However, then following the argument of theorem **3.5.2**, there is a contradiction. Fix some subset $B \subset \mathcal{X}$ with $0 < \mu(B) < \infty$. Take any unit norm $\psi \in \mathsf{H}$. Since Ψ is positive, by definition, the equivalence class $\Psi(\psi\psi^* 1_B)$ has a positive member, call it g_ψ. There must be some $\psi \neq \xi$ such that $A = \{g_\psi > 0\} \cap \{g_\xi > 0\}$ has strictly positive ν-measure. Since Ψ is an isometry on the positive cone,

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{Z}} g_\psi\, d(\mu \times \eta) &= \mu(B)\, \mathrm{tr}\ I\psi\psi^* = \mu(B) = \mu(B)\, \mathrm{tr}\ \psi\psi^*\psi\psi^* \\ &= \int_{\mathcal{X}\times\mathcal{Z}} (1_{\mathcal{X}} \otimes \langle T\psi, \psi\rangle)\, g_\psi\, d(\mu \times \eta)\end{aligned} \tag{3.43}$$

Hence, $\langle T\psi, \psi\rangle \le \mathrm{tr}\ T = 1$ must be equal to 1 almost everywhere with respect to η when restricted to $\{g_\psi > 0\} \supset A$. By a similar argument, $\langle T\xi, \xi\rangle \le \mathrm{tr}\ T = 1$ must be equal to 1 almost everywhere with respect to η when restricted to $\{g_\xi > 0\} \supset A$. These conditions are impossible to satisfy. □

A construction for option II

For option **II**, there is also a way to embed quantum behavior into a purely classical, but contextual, model. For any Hilbert space H, let $\mathbb{B}_{\mathsf{H}}$ be the closed unit ball within H. Equip $\mathbb{B}_{\mathsf{H}}$ with the weak topology; denote the resulting space by $\mathbb{B}_{\mathsf{H}}^{\text{weak}}$. Since Hilbert spaces are reflexive, with $\mathsf{H}^* \cong \mathsf{H}$ by Riesz's theorem [63], then this is the same as the weak* topology, so by Alaoglu's theorem [64], $\mathbb{B}_{\mathsf{H}}^{\text{weak}}$ is a compact space. Furthermore, it is Hausdorff since, given any distinct points b_0, b_1, there are separating weak neighborhoods $\mathcal{N}\left(b_0; b_1 - b_0; \frac{1}{2}\|b_0 - b_1\|^2\right)$ and $\mathcal{N}\left(b_1; b_1 - b_0; \frac{1}{2}\|b_0 - b_1\|^2\right)$.

Then, for any compact, Hausdorff space $\mathcal{X}$, we have the following theorem:

Theorem 3.5.5 There is a positive, isometric isomorphism between $\mathcal{C}(\mathcal{X}; \mathcal{K}(\mathsf{H}))$ and a closed, linear subspace of $\mathcal{C}\left(\mathcal{X} \times \mathbb{B}_{\mathsf{H}}^{\text{weak}}\right)$.

[24] A frame function is *regular* if it is given by $\psi \to \langle T\psi, \psi\rangle$ for some trace-class, self-adjoint operator T.

Proof Consider the map $\Psi : \mathcal{C}(\mathcal{X}; \mathcal{K}(\mathsf{H})) \to \mathcal{C}\left(\mathcal{X} \times \mathbb{B}_{\mathsf{H}}^{\text{weak}}\right)$ given by

$$(\Psi\varphi)(x, b) = \langle \varphi(x) b, b \rangle \tag{3.44}$$

The map Ψ is clearly both positive and an isometry. From the latter property, its image is a closed, linear subspace. The functions in the image of the map are also clearly continuous in x; the following lemma shows that they are continuous in b, so, as claimed above, they are indeed continuous functions. □

Lemma 3.5.6 For any compact operator $\phi \in \mathcal{K}(\mathsf{H})$, the map $\Phi : \mathbb{B}_{\mathsf{H}}^{\text{weak}} \to \mathbb{R}$ given by $\Phi(b) = \langle \phi b, b \rangle$ is continuous.

Proof First take the case of ϕ positive and rank one, so $\phi = \psi\psi^*$ for some $\psi \in \mathsf{H}$. Then, given any $b_0 \in \mathbb{B}_{\mathsf{H}}$ and $\varepsilon > 0$, there is a weak neighborhood $\mathcal{N}\left(b_0; \psi; \sqrt{|\langle\psi, b_0\rangle|^2 + \varepsilon} - |\langle\psi, b_0\rangle|\right)$ such that for every b in the neighborhood, using the triangle inequality repeatedly,

$$\begin{aligned} |\Phi(b) - \Phi(b_0)| = \left||\langle\psi, b\rangle|^2 - |\langle\psi, b\rangle|^2\right| &\le (|\langle\psi, b\rangle| + |\langle\psi, b_0\rangle|)\, |\langle\psi, b\rangle - \langle\psi, b_0\rangle| \\ &< (|\langle\psi, b\rangle| + |\langle\psi, b_0\rangle|) \left(\sqrt{|\langle\psi, b_0\rangle|^2 + \varepsilon} - |\langle\psi, b_0\rangle|\right) \\ &< \left(2\,|\langle\psi, b_0\rangle| + \sqrt{|\langle\psi, b_0\rangle|^2 + \varepsilon} - |\langle\psi, b_0\rangle|\right)\left(\sqrt{|\langle\psi, b_0\rangle|^2 + \varepsilon} - |\langle\psi, b_0\rangle|\right) \end{aligned} \tag{3.45}$$

which is equal to ε.

Since finite rank operators are norm-dense in $\mathcal{K}(\mathsf{H})$, for any $\phi \in \mathcal{K}(\mathsf{H})$ and $\varepsilon > 0$ there is are finite collections $\{\psi_j\} \subset \mathbb{S}_{\mathsf{H}}$ (for $\mathbb{S}_{\mathsf{H}}$ the unit sphere in H) and $\{a_j\} \subset \mathbb{R}$ such that $\left\|\phi - \sum_{j=1}^{n} a_j \psi_j \psi_j^*\right\|_{\text{op}} \le \frac{1}{2}\varepsilon$. Then, by the triangle equality and the above result, for every b in the neighborhood

$$\bigcap_{j=1}^{n} \mathcal{N}\left(b_0; \psi_j; \sqrt{|a_j| |\langle\psi_j, b_0\rangle|^2 + \frac{1}{2n}\varepsilon} - \sqrt{|a_j|}|\langle\psi, b_0\rangle|\right) \tag{3.46}$$

we have $|\Phi(b) - \Phi(b_0)| < \varepsilon$. □

Comment Since $\mathbb{B}_{\mathsf{H}} \times \mathbb{B}_{\mathsf{J}} \ncong \mathbb{B}_{\mathsf{H}\otimes\mathsf{J}}$ if neither H nor J are trivial, this construction is necessarily contextual. Therefore, it does not violate Bells' inequality (see §6.3 below). Also, for H finite-dimensional, it is possible to use $\mathbb{S}_{\mathsf{H}}$ instead of $\mathbb{B}_{\mathsf{H}}$ and the norm topology (which is equivalent to the weak topology in finite dimensions). In this case, it is also possible to further reduce the space by the equivalence relation on $\mathbb{S}_{\mathsf{H}}$ used in the preceding construction for option **I**.

3.6 Embedding classical models into quantum ones

For σ-algebras generated by a countable collection of atoms, using option **I'** it is always possible to duplicate classical behavior using quantum nodes; one simply embeds $L^1(\mathcal{X}; \mu)$

as the diagonal operators in $\mathcal{D}(L^2(\mathcal{X};\mu))^+$. Similarly, using option **II**, if $\mathcal{X}$ is a finite set of points with the discrete topology, $\mathcal{C}(\mathcal{X})$ can be embedded as the diagonal operators in $\mathcal{K}(L^2(\mathcal{X};\text{counting measure}))^+$. However, for continuous sets with finer σ-algebras, these are not options since there are no diagonal operators. One alternative is the map η_*^{-1} into equivalence classes of trace-class operators with common diagonals defined in **B5.19** and used in §5.1 below.

3.7 Classical physics

Classical mechanics

Another approach, which is that of classical mechanics, is to not completely duplicate quantum behavior, but instead approximate it. The key is to recognize that $L^1(T^*\mathcal{X};\Omega)^+$ intersects the Hilbert space $L^2(T^*\mathcal{X};\Omega)$. Similarly, $\mathcal{D}(L^2(\mathcal{X};\mu))^+$ intersects the Hilbert space[25] $\mathcal{S}_2(L^2(\mathcal{X};\mu))$, which is isomorphic to $L^2(\mathcal{X}\times\mathcal{X};\mu\times\mu)$. Therefore, if we have an isomorphism Ψ between $L^2(T^*\mathcal{X};\Omega)$ and $L^2(\mathcal{X}\times\mathcal{X};\mu\times\mu)$, we can associate some of the elements of $L^1(T^*\mathcal{X};\Omega)^+$ with those in $\mathcal{D}(L^2(\mathcal{X};\mu))^+$.

For $\mathcal{X} = \mathbb{R}^n$ and μ mutually absolutely continuous with respect to Lebesgue measure λ, this is indeed possible, with $f(q,p) = (\Psi\rho)(q,p)$ given by

$$\frac{1}{(\pi\hbar)^n}\int_{u\in\mathbb{R}^n}\exp\left(-\frac{2\imath p\bullet u}{\hbar}\right)\rho(q+u,q-u)\sqrt{\frac{d\mu}{d\lambda}}(q+u)\sqrt{\frac{d\mu}{d\lambda}}(q+u)\,d\lambda(u) \tag{3.47}$$

and inverse $\rho(x,x') = (\Psi^{-1}f)(x,x') = (\pi\hbar)^n(\Psi^*f)(x,x')$ given by

$$\sqrt{\frac{d\lambda}{d\mu}}(x)\sqrt{\frac{d\lambda}{d\mu}}(x')\int_{p\in\mathbb{R}^n}\exp\left(\frac{\imath p\bullet(x-x')}{\hbar}\right)f\left(\frac{x+x'}{2},p\right)d\lambda(p) \tag{3.48}$$

Intuitively, $f\in L^1(T^*\mathcal{X};\Omega)^+$ will be paired with some $\rho\in\mathcal{D}(L^2(\mathcal{X};\mu))^+$ if ρ is very close to being diagonal and varies slowly along the diagonal whereas f obeys a local version of Heisenberg's uncertainty relation, with sharp features in q being spread out in p and vice versa. Also, these pairs exist for the ground state of the harmonic oscillators, despite ρ not being very close to being diagonal. Note that when these pairs exist, as a map from $\mathcal{D}(L^2(\mathcal{X};\mu))^+$ to $L^1(T^*\mathcal{X};\Omega)^+$, Ψ is actually an isometry.

Two observations arise from this. Firstly, if we believe that a quantum description gives better predictions and that a formulation in terms of classical physics is merely an approximation, then the classical predictions are completely untrustworthy for non-integrable classical systems where f rapidly develops fine tendrils under time evolution. Evolving the quantum analogue of the system and then using Ψ would give very different behavior. Either the fine tendrils would not form or the correlation would break down, with $f = \Psi\rho$ no longer

[25] $\mathcal{S}_2$ denotes the Hilbert-Schmidt operators.

being positive. Secondly, classical physics is simpler than quantum mechanics only because the time evolution for a Hamiltonian system is so simple, being merely a point transformation on a finite dimensional space, $T^*\mathcal{X} \to T^*\mathcal{X}$. Working with full stochastic generality using non-Hamiltonian systems in classical physics, the analogous quantum description is no more complicated.

The preceding is evidently dependent on the existence of Lebesgue measure. Since there is no infinite-dimensional analogue of Lebesgue measure [48], it cannot be extended directly to field theories. Instead, some measure which itself depends on $\hbar$ must be utilized.

Distinguishable versus indistinguishable particles

It is usually said that there are no indistinguishable particles classically because it is possible to follow trajectories, which are the time-evolution point transformations $T^*\mathcal{X} \to T^*\mathcal{X}$. The above analysis gives a different interpretation that is valid in the more general case where there are no trajectories: if indistinguishability is important in a model, so the density matrix ρ is taken symmetric (or anti-symmetric) under interchange,

$$\rho(x_1, x_2, x_1', x_2') = \pm\rho(x_2, x_1, x_1', x_2') = \pm\rho(x_1, x_2, x_2', x_1') = \rho(x_2, x_1, x_2', x_1') \tag{3.49}$$

then it is necessarily far from being diagonal, so there is likely no partner in $L^1(T^*\mathcal{X}; \Omega)^+$ to ρ and a classical model is invalid.

A similar situation holds for the case of σ-algebras generated by a countable collection of atoms. If there is classical indistinguishability (such as dollars in a bank account), the measures assigned to certain atoms by any allowed measure must be the same. No issues are raised by embedding the measure as the diagonal in $\mathcal{D}(L^2(\mathcal{X}; \mu))^+$. However, trying to recast this as quantum indistinguishability then creates off-diagonal entries without classical interpretation.

3.8 Additional structures for Bayesian networks?

It may be possible that additional mathematical structures besides measures and density matrices could be used as the base and target spaces of linear maps to calculate probabilities using Bayesian networks. To avoid topological complications, only the finite-dimensional case will be sought. Looking through the properties of measures and density matrices that are essential, we have the following properties required for an additional mathematical structure, given by a collection $\mathcal{A}$ of finite-dimensional, normed vector spaces: *(i)*for any vector spaces $\mathsf{A}, \mathsf{B} \in \mathcal{A}$, the vector space $\mathsf{A} \otimes \mathsf{B}$ is also in $\mathcal{A}$; *(ii)* each $\mathsf{B} \in \mathcal{A}$ has a positive cone, B^+ that is generating[26]; *(iii)* each $\mathsf{B} \in \mathcal{A}$ has the quasi-AL-property that $\|\mathbf{x}\| + \|\mathbf{y}\| = \|\mathbf{x} + \mathbf{y}\|$ for any $\mathbf{x}, \mathbf{y} \in \mathsf{B}^+$; and *(iv)* for any $\mathbf{x} \in \mathsf{A}^+$, $\mathbf{y} \in \mathsf{B}^+$, $\mathbf{x} \otimes \mathbf{y}$ is in $(\mathsf{A} \otimes \mathsf{B})^+$ with $\|\mathbf{x} \otimes \mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\|$. In addition, *(v)* it is required that there be some nontrivial (other than the identity) maps

[26] Any $\mathbf{x}$ in each $\mathsf{B} \in \mathcal{A}$ can be written as $\mathbf{y} - \mathbf{z}$ for some $\mathbf{y}, \mathbf{z} \in \mathsf{B}$.

between these vector spaces that are simultaneously both completely positive and completely bounded.

Evidently, for measures $\mathcal{A}$ is a collection of n-dimensional vector spaces for n a natural number, $\{1, 2, 3, \ldots\}$. For density matrices it is a collection of n-dimensional vector spaces for n a square number, $\{1, 4, 9, \ldots\}$. The question, which is open, is whether there are a collection of vector spaces $\mathcal{A}$ satisfying the preceding properties other than spaces isomorphic to measures, density matrices, or various tensor products of measures and density matrices. Obviously, this question is of vital interest–if the answer is negative, then we are limited to linear maps on measures, density matrices, and their tensor products in formulating Bayesian networks; if the answer is affirmative, then there is the immediate additional question of whether or not the additional structure is of practical utility in making predictions.

Chapter 4

Constructing the networks

4.1 Graphical constructs and rules

Recall the graphical model simply gives the form for the needed information and illustrates constraints. The actual needed data is given in an accompanying table. In the following, first the rules for the graphical model are given, then what accompanying data needs to be supplied depending on which option, **I'** or **II**, is chosen.

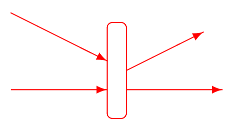
Nodes and terminators

Nodes that are observable, with the traditional emphasis on the transition probability function, will be indicated in black. Hidden classical nodes, where the linear map on measures is taken to be fundamental, will be red. Quantum nodes, whose linear maps involve non-trivial spaces of density matrices, will be blue. Arrows share the color of their parent. Observable nodes that are being incorporated into the calculation of probability will be crossed. Those that are being marginalized will be left open. Those that are being conditioned on will be filled. Hidden (classical or quantum) nodes will always be left open. To make it visually clear whether or not a graph is complete and represents the calculation of a number (the probability) or is just a graph fragment with unresolved maps, terminal nodes which are either maps from the trivial space (giving initialization) or to the trivial space (giving the evaluation map) will be indicated by half-circle-half-box shapes, with the half-box being the terminating end.

(4.1)

Splitters

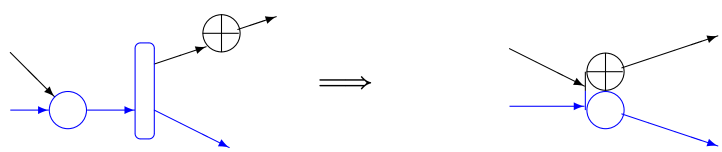
As has been pointed out (see §2.4), the usual Bayesian network construction depends heavily on the ability to copy, with each child node receiving a copy. Without this ability (see §3.2 and §3.4), an additional graphical structure is required, the *splitter*. This is an oval with one or more arrows coming in and more than one arrow coming out–if only one arrow comes out, the splitter can be dispensed with and all the incoming arrows can be redirected to the following node.



(4.2)

Pince-nez

There is a certain construction that occurs so often for joining hidden nodes to observable ones that it is replaced by its own structure, the *pince-nez*. As the name suggests, this is two circles, one an observable node, the other a hidden one, joined by a bar.



(4.3)

The neutral term *pince-nez* is employed since the concept of "measurement"[1] applied to hidden nodes is so fraught with metaphysical connotations.

4.2 Data for the structures using Option I'

Using **option I'**, for each hidden node, the most general data is

$$(\mathcal{I}, \mu, \mathsf{H}; \mathcal{O}, \nu, \mathsf{J}; L)$$

where $\mathcal{I}$ is the input set, μ the accompanying base measure, H the input Hilbert space, $\mathcal{O}$ the output set, ν the accompanying base measure, J the output Hilbert space, and L a completely-positive, norm-preserving (on the positive cone) map in $\mathcal{B}(L^1(\mathcal{I};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{O};\nu;\mathcal{D}(\mathsf{J})))$. If either Hilbert spaces are the trivial space $\mathbb{C}$, it can be left off the list. Similarly, if either base measure is trivial, for instance, if $\mu(\varnothing) = 0$, $\mu(\mathcal{I}) = 1$ with no other sets measurable, then it and its set can be left out as well.

[1]We will always place "measurement" in quotation marks to emphasize its problematic nature.

For each splitter, the most general data is

$$(\mathcal{I}, \mu, \mathsf{H}; \mathcal{O}_1, \nu_1, \mathsf{J}_1, \ldots, \mathcal{O}_m, \nu_m, \mathsf{J}_m; \Psi)$$

where $\mathcal{I} \cong \mathcal{O}_1 \times \cdots \times \mathcal{O}_m$, $\mu \cong \nu_1 \times \cdots \times \nu_m$, and $\mathsf{H} \cong \mathsf{J}_1 \otimes \cdots \otimes \mathsf{J}_m$, with Ψ giving the isomorphisms, which should be trivial except for possible permutations to make sure everything goes to the correct place, with more complicated behavior placed in separate nodes. Once again, trivial entries can be left out.

For each pince-nez, the most general data is

$$(\mathcal{I}, \mu, \mathsf{H}; \mathcal{R}, \tau; \mathcal{O}, \nu, \mathsf{J}; L)$$

where $\mathcal{I}$ is the input set, μ the accompanying base measure, H the input Hilbert space, $\mathcal{R}$ the observable set, τ the accompanying base measure, $\mathcal{O}$ the output set, ν the accompanying base measure, J the output Hilbert space, and L a completely-positive, norm-preserving (on the positive cone) map in $\mathcal{B}(L^1(\mathcal{I}; \mu; \mathcal{D}(\mathsf{H})), L^1(\mathcal{R} \times \mathcal{O}; \tau \times \nu; \mathcal{D}(\mathsf{J})))$. Again, trivial entries can be left out (however, if τ is trivial, then the pince-nez is actually just a node).

4.3 Data for the structures using Option II

Using **option II**, for each hidden node, the most general data is

$$(\mathcal{I}, \mathsf{H}; \mathcal{O}, \mathsf{J}; L)$$

where $\mathcal{I}$ is the input compact space, H the input Hilbert space, $\mathcal{O}$ the output compact space, J the output Hilbert space, and L a completely-positive, norm-preserving (on the positive cone) map in $\mathcal{B}(\mathcal{C}(\mathcal{O}; \mathcal{K}(\mathsf{J})), \mathcal{C}(\mathcal{I}; \mathcal{K}(\mathsf{H})))$. As previously noted, the spaces can be locally-compact if L is further restricted. Trivial entries can be left off the list.

For each splitter, the most general data is

$$(\mathcal{I}, \mathsf{H}; \mathcal{O}_1, \mathsf{J}_1, \ldots, \mathcal{O}_m, \mathsf{J}_m; \Psi)$$

where $\mathcal{I} \cong \mathcal{O}_1 \times \cdots \times \mathcal{O}_m$ with the product topologies agreeing as well, and $\mathsf{H} \cong \mathsf{J}_1 \otimes \cdots \otimes \mathsf{J}_m$, with Ψ giving the isomorphisms, which should be trivial except for possible permutations to make sure everything goes to the correct place, with more complicated behavior placed in separate nodes. Once again, trivial entries can be left out.

For each pince-nez, the most general data is

$$(\mathcal{I}, \mathsf{H}; \mathcal{R}; \mathcal{O}, \mathsf{J}; L)$$

where $\mathcal{I}$ is the input compact space, H the input Hilbert space, $\mathcal{R}$ the observable compact space, $\mathcal{O}$ the output compact space, J the output Hilbert space, and L a completely-positive, norm-preserving (on the positive cone) map in $\mathcal{B}(\mathcal{C}(\mathcal{R} \times \mathcal{O}; \mathcal{K}(\mathsf{J})), \mathcal{C}(\mathcal{I}; \mathcal{K}(\mathsf{H})))$. As previously noted (see §3.2), the spaces can be locally-compact if L is further restricted. Again, trivial entries can be left out (however, if $\mathcal{R}$ is trivial in the sense of being the one-point set, then the pince-nez is actually just a node).

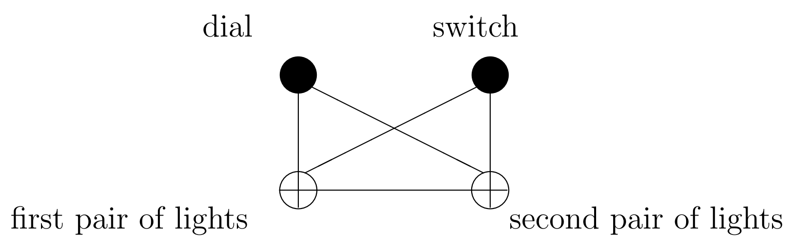
4.4 Example-double-slit experiment

So far, the exposition has been quite abstract. To make things more concrete and demonstrate how to use Bayesian networks with linear maps (both classical and quantum) in practice, consider the following example. There is a black-box with a dial which can be set to any position θ on a circle, a switch with two settings, and four lights, with lights one and two forming one pair and lights three and four forming another. After study, it is determined to have the following properties: if the switch is off, periodically one of the second pair of lights flashes while the first pair never flashes, whereas, if the switch is on, one of each pair flashes. After more study, it is determined the behavior of each round is independent and, if the switch is off, with probability $\cos^2\theta$ the third light alone flashes and with probability $\sin^2\theta$ the fourth light alone flashes. If the switch is on, the joint probabilities are as follows:

first on, third on, others off	$\frac{1}{4}$
first on, fourth on, others off	$\frac{1}{4}$
second on, third on, others off	$\frac{1}{4}$
second on, fourth on, others off	$\frac{1}{4}$

(4.4)

with the other probabilities being zero. Presenting the information in the manner of the preceding joint probabilities corresponds to the following graphical model:

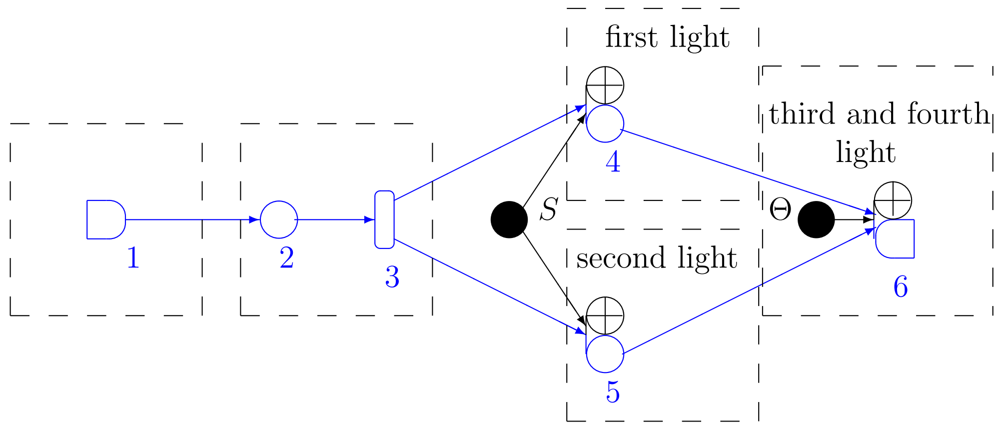


(4.5)

Suppose the box is now opened and is found to be composed of five modules. A cable runs from the first module to the second. Two cables come from the second module, one each running to a pair of seemingly identical modules, each of which has one of the first pair of lights. The switch activates switches on the pair in unison. These modules have cables running to the last, which has the dial as a control. If possible, given the joint probabilities, we should confine ourselves to models which are consistent with the constraints implied by these observations because we would like to be able to predict what would happen if these modules were rewired or taken out and placed in a different context.

4.5 Quantum model

Consider first a quantum model with graphical model:



(4.6)

The blue 1,2,3,4,5,6 are for reference. The information presented in the form indicated by the graphical model is as follows:

node 1 $(;\mathbb{C}^2;L_1)$, where L_1 is the constant map with value $\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$.

node 2 $(\mathbb{C}^2;\mathbb{C}^4;L_2)$, where $L_2(\rho)$ is

$$\begin{bmatrix} 0 & 0 \quad 0 & 0 \\ 0 & & 0 \\ 0 & \rho & 0 \\ 0 & 0 \quad 0 & 0 \end{bmatrix} \tag{4.7}$$

splitter 3 $(\mathbb{C}^4;\mathbb{C}^2,\mathbb{C}^2)$

pince-nez 4

$$\left(\left\{ \begin{array}{l} \text{switch on,} \\ \text{switch off} \end{array} \right\}, \begin{array}{l} \text{counting} \\ \text{measure} \end{array}, \mathbb{C}^2; \left\{ \begin{array}{l} \text{light one on and switch on,} \\ \text{light one on and switch off,} \\ \text{light one off and switch on,} \\ \text{light one off and switch off,} \end{array} \right\}, \begin{array}{l} \text{counting} \\ \text{measure} \end{array}; \mathbb{C}^2; L_4 \right) \tag{4.8}$$

where $L_4(\tau)$ is

$$\begin{cases} \tau|_{\text{switch off}} & \text{if light one off and switch off} \\ 0 & \text{if light one on and switch off} \\ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \tau|_{\text{switch on}} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} & \text{if first on and switch on} \\ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tau|_{\text{switch on}} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} & \text{if first off and switch on} \end{cases} \tag{4.9}$$

pince-nez 5 Same as for pince-nez 4 except the second light replaces the first light.

pince-nez 6

$$\left([0, 2\pi), \begin{matrix} \text{Lebesgue} \\ \text{measure} \end{matrix}, \mathbb{C}^4; \left\{ \begin{matrix} \text{third and fourth on,} \\ \text{third on,} \\ \text{fourth on,} \\ \text{third and fourth off} \end{matrix} \right\}, \begin{matrix} \text{counting} \\ \text{measure} \end{matrix};; L_6 \right)$$

where $L_6(\tau)$ is

$$\int_0^{2\pi} \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix} \tau(\theta) \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} d\theta \text{ if both third and fourth on} \tag{4.10}$$

$$\int_0^{2\pi} \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} e^{\imath\theta} & \frac{1}{\sqrt{2}} e^{-\imath\theta} & 0 \end{bmatrix} \tau(\theta) \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} e^{-\imath\theta} \\ \frac{1}{\sqrt{2}} e^{\imath\theta} \\ 0 \end{bmatrix} d\theta \text{ if third on}$$

$$\int_0^{2\pi} \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} e^{\imath\theta} & -\frac{1}{\sqrt{2}} e^{-\imath\theta} & 0 \end{bmatrix} \tau(\theta) \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} e^{-\imath\theta} \\ -\frac{1}{\sqrt{2}} e^{\imath\theta} \\ 0 \end{bmatrix} d\theta \text{ if fourth on}$$

$$\int_0^{2\pi} \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \tau(\theta) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} d\theta \text{ if both third and fourth off}$$

Using this information, the joint probability, given that the switch is off and the dial is set to θ, for both the first pair being off, the third light being on, and the fourth light being off, is then given by, for positive, unit-norm $g \in L^1([0, 2\pi), \text{Lebesgue})$,

$$L_6 \left((L_4 \otimes L_5)\, (L_2\, (L_1)\, 1_{\text{switch off}}) \big|_{\substack{\text{first pair is off,} \\ \text{switch is off}}} \; g \right) \Bigg|_{\substack{\text{third is on,} \\ \text{fourth is off}}} \tag{4.11}$$

$$= \int_0^{2\pi} \left| \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} e^{\imath\theta'} & \frac{1}{\sqrt{2}} e^{-\imath\theta'} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \right|^2 g(\theta')\, d\theta' = \int_0^{2\pi} \cos^2\theta'\, g(\theta')\, d\theta' \tag{4.12}$$

For g sufficiently peaked about θ, the result is approximately $\cos^2\theta$.

Similarly, the joint probability, given that the switch is on and the dial is set to θ, for both the first and the third light being on, with the second and fourth being off, is then given by

$$L_6 \left((L_4 \otimes L_5)\, (L_2\, (L_1)\, 1_{\text{switch on}}) \big|_{\substack{\text{first is on,} \\ \text{second is off} \\ \text{switch is on}}} \; g \right) \Bigg|_{\substack{\text{third is on,} \\ \text{fourth is off}}} \tag{4.13}$$

$$= \int_0^{2\pi} \left| \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} e^{\imath\theta'} & \frac{1}{\sqrt{2}} e^{-\imath\theta'} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix} \right|^2 g(\theta')\, d\theta' = \frac{1}{4} \tag{4.14}$$

The other joint probabilities can be calculated similarly and agree with the joint probabilities originally determined for the black box.

Comments on the use of option I'

Note the use of option **I'** leads to extra complexity concerning the dial position Θ, forcing the introduction of a function g peaked at θ, although in this example we are not concerned about modelling the uncertainty of that value. In this case, using option **II** would have been slightly simpler. Then, instead of using counting measure as the base measure on the discrete sets, the discrete topology would be used to make the discrete sets into compact spaces.

4.6 The Bayesian network approach versus the standard textbook approach

Interpreting the switch as controlling the operation of a position "measurement" at two slits and the dial as selecting a point on a backing screen for another position "measurement", there is a well-known standard textbook approach using wavefunctions, projectors, and Bohm's postulate that replicates the outcome. By comparison, the Bayesian network for this particular problem may appear cumbersome; however, that is largely a result of familiarity with the firmer. With the network approach, there are none of the seemingly ad hoc rules for dealing with quantum systems; instead, only the simple conditions of positivity, linearity and potential universality. Also, the graphical model is highly intuitive and guides the writing of the correct expression (4.13). This is of great importance in dealing with more complicated systems.

In addition, as will be illustrated by more complicated examples to follow (see §7.4), the network approach is far more flexible. Of particular interest, in some special cases it allows the needed Hilbert spaces to be kept to reasonable sizes in the course of the calculation instead of ballooning exponentially. To be more explicit, for a Hilbert space of dimension n, the space of operators has dimension of order n^2, and the space of maps on these operators has dimension of order n^4, which is why it appears cumbersome. However, if, by utilizing the flexibility of the network approach, one is able to avoid dealing with Hilbert spaces of dimension n^N and operator spaces of dimension n^{2N}, where N is the number of particles, the potential savings is tremendous (as will be seen in §7.4).

4.7 Classical hidden model

For the same black box, now consider a classical hidden model with the same graphical model (4.6) (except now the hidden nodes and connecting arrows will be red). For simplicity, functions on discrete spaces indexed by numbers will be given as column vectors. The information presented in the form indicated by the graphical model is as follows:

node 1 $\left(; \{1,2\}, \begin{array}{c}\text{counting}\\ \text{measure}\end{array} ; L_1 \right)$, where L_1 is the constant map with value $\begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$.

node 2 $\left(\{1,2\}, \begin{array}{c}\text{counting}\\ \text{measure}\end{array} ; \{(1,1),(1,2),\ldots,(4,4)\}, \begin{array}{c}\text{counting}\\ \text{measure}\end{array} ; L_2 \right)$, where $L_2(\rho)$ is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \rho_1 & 0 & 0 & 0 & 0 & 0 & \rho_2 & 0 & 0 & 0 \end{bmatrix}^{\mathrm{T}} \tag{4.15}$$

splitter 3 $(\{(1,1),(1,2),\ldots,(4,4)\};\{1,2,3,4\},\{1,2,3,4\})$ with the obvious identification.

pince-nez 4

$$\left(\left\{\begin{array}{l}\text{switch on,}\\ \text{switch off}\end{array}\right\}\times\{1,2,3,4\},\ \begin{array}{l}\text{counting}\\ \text{measure}\end{array};\right. \tag{4.16}$$
$$\left.\left\{\begin{array}{l}\text{light one on and switch on,}\\ \text{light one on and switch off,}\\ \text{light one off and switch on,}\\ \text{light one off and switch off,}\end{array}\right\},\ \begin{array}{l}\text{counting}\\ \text{measure}\end{array};\{1,2,3,4\},\ \begin{array}{l}\text{counting}\\ \text{measure}\end{array};L_4\right)$$

where $L_4(\tau)$ is

$$\begin{cases} \tau|_{\text{switch off}} & \text{if light one off and switch off}\\ 0 & \text{if light one on and switch off}\\ \begin{bmatrix} 0 & 0 & 0 & 0\\ 1 & 1 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}\tau|_{\text{switch on}} & \text{if first on and switch on}\\ \begin{bmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 1\end{bmatrix}\tau|_{\text{switch on}} & \text{if first off and switch on}\end{cases} \tag{4.17}$$

pince-nez 5 Same as for pince-nez 4 except the second light replaces the first light.

pince-nez 6

$$\left(\{(1,1),(1,2),\ldots,(4,4)\}\times[0,2\pi),\ \begin{array}{l}\text{counting}\\ \text{measure}\end{array}\times\begin{array}{l}\text{Lebesgue}\\ \text{measure}\end{array};\right. \tag{4.18}$$
$$\left.\left\{\begin{array}{l}\text{third and fourth on,}\\ \text{third on,}\\ \text{fourth on,}\\ \text{third and fourth off}\end{array}\right\},\ \begin{array}{l}\text{counting}\\ \text{measure}\end{array};;L_6\right)$$

where $L_6(\tau)$ is

$$\int_0^{2\pi}\left[\begin{array}{cccccccccccccccc} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\end{array}\right]\tau(\theta)\,d\theta \tag{4.19}$$

if both third and fourth lights are on,

$$\int_0^{2\pi}\left[\begin{array}{cccccccccccccccc} 0 & 0 & 0 & 0 & 0 & 0 & \cos^2\theta & \frac{1}{2} & 0 & 0 & 0 & 0 & \cos^2\theta & \frac{1}{2} & 0 & 0\end{array}\right]\tau(\theta)\,d\theta \tag{4.20}$$

if the third light is on and the fourth is off,

$$\int_0^{2\pi} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \sin^2\theta & \frac{1}{2} & 0 & 0 & 0 & 0 & \sin^2\theta & \frac{1}{2} & 0 & 0 \end{bmatrix} \tau(\theta)\, d\theta \tag{4.21}$$

if the fourth light is on and the third is off, and

$$\int_0^{2\pi} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \tau(\theta)\, d\theta \tag{4.22}$$

if both third and fourth lights are off.

Using this information, the joint probability, given that the switch is off and the dial is set to θ, for both the first pair being off, the third light being on, and the fourth light being off, is then given by (4.11), which is $\int_0^{2\pi} \cos^2\theta'\, g(\theta')\, d\theta' \approx \cos^2\theta$. Similarly, the joint probability, given that the switch is on and the dial is set to θ, for both the first and the third light being on, with the second and fourth being off, is then given by (4.13), which is $\int_0^{2\pi} \frac{1}{4} g(\theta')\, d\theta' = \frac{1}{4}$. The other joint probabilities can be calculated similarly and agree with the joint probabilities originally given for the black box.

4.8 What is a quantum system?

For the two-slit experiment, which only required finite-dimensional linear algebra in the quantum model (apart from the already commented on problem of inputting the dial setting Θ), a classical model with the same behavior also only required finite-dimensional linear algebra. This is atypical. As we have seen in §3.5, a classical model that duplicates the behavior of a quantum model is generally far more complicated. Also, these classical models are generally inherently contextual.

A reason why the name "quantum system" could still be applied to this example is universality: while both models have been constructed to be potentially universal, testing will reveal if they fail in this regard. If the black boxes were opened to reveal a laser, beam-splitters, photon detectors, and so on, from experience we would have a lot of confidence that the modules in the quantum model are universal, whereas we would have very little confidence for the modules in the classic model to have that property. Conversely, if the boxes were opened to reveal regular computer circuits, the situation would be reversed.

Chapter 5

Relation to textbook quantum mechanics

5.1 Textbook rules for quantum mechanics

The quantum Bayesian network structure developed so far has constraints only arising from the requirements of positivity, linearity, and potential universality. It is not obvious that it has any connection necessarily to what is usually thought of as quantum mechanics. Using option **I'**, that is not the situation, as is shown below.

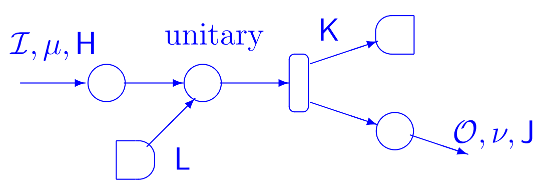
However, while the following justifies the usage of the name *quantum* for the extended Bayesian networks, it should not be taken as a justification of them–quite the opposite. The Bayesian network approach is predicated on a very reasonable basis. The textbook approach to quantum mechanics is only comfortable due to familiarity; on its own merits its rules are incomprehensible and unmotivated. Hence, the following is better taken as a justification of the textbook approach. In other words, if quantum behavior had first been discovered as forecasting arising from a reasonable extension of Bayesian networks, and the usual list of textbook rules were later discovered picking out a certain subset of networks that were sufficient to model any situation, no one would take those rules as primary.

Rule one

Textbook quantum mechanics imposes an additional rule: only networks in the form of chains are permitted and all the Hilbert spaces along a chain must be the same (apart for terminating trivial spaces). (This rule may arise from imagining that we have some sort of universal, initial-value, dynamical system). Mathematically, this rule puts no additional constraints on the network formulation; it is always possible to obey this by working in a sufficiently large Hilbert space and waiting until the end of the calculation to perform the reduced traces arising from terminal nodes. However, doing this in practice is unnecessarily difficult–the whole point of Bayesian networks is to try to find a model made of simple and (hopefully) universal parts where the calculations can be done in a manageable manner. Therefore, we will discard this rule.

Rule two

Another rule from textbook quantum mechanics is: the linear map for a quantum node can only be of the form $L\rho = U\rho U^*$, where U is unitary. As we prove in theorem **5.1.1** below, this is no limitation at all mathematically if option **I'** is employed. If option **II** is used, the theorem still holds if H is separable and if there is some strictly positive[1], finite Radon measure on $\mathcal{X}$. For any node with data $(\mathcal{I}, \mu, \mathsf{H}; \mathcal{O}, \nu, \mathsf{J}; L)$, the map L can be represented as the sequence of three operations, using some auxiliary Hilbert space K: (**i**) injecting $\rho \in L^1(\mathcal{I}; \mu; \mathcal{D}(\mathsf{H}))$ as a density matrix on a Hilbert space $\mathsf{K} \otimes L^2(\mathcal{O}; \nu) \otimes \mathsf{J}$ utilizing a partial isometry; (**ii**) taking a reduced trace of the density matrix over K; (**iii**) mapping the resulting density matrix on $L^2(\mathcal{O}; \nu) \otimes \mathsf{J}$ into $L^1(\mathcal{O}; \nu; \mathcal{D}(\mathsf{J}))$. By introducing another auxiliary Hilbert space L with some fixed template density-matrix on it, where $\mathsf{L} \otimes L^2(\mathcal{I}; \mu) \otimes \mathsf{L} \cong \mathsf{K} \otimes L^2(\mathcal{O}; \nu) \otimes \mathsf{J}$, the partial isometry can be upgraded to an unitary operator. In terms of the graphical model, this means replacing a node with the graph fragment:



(5.1)

Of course, just because one can do this does not mean one must, or that one should. To obey this rule, a great deal of extra computation and irrelevant, arbitrary choices must be made to no benefit. Therefore, this rule will also be discarded.

Rule three

Yet another rule from textbook quantum mechanics is: the linear map for a terminal pince-nez (a destructive "measurement") can only be of the form $\int_B L\rho\, d\nu = \mathrm{tr}\ E_B \rho$, where $\{E_B\}_{B \in \mathcal{E}}$ are a complete set of mutually commuting, orthogonal projectors[2]. For the other pince-nez, only σ-algebras generated by a countable set of atoms are allowed, and the map must be of

[1]Every open set gets measure greater than zero.

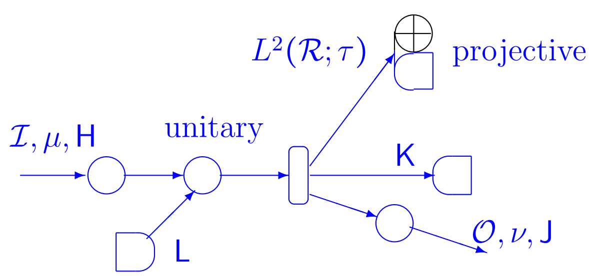
[2]$\{E_B\}_{B \in \mathcal{E}}$ is a *complete set of mutually commuting, orthogonal projectors* for a σ-algebra $\mathcal{E}$ of subsets of a set $\mathcal{X}$ if they are all mutually commuting, orthogonal projectors with $E_\varnothing = 0$, $E_\mathcal{X} = I$, and $E_{\bigcup_j B_j} = \sum_j E_{B_j}$ for all countable, disjoint collections $\{B_j\} \subset \mathcal{E}$, with convergence of the sum in the weak* topology induced on $\mathcal{B}(\mathsf{H})$ by its being dual to the trace-class operators $\mathcal{S}_1(\mathsf{H})$. In this case, convergence in the ultrastrong-operator, strong-operator, ultraweak-operator (same as weak*), and weak-operator topologies on $\mathcal{B}(\mathsf{H})$ are all equivalent (see **A2.1**).

the form

$$\int_B L\rho\, d\nu = \sum_j E_{A_j}\rho E_{A_j} \text{ for } \{A_j\} \text{ a partition of } B \tag{5.2}$$

This rule can also be stated in the form that every "measurement" has an associated operator, which are the coordinates r (taking values as coordinates in the discrete case) for $\mathcal{R}$ acting on $L^2(\mathcal{R};\tau)$ in this formulation. An even stronger formulation requires the result of any "measurement" to be an eigenvalue of the associated operator [15]. This implicitly assumes any results accrued to unions of atoms is due to post-"measurement" garbling and not inherent in the "measurement". Furthermore, if the eigenspace corresponding to any eigenvalue has dimension greater than one, this formulation fails to distinguish among the myriad of possible complete sets of mutually commuting, orthogonal projectors consistent with its prescription, giving rise to the "three-box paradox" [2] if there is post-conditioning.

Again employing option **I'**, as we prove in theorem **5.1.1** below, this third textbook rule also imposes no mathematical limitation. Using option **II**, the theorem still holds with the same limitations given before. For any pince-nez with data $(\mathcal{I},\mu,\mathsf{H};\mathcal{R},\tau;\mathcal{O},\nu,\mathsf{J};L)$, the map L can be represented as the sequence of four operations, using some auxiliary Hilbert space K: (**i**) injecting $\rho \in L^1(\mathcal{I};\mu;\mathcal{D}(\mathsf{H}))$ as a density matrix on a Hilbert space $\mathsf{K}\otimes L^2(\mathcal{R}\times\mathcal{O};\tau\times\nu)\otimes\mathsf{J}$ utilizing a partial isometry; (**ii**) making a projective measurement on the $L^2(\mathcal{R};\tau)$ portion; (**iii**) taking a reduced trace of the density matrix over K; and (**iv**) mapping the resulting density matrix on $L^2(\mathcal{O};\nu)\otimes\mathsf{J}$ into $L^1(\mathcal{O};\nu;\mathcal{D}(\mathsf{J}))$. By introducing another auxiliary Hilbert space L with some fixed template density-matrix on it, where $\mathsf{L}\otimes L^2(\mathcal{I};\mu)\otimes\mathsf{L}\cong\mathsf{K}\otimes L^2(\mathcal{R}\times\mathcal{O};\tau\times\nu)\otimes\mathsf{J}$, the partial isometry can be upgraded to an unitary operator. In terms of the graphical model, this means replacing a pince-nez with the graph fragment:



(5.3)

Once again, just because one can do this does not mean one must, or that one should. To obey this rule, a great deal of extra computation and irrelevant, arbitrary choices must be made to no benefit. Therefore, this rule will also be discarded.

The main theorem

The statement of the aforementioned theorem is:

Theorem 5.1.1 If $L\in\mathcal{B}\left(L^1(\mathcal{I};\mu;\mathcal{D}(\mathsf{H})^+),L^1(\mathcal{O};\nu;\mathcal{D}(\mathsf{J})^+)\right)$ is completely positive, then it can be written as

$$L\rho = \Theta\circ\left(\mathrm{tr}_{\mathsf{K}}\otimes I_{\mathcal{B}(L^2(\mathcal{O};\nu))}\otimes I_{\mathcal{B}(\mathsf{J})}\right)\left(V(\tilde{\eta}_*^{-1}\rho)V^*\right)$$

for some Hilbert space K, some map $\tilde{\eta}_*^{-1} : L^1(\mathcal{I};\mu;\mathcal{D}(\mathsf{H})^+) \to \mathcal{D}\left(L^2(\mathcal{I};\mu)\otimes\mathsf{H}\right)^+$, some partial isometry $V : L^2(\mathcal{I};\mu)\otimes\mathsf{H} \to \mathsf{K}\otimes L^2(\mathcal{O};\nu)\otimes\mathsf{J}$, and some map $\Theta : \mathcal{D}\left(L^2(\mathcal{O};\nu)\otimes\mathsf{J}\right)^+ \to L^1(\mathcal{O};\nu;\mathcal{D}(\mathsf{J})^+)$. If all the Hilbert spaces are finite dimensional, the dimension of K is less than or equal to the product of dimensions $\dim\mathsf{J}\dim\mathsf{H}\dim L^2(\mathcal{I};\mu)$.

Comments Writing L in the preceding form gives the behavior asserted above for the pince-nez since, rewriting $\mathcal{O}\to\mathcal{R}\times\mathcal{O}$ and $\nu\to\tau\times\nu$, then for any τ-measurable subset $B\subset\mathcal{R}$, the map Θ is such that

$$\int_B L\rho\, d\tau = \Theta'\left(\mathrm{tr}_{\mathsf{K}}\otimes P(B)\otimes I_{\mathcal{B}(L^2(\mathcal{O};\nu))}\otimes I_{\mathcal{B}(\mathsf{J})}\right)\left(V(\tilde{\eta}_*^{-1}\rho)V^*\right) \tag{5.4}$$

where $P(B)$ is the functional on $\mathcal{D}(L^2(\mathcal{O};\nu))$ given by $P(B)\tau = \int_{x\in B}\tau(x,x)$ $d\nu(x)$ and Θ' is some map from $\mathcal{D}\left(L^2(\mathcal{O};\nu)\otimes\mathsf{J}\right)^+$ to $L^1(\mathcal{O};\nu;\mathcal{D}(\mathsf{J})^+)$. The assertion for the node follows since a node is a pince-nez with trivial base measure τ for $\mathcal{R}$.

From **C5.12**, the theorem also holds for option **II** as long as the given conditions are held: H is separable (so there is some strictly positive density matrix on it) and there is some strictly positive, finite Radon measure μ on $\mathcal{X}$ (without strict positivity, some of the needed inverses will not exist). Note that from **C5.12**, it is not possible to choose ν arbitrarily; it depends on both μ and the map L. Then we have

$$L(1_B\otimes f) = \tilde{\eta}^{-1}\left(V^*\left(I_{\mathsf{K}}\otimes\pi(1_B)\otimes\eta(f)\right)V\right) \tag{5.5}$$

where $\pi(g)$ is the operator on $L^2(\mathcal{R};\tau)$ given by pointwise multiplication by g for any $g\in L^\infty(\mathcal{R};\tau)$), $\tilde{\eta}^{-1} : \mathcal{B}\left(L^2(\mathcal{I};\mu)\otimes\mathsf{H}\right)^+ \to L^\infty(\mathcal{X};\mu;\mathcal{K}(\mathsf{H}))$–which contains $\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H}))$–and $\eta : \mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J}))^+ \to \mathcal{B}(L^2(\mathcal{O};\nu)\otimes\mathsf{J})^+$.

The following proof, together with a following lemma and several supporting propositions in the appendix, is quite lengthy. The main idea is to recognize that the dual space to density-matrix-valued, L^1-functions is a von Neumann algebra[3]. Then one result by Stinespring [88] on completely-positive maps on C^*-algebras and another result by Sakai [79] on representations of von Neumann algebras provide the core. Most of the rest is just the tedious work of proving the preconditions necessary for these two results apply.

Proof Start with any such L. Extend L to $L^1(\mathcal{I};\mu;\mathcal{S}_1(\mathsf{H}))$ using **B5.15** and **B5.16**[4]. L induces an adjoint map $L^* : L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^* \to L^1(\mathcal{I};\mu;\mathcal{S}_1(\mathsf{H}))^*$ by $L^*(\Phi)\rho = \Phi(L\rho)$. By **B5.19**, there is a weakly-continuous, isometric bijection $\eta : L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^* \to \mathcal{W}(\mathcal{O};\nu;\mathsf{J})$, where $\mathcal{W}(\mathcal{O};\nu;\mathsf{J})\subset\mathcal{B}(L^2(\mathcal{O};\nu)\otimes\mathsf{H})$ is a von Neumann algebra with unit. The bijection therefore induces a C^*-algebra structure on $L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^*$ with W^*- representation[5] $(\eta, L^2(\mathcal{O};\nu)\otimes\mathsf{J})$. Also by **B5.19**, there is a weakly-continuous, isometric bijection

[3] A *Banach algebra* is a Banach space equipped with an associative, distributive product satisfying $\|ab\|\le\|a\|\|b\|$. A **-algebra* is a Banach algebra with an antilinear involution. A *C^*-algebra* is a *-algebra where $\|a^*a\| = \|a\|^2$. A *von Neumann algebra* (also termed a *W^*-algebra*) is a C^*-algebra which, as a Banach space, is a dual space.

[4] $\mathcal{S}_1(\mathsf{H})$ denotes all trace-class operators on H.

[5] A *W^*-representation* has a weakly-continuous map.

$\tilde{\eta} : L^1(\mathcal{I};\mu;\mathcal{S}_1(\mathsf{H}))^* \to \mathcal{W}(\mathcal{I};\mu;\mathsf{H}) \subset \mathcal{B}(L^2(\mathcal{I};\mu)\otimes\mathsf{H})$. By **B5.20**, $\tilde{\eta}\circ L^*$ is completely-positive (in the sense required for Stinespring's theorem). Therefore, by Stinespring's theorem[88], $\tilde{\eta}\circ L^*(\Phi) = V^*\zeta(\Phi)V$ for some Hilbert space M, some representation (ζ,M), and some $V \in \mathcal{B}(L^2(\mathcal{I};\mu)\otimes\mathsf{H},\mathsf{M})$.

By lemma **5.1.2** below, (ζ,M) is a W^*-representation. Therefore, since η is faithful, (ζ,M) is related to $(\eta, L^2(\mathcal{O};\nu)\otimes\mathsf{J})$ (up to unitary equivalence) by an *amplification* composed with an *induction*[79]: $\zeta(\Phi) = W^*E\left(I_{\mathsf{K}}\otimes\eta(\Phi)\right)W$ for some Hilbert space K with $\mathsf{M} = \mathsf{K}\otimes L^2(\mathcal{O};\nu)\otimes\mathsf{J}$, some orthogonal projector E in the commutant of $I_{\mathsf{K}}\otimes\mathcal{W}(\mathcal{O};\nu;\mathsf{J})$, and some unitary operator $W\in\mathcal{B}(\mathsf{M})$. However, W and E can be absorbed into a redefinition of V, $V\to EWV$, so ζ can be taken to be simply the amplification of η, $\zeta\Phi = I_{\mathsf{K}}\otimes(\eta\Phi)$.

Then, for any $\Phi\in L^1(\mathcal{I};\mu;\mathcal{D}(\mathsf{H}))^*$ and $\rho\in L^1(\mathcal{I};\mu;\mathcal{D}(\mathsf{H})^+)$,

$$\begin{aligned}\Phi\, L\rho &= (L^*\Phi)\rho = ((\tilde{\eta}^{-1}\circ\tilde{\eta}\circ L^*)\Phi)\rho = \left(V^*\left(I_{\mathsf{K}}\otimes\eta\Phi\right)V\right)\tilde{\eta}_*^{-1}\rho \\ &= \left(I_{\mathsf{K}}\otimes\eta\Phi\right)V(\tilde{\eta}_*^{-1}\rho)V^* \\ &= (\eta\Phi)\left((\mathrm{tr}_{\mathsf{K}}\otimes I_{\mathcal{B}(L^2(\mathcal{O};\nu))}\otimes I_{\mathcal{B}(\mathsf{J})})V(\tilde{\eta}_*^{-1}\rho)V^*\right) \\ &= \Phi\left(\Theta\circ(\mathrm{tr}_{\mathsf{K}}\otimes I_{\mathcal{B}(L^2(\mathcal{O};\nu))}\otimes I_{\mathcal{B}(\mathsf{J})})V(\tilde{\eta}_*^{-1}\rho)V^*\right)\end{aligned}\tag{5.6}$$

where Θ and $\tilde{\eta}_*^{-1}$ are defined in **B5.19** (note $\tilde{\eta}_*^{-1}$ is only defined up to equivalence class, but the choice of element within this class is of no consequence). Since Φ was arbitrary, the desired form has been demonstrated.

Since L is an isometry on the positive cone, for any ρ in this cone with $\|\rho\| = 1$,

$$\begin{aligned}1 = \|L\rho\| &= \Phi\left(\Theta\circ(\mathrm{tr}_{\mathsf{K}}\otimes I_{\mathcal{B}(L^2(\mathcal{O};\nu))}\otimes I_{\mathcal{B}(\mathsf{J})})V(\tilde{\eta}_*^{-1}\rho)V^*\right) \\ &= \tilde{\eta}^{-1}\left(V^*\left(I_{\mathsf{K}}\otimes\eta\Phi\right)V\right)\rho\end{aligned}\tag{5.7}$$

where Φ is the functional $\varphi\to\int_{\mathcal{O}}\mathrm{tr}_{\mathcal{J}}\varphi\,d\nu$. However, then $\eta\Phi = I_{L^2(\mathcal{O};\nu)}\otimes I_{\mathsf{J}}$, so we have

$$1 = \tilde{\eta}^{-1}\left(V^*V\right)\rho\tag{5.8}$$

In order for $\tilde{\eta}^{-1}\left(V^*V\right)$ to be the functional $\rho\to\int_{\mathcal{I}}\mathrm{tr}_{\mathcal{H}}\rho\,d\mu$, it must be that $V^*V = I_{L^2(\mathcal{I};\mu)}\otimes I_{\mathsf{H}}$, so V is a partial isometry.

Furthermore, if all the Hilbert spaces are finite-dimensional, it is possible to bound the dimension of K. By **B5.21**, the Hilbert space $L^2(\mathcal{O};\nu)\otimes\mathcal{S}_2(\mathsf{J})\cong L^2(\mathcal{O};\nu;\mathcal{S}_2(\mathsf{J}))$ is relatively dense within $\mathcal{W}(\mathcal{O};\nu;\mathsf{J})$ in the ultrastrong-operator topology. Therefore, given any $\sum_{j=1}^m\Phi_j\otimes\phi_j\in L^1\left(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J})\right)^*\otimes\left(L^2(\mathcal{I};\mu)\otimes\mathsf{H}\right)$ and $\varepsilon>0$, there is a subset $\{\Psi_j\}\subset\eta^{-1}\left(\mathcal{W}\cap L^2(\mathcal{O};\nu)\otimes\mathcal{S}_2(\mathsf{J})\right)$ such that[6]

$$\begin{aligned}\varepsilon &> \sum_{j,k=1}^m\left((\Phi_k-\Psi_k)^\star\left(\Phi_j-\Psi_j\right)\right)L\left(\tilde{\eta}_*(\phi_j\phi_k^*)\right) \\ &= \sum_{j,k=1}^m\left\langle(\tilde{\eta}\circ L^*)\left((\Phi_k-\Psi_k)^\star\left(\Phi_j-\Psi_j\right)\right)\phi_j,\phi_k\right\rangle\end{aligned}\tag{5.9}$$

[6]To avoid confusion with the adjoint map $\Phi^*:\mathbb{C}\to L^1\left(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})\right)^*$, $\star$ is used to indicate the conjugate within the von Neumann algebra, with $\Phi^\star = \eta^{-1}\left(\eta(\Phi)^*\right)$. However, if the dual pairs $\langle w,z\rangle$ for $w,z\in\mathbb{Z}$ and $\langle L,\Phi\rangle$ are taken complex, with values $w\bar{z}$ and $\Phi^\star L$ respectively, then $\Phi^\star z = z\Phi^\star$.

From the construction of M in the proof to Stinespring's theorem[88] as the completion of certain quotient space of $L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^* \otimes (L^2(\mathcal{I};\mu)\otimes \mathsf{H})$, this is equivalent to

$$\left\| \left[\sum_{j=1}^m \Phi_j \otimes \phi_j\right] - \left[\sum_{k=1}^m \Psi_k \otimes \phi_k\right]\right\|^2 < \varepsilon \tag{5.10}$$

where $[\cdot]$ indicates the equivalence class. Hence, letting ι be the quotient map, the bounded, linear map $\iota \circ (\eta^{-1} \otimes I_{L^2(\mathcal{I};\mu)} \otimes I_{\mathsf{H}})$ sends $(\mathcal{W} \cap L^2(\mathcal{O};\nu)\otimes \mathcal{S}_2(\mathsf{J}))\otimes (L^2(\mathcal{I};\mu)\otimes \mathsf{H})$ to a norm-dense subset of M. Since $L^2(\mathcal{O};\nu)\otimes\mathcal{S}_2(\mathsf{J})\otimes(L^2(\mathcal{I};\mu)\otimes\mathsf{H})$ is a Hilbert space, the dimension of M is necessarily less than or equal to the product of dimensions:

$$(\dim \mathsf{J})^2 \cdot \dim \mathsf{H} \cdot \dim L^2(\mathcal{I};\mu)\cdot \dim L^2(\mathcal{O};\nu) \tag{5.11}$$

consequently, the dimension of K is less than or equal to $\dim \mathsf{J}\cdot \dim\mathsf{H}\cdot\dim L^2(\mathcal{I};\mu)$. □

Lemma 5.1.2 ζ is weak*-continuous.

Proof Let

$$\Omega : L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^* \otimes (L^2(\mathcal{I};\mu)\otimes\mathsf{H}) \otimes L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^* \otimes (L^2(\mathcal{I};\mu)\otimes\mathsf{H})^* \to L^1(\mathcal{O};\mu;\mathcal{S}_1(\mathsf{H})) \tag{5.12}$$

be given by

$$\Omega(\Phi\otimes\phi\otimes\Phi'\otimes\phi'^*) = \eta_*\left(\eta(\Phi)(\eta_*^{-1}\circ L(\tilde{\eta}_*(\phi\otimes\phi'^*))\eta(\Phi')\right) \tag{5.13}$$

By a similar argument to that employed in **B5.20**, since L is completely-positive, Ω is positive in the sense that for any $\xi \in L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^*\otimes L^2(\mathcal{I};\mu)\otimes\mathsf{H}$, $\Omega(\xi\otimes\xi^*)\geq 0$. More generally, we have

$$\begin{aligned}\int_{\mathcal{O}} \mathrm{tr}_{\mathsf{J}}\Omega\left(\left(\sum_{j=1}^m \Phi_j\otimes\psi_j\right)\otimes\left(\sum_{k=1}^n \Phi'^{\star}_k\otimes\psi'^{*}_k\right)\right)d\nu &= \sum_{j=1}^m\sum_{k=1}^n(\Phi'^{\star}_k\Phi_j)L(\tilde{\eta}_*(\phi_j\otimes\phi'^{*}_k) \\ &= \sum_{j=1}^m\sum_{k=1}^n\left\langle(\tilde{\eta}^{-1}\circ L^*)(\Phi'^{\star}_k\Phi_j)\psi_j,\psi'_k\right\rangle\end{aligned} \tag{5.14}$$

By the construction of the Hilbert space M as a certain quotient space of $L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^*\otimes L^2(\mathcal{I};\mu)\otimes\mathsf{H}$ in the proof of Stinespring's theorem[88], this is equal to the inner-product

$$\left\langle\left[\sum_{j=1}^m\Phi_j\otimes\psi_j\right],\left[\sum_{k=1}^n\Phi'_k\otimes\psi'_k\right]\right\rangle \tag{5.15}$$

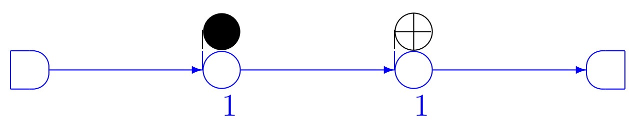
Therefore, for any $\xi\in[0]$, $\Omega(\xi\otimes\xi^*)=0$. By the positivity of Ω, $\Omega((\chi+z\xi)\otimes(\chi+z\xi)^*)\geq 0$ for any $\chi\in L^1(\mathcal{O};\nu;\mathcal{S}_1(\mathsf{J}))^*\otimes L^2(\mathcal{I};\mu)\otimes\mathsf{H}$ and $z\in\mathbb{C}$. Taking z on the real axis, it follows that $\Omega(\xi\otimes\chi^*)+\Omega(\chi\otimes\xi^*)=0$. Taking z on the imaginary axis, it follows that $\Omega(\xi\otimes\chi^*)-\Omega(\chi\otimes\xi^*)=0$. Therefore, $\Omega(\xi\otimes\chi^*)=\Omega(\chi\otimes\xi^*)=0$.

Hence, $\Omega(\chi \otimes \tau^*)$ only depends on the equivalence classes $[\chi], [\tau]$, so by continuity it induces a map $\zeta_* : \mathsf{M} \otimes \mathsf{M}^* \to L^1(\mathcal{O}; \nu; \mathcal{S}_1(\mathsf{J}))$. By the spectral theorem for compact operators, linearity, and continuity, ζ_* extends to a map $\zeta_* : \mathcal{S}_1(\mathsf{M}) \to L^1(\mathcal{O}; \nu; \mathcal{S}_1(\mathsf{J}))$. This map is positive, is an isometry on the positive cone (hence is bounded with operator norm two) by (5.14), and satisfies $\mathrm{tr}\, \zeta(\Phi)\rho = \Phi(\zeta_*(\rho))$ for any $\Phi \in L^1(\mathcal{O}; \nu; \mathcal{S}_1(\mathsf{J}))$ and $\rho \in \mathcal{S}_1(\mathsf{M})$ by the construction of $\zeta : L^1(\mathcal{O}; \nu; \mathcal{S}_1(\mathsf{J})) \to \mathcal{B}(\mathsf{M})$ in the proof of Stinespring's theorem[88]. Hence, $\zeta = (\zeta_*)^*$. □

5.2 Weak measurements

Reproducibility as a criteria

One motivation for the restriction to projective maps, emphasized by Dirac [16], is reproducibility: if two successive pince-nez have the same data (and the incoming and outgoing data are compatible), then conditioning on the observation for the first being in some measurable subset $B \subset \mathcal{R}$, the probability measure for the second observation has the property that on measurable sets that do not intersect B, it gives zero, whereas on measurable sets that contain B, it gives one. The Bayesian network for this set-up is:



(5.16)

where the repeated '1's indicate the data for the two pince-nez are the same.

While this criteria looks superficially attractive–as is implied in the saw "measure twice, cut once"–its flaw is readily apparent. Suppose I were asked to see whether or not a light in another room were off. If I see the light is off, I return and report it being off. If I see it is on, I turn it off, then return and also report it being off. This is repeatable, but does not correspond to what we generally mean by the measurement of the state of the light. While this example may seem obtuse, the point is that mathematically projective maps in quantum mechanics act in this manner–they have a significant effect even if the results are ignored.

Weak versus strong maps for pince-nez

Building from the preceding example, if the incoming and outgoing data for a pince-nez are compatible, then the map L is termed *strong* if, when marginalizing over the observation, the result (using option **I'**), $\int_{\mathcal{R}} L(\rho)\, d\tau$, is far from ρ for all but a small subset of ρ's. Conversely, the map L is termed *weak* if $\int_{\mathcal{R}} L(\rho)\, d\tau$ is close to ρ for a large subset of ρ's. Obviously, there is a continuum of possibilities; the strength could be varied depending on an input parameter. For instance, in the double-slit example previously considered, the switch control could be replaced by a continuous slider that would slowly vary the joint probabilities of the output.

Now we can easily understand the misconception leading to the repeatability criteria; in the classic case, restriction maps (using diagonal projectors) are both repeatable and the weakest possible, having no effect if their observations are ignored. In the quantum case, however, while projective maps (using orthogonal projectors) are still repeatable, they can also be very strong.

Information as a criteria

Another misconception is that the result of any pince-nez map can be duplicated by a projective map followed by a garbling of the observation data[7]. If this were the case, then, by convexity, the information[8] using a projective map would always be higher than for any other map in every situation. Consequently, the claim by Kochen and Specker that projective "measurements" give maximal knowledge [31] would be correct. However, it is easy to see this is false. For instance, the matrices

$$\phi_1 = \begin{bmatrix} \frac{2}{3} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}, \phi_2 = \begin{bmatrix} \frac{1}{6} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{3} \end{bmatrix}, \phi_3 = \begin{bmatrix} \frac{1}{6} & -\frac{1}{5} \\ -\frac{1}{5} & \frac{1}{3} \end{bmatrix} \tag{5.17}$$

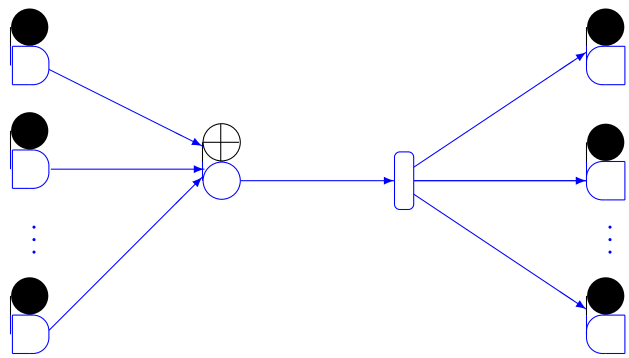
are all positive, sum to the identity, and do not mutually commute. The map $L : \mathcal{D}(\mathbb{C}^2) \to \mathbb{R}^3$ given by $L(\rho) = (\mathrm{tr}\ \phi_1\rho, \mathrm{tr}\ \phi_2\rho, \mathrm{tr}\ \phi_3\rho)$ is completely-positive and norm-preserving (on the positive cone), yet cannot be duplicated by any garbling of a projective map, $\sum_{k=1}^{2} g_{jk}\mathrm{tr}\ E_k\rho$, for stochastic matrix $G = [g_{jk}]$ and commuting, complete, orthogonal projectors $\{E_1, E_2\}$ on $\mathbb{C}^2$ since the matrices $\{g_{11}E_1 + g_{12}E_2, g_{21}E_1 + g_{22}E_2, g_{31}E_1 + g_{32}E_2\}$ necessarily mutually commute.

Also, even if it were possible to duplicate the behavior of a particular pince-nez map by an projective map followed by a garbling of the observation data when the module is used in isolation, if the module is then inserted into a larger Bayesian network where conditioning is taking place, the difference in behavior between the particular pince-nez map and the projective map followed by a garbling may become quite large. In particular, for so-called "weak-measurements" [2], which employ both strong and weak pince-nez maps together with both pre- and post-conditioning, the information can be very large. The typical Bayesian

[7] For instance, Dirac seemed to believe "measurements" were always inherently projective. [16]

[8] Information in the sense of the Shannon definition [85] for the probability measure on the observation set $\mathcal{R}$.

network for one of these has the form:

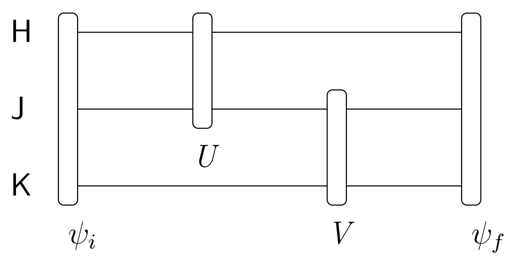


(5.18)

There is both pre- and post-conditioning; those pince-nez have projective maps. The central pince-nez has a map that uses "fuzzy" projectors with spread of order $\sqrt{N}$ for N particles. Hence, the relative standard deviation in the values assigned to the observation set is only of order $\frac{1}{\sqrt{N}}$, so the information is high; furthermore, the center of the peak of the distribution can be in unexpected ranges of the observation set, as has been well-publicized recently [9].

5.3 Comparison of Bayesian networks to quantum circuits and tensor networks

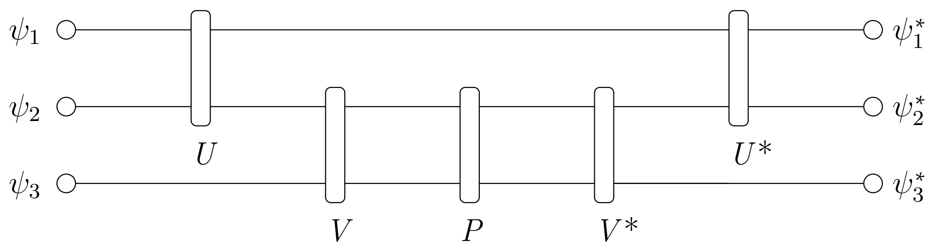
A commonly-used, alternative graphical approach for quantum systems is quantum circuits [40]. For example consider the following quantum circuit:



(5.19)

Then for Hilbert spaces H, J, and K, termed the *quantum channels*, and initial state $\psi_i \in \mathsf{H} \otimes \mathsf{J} \otimes \mathsf{K}$, the final state ψ_f is given by $(I_{\mathsf{H}} \otimes V)(U \otimes I_{\mathsf{K}})\psi_i$ for some unitary $U : \mathsf{H} \otimes \mathsf{J} \to \mathsf{H} \otimes \mathsf{J}$ and $V : \mathsf{J} \otimes \mathsf{K} \to \mathsf{J} \otimes \mathsf{K}$. Quantum circuits are obviously directly related to the usual textbook quantum mechanics working with wavefunctions. Following the first textbook rule from above, there is a single overall Hilbert space, the tensor product of all the Hilbert spaces for each quantum channel, that is used throughout. Only unitary operators can be accommodated, following the second textbook rule above.

It is possible to incorporate projective "measurements" into the quantum circuit, giving a tensor network, but this requires doubling the graph–clearly not an efficient approach graphically. For instance, suppose the initial state is a product state, $\psi_i = \psi_1 \otimes \psi_2 \otimes \psi_3$ and that a projective "measurement" is made on the final state, where the projector is of the form $I_{\mathsf{H}} \otimes P$. Then the calculation of the probability $\|I_{\mathsf{H}} \otimes P\psi_f\|^2 = \langle I_{\mathsf{H}} \otimes P\psi_f, \psi_f\rangle$ is represented by the tensor network



(5.20)

The tensor network has some of the advantages of the Baysian network in that once it is set up, it is possible to look for computational shortcuts.

On the other hand, the Bayesian network has many advantages. It does not depend on the incomprehensible and unmotivated textbook rules, but instead stands on its own reasonable basis. It is graphically more efficient in not having to duplicate itself to include observations. It is also potentially more efficient in allowing non-unitary nodes and non-projective pince-nez, hence avoiding the need to introduce auxiliary spaces, and in having the splitter construction, so it is not necessary to maintain all the quantum circuits throughout the diagram. The Bayesian network indicates which observable nodes are being marginalized or conditioned on and works seamlessly with the usual observable Bayesian networks that are already in common use, so it easily allows models with information coming from previous observations or random factors (such as coin flips). For instance, a tensor network coming from a quantum circuit for the "weak measurement" example (5.18) would be far more complicated, with the need for auxiliary spaces, yet would still not be able to indicate the post-conditioning graphically (the pre-conditioning could be incorporated into the initial values for the wavefunction).

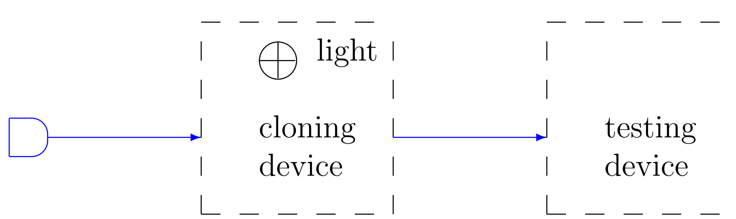
Chapter 6

Further examples

6.1 No-cloning–classical and quantum

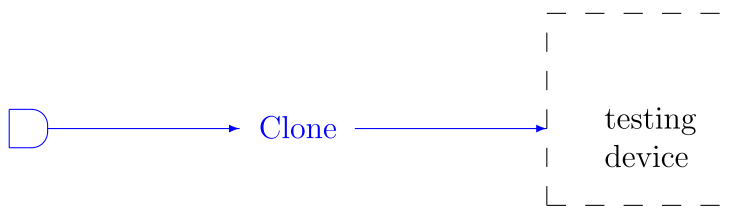
No-cloning is taken as a hallmark of quantum mechanics. However, it is a property of Bayesian networks more generally, holding even if all the Hilbert spaces are trivial. The non-linear map Clone sends the density matrix-valued vector measure ρ to $\rho \otimes \rho$. This cannot be implemented by any device that can be modelled by a Bayesian network since such a network, no matter how complicated, altogether gives rise to a linear map on ρ by the principle of linearity.

One may still consider something a cloning device if it does not always clone, but only clones conditioned on a observation; for instance, suppose there is a green light for success and a red light for failure. However, to be a cloning device, it must have a finite probability of success for at least two distinct inputs, μ and ν. By the principle of linearity, then every convex combination $t\mu + (1-t)\nu$ of the two has a finite probability for success. Now consider the following graphical model:



(6.1)

For the device to be cloning, then conditioning on the light being green, the joint probability for the observations on the testing device must be the same as those for the graphical model with the Clone map



(6.2)

with the same corresponding data for the testing device for each model. For every possibility of what is put in the dashed boxes and what the corresponding information is, there are some linear maps L from $\mathcal{D}(\mathsf{H})$-valued vector measures on $\mathcal{X}$ to $\mathcal{D}(\mathsf{H}\otimes\mathsf{H})$-valued vector measures on $\mathcal{X}\times\mathcal{X}$ and K, M from the latter space to $[0,1]$ such that the joint probability (conditioning on the light being green) using the purported cloning device is $\frac{(K\circ L)\rho}{(M\circ L)\rho}$, whereas the joint probability using the mathematical Clone map is $K\text{Clone}(\rho) = K(\rho\times\rho)$. With the above assumption, $(M\circ L)(t\mu+(1-t)\nu)) \neq 0$; however, then we must have

$$(M\circ L)(t\mu+(1-t)\nu)K((t\mu+(1-t)\nu)\times(t\mu+(1-t)\nu)) = (K\circ L)(t\mu+(1-t)\nu) \tag{6.3}$$

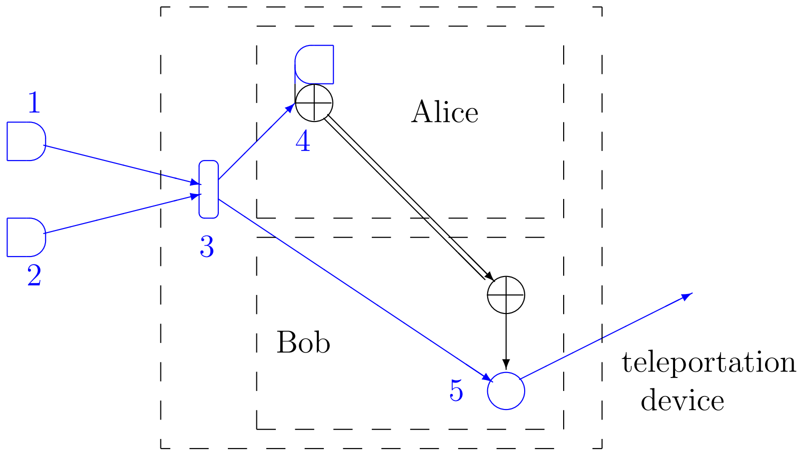
for all $t\in[0,1]$, which is impossible.

6.2 Teleportation-classical and quantum

Teleportation is also generally taken as a hallmark of quantum mechanics. However, as will be shown, teleportation is also possible in classical hidden models.

A device is a teleportation device if, when modelled by a Bayesian network, for a fixed set $\mathcal{X}$, base measure μ, Hilbert space H, and template density-matrix-valued function $\sigma\in L^1(\mathcal{X}\times\mathcal{X};\mu\times\mu;\mathcal{D}(\mathsf{H}\otimes\mathsf{H})^+)$, it takes the product function $\rho\otimes\sigma\in L^1(\mathcal{X}\times\mathcal{X}\times\mathcal{X};\mu\times\mu\times\mu;\mathcal{D}(\mathsf{H}\otimes\mathsf{H}\otimes\mathsf{H})^+)$ back to ρ for any $\rho\in L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})^+)$ and where a splitter is used so only Alice gets the $(\mathcal{X},\mu,\mathsf{H})$ that ρ lives on, Bob gets the final ρ, and there are no hidden connections between Alice and Bob, but only classic, observable information. Note that because the first textbook rule that the overall Hilbert space is the same throughout is unnecessary for the Bayesian network formulation, there is no need to have another output from Alice (whose value is irrelevant in the context of being a teleportation device)–this leads to simplification in the required calculations.

There is a Bayesian network for such a device using quantum nodes (based on, but extending, the calculations of Bennett, Brassard, Crépeau, Josza, Peres, and Wootters [5]) if $\mathcal{X}=\{1,2,\dots,m\}$, μ is the counting measure, H is $\mathbb{C}^n$, and σ is the density-matrix-valued function (written as a column vector with m^2 entries, each a $n^2\times n^2$-matrix) $\frac{1}{m}\text{ Vec } I_m\otimes\left(\frac{1}{n}\text{ Vec } I_n(\text{ Vec } I_n)^{\mathrm{T}}\right)$. Note σ is the Kronecker product of Copy applied to the uniform distribution with the usual maximally-entangled state. The graphical model is:



(6.4)

The double arrow indicates Bob's observations are identical to Alice's (perfect communication). The information presented in the form according to the graphical model is:

node 1 $\left(; \{1,2,\ldots,m\}, \begin{matrix}\text{counting}\\\text{measure}\end{matrix}, \mathbb{C}^n; L_1\right)$ where L_1 is the constant map with value ρ.

node 2 $\left(; \{1,2,\ldots,m\} \times \{1,2,\ldots,m\}, \begin{matrix}\text{counting}\\\text{measure}\end{matrix}, \mathbb{C}^n \otimes \mathbb{C}^n; L_2\right)$ where L_2 is the constant map with value σ.

splitter 3 $\Big(\{1,2,\ldots,m\}^{\times 3}, \begin{matrix}\text{counting}\\\text{measure}\end{matrix}, \mathbb{C}^{n^3};$

$$\{1,2,\ldots,m\} \times \{1,2,\ldots,m\}, \begin{matrix}\text{counting}\\\text{measure}\end{matrix}, \mathbb{C}^n \otimes \mathbb{C}^n, \{1,2,\ldots,m\}, \begin{matrix}\text{counting}\\\text{measure}\end{matrix}, \mathbb{C}^n\Big)$$

pince-nez 4 $\Big(\{1,2,\ldots,m\} \times \{1,2,\ldots,m\}, \begin{matrix}\text{counting}\\\text{measure}\end{matrix}, \mathbb{C}^n \otimes \mathbb{C}^n;$

$$\{1,2,\ldots,m\} \times \{1,2,\ldots,n\} \times \{1,2,\ldots,n\}, \begin{matrix}\text{counting}\\\text{measure}\end{matrix};; L_4\Big)$$

Let Ω be the $n \times n$-matrix with the nth roots of unity along its diagonal, S be the $n \times n$-shift-matrix with a one in the upper-right corner and ones on the subdiagonal, and Q be the $m \times m$-shift-matrix with a one in the upper-right corner and ones on the subdiagonal. For $k, l \in \{1,2,\ldots,n\}$, let w_{kl} be the column vector $(I_n \otimes (\Omega^{l-1} S^{k-1}))$ Vec I_n. Then for $(j,k,l) \in \{1,2,\ldots,m\} \times \{1,2,\ldots,n\} \times \{1,2,\ldots,n\}$, $L_4(\tau)_{jkl}$ is given by

$$\left(\operatorname{Vec} Q^{j-1}\right)^{\mathrm{T}} \left((I_{m^2} \otimes w_{kl}^*) \tau (I_{m^2} \otimes w_{kl})\right) \tag{6.5}$$

node 5 $$\Big(\{1,2,\ldots,m\}\times\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}\times\{1,2,\ldots,n\},\ \begin{matrix}\text{counting}\\ \text{measure}\end{matrix},\mathbb{C}^n;$$

$$\{1,2,\ldots,m\},\ \begin{matrix}\text{counting}\\ \text{measure,}\end{matrix}\ \mathbb{C}^n;L_5\Big)$$

The incoming density-matrix-valued function τ will be considered as a $m{\times}n{\times}n$-array, indexed by j,k,l, of column vectors of m entries each, where each entry is a $n\times n$-matrix. With this convention, $L_5(\tau)$ is given by

$$\sum_{j=1}^{m}\sum_{k,l=1}^{n}((Q^{j-1})^{\mathrm{T}}\otimes I_n)\Big((I_m\otimes(\Omega^{l-1}(S^{k-1})^{\mathrm{T}}))\tau_{jkl}(I_m\otimes(S^{k-1}\overline{\Omega}^{l-1})\Big) \tag{6.6}$$

Then, for any following testing device, the incoming density-matrix-valued function from Bob if Alice observes $(j,k,l)\in\{1,2,\ldots,m\}\times\{1,2,\ldots,n\}\times\{1,2,\ldots,n\}$ is

$$L_5\Big(((L_4\otimes I_{\mathcal{B}(L^1(\mathcal{X};\mu;\mathsf{H}))}))(L_1\otimes L_2))\big|_{j,k,l}1_{\{(j,k,l)\}}\Big) \tag{6.7}$$

This apparently complicated expression is just ρ, regardless of which particular values of j,k,l were observed by Alice and sent to Bob.

The classical case

Note the preceding is meaningful for the case of trivial Hilbert space, $n=1$, so teleportation is not a quantum phenomenon. Another way to achieve the same output in this purely-classical case is to not have a template shared by Alice and Bob, but rather to have the pince-nez for Alice have simply the identity for its map (so the hidden node is not really hidden, but is actually observable). Then Alice sends her information to Bob, who makes a copy using the information. Finally, Alice and Bob forget what the information was (so their nodes are marginalized). The result is that the incoming probability distribution to Alice is the same as that outgoing from Bob.

However, this second approach is not teleportation because of the forgetting step. Also, for the classical case of the teleportation given above, an eavesdropper to the information sent from Alice to Bob without access to the template cannot replicate the probability distribution, whereas in the second, non-teleportation approach, the eavesdropper could not only replicate the probability distribution by making his own copy, then forgetting, but by not forgetting, would actually have more information.

Does a teleportation device really teleport?

Note the graphical model for the teleportation device is simply a graph fragment since the output from Bob is not terminated. One may hope its behavior is universal, but is always possible that additional testing will reveal it is not. For instance, before the discovery of quantum mechanics, it may have been believed that the classical teleportation discussed above was "true" teleportation. It is possible that additional mathematical structures besides measures and density matrices (see §3.8 above) can be used to calculate probabilities in Bayesian networks and that these are found to be useful in practice–then the current belief that quantum telportation is "true" teleportation will also be shown false.

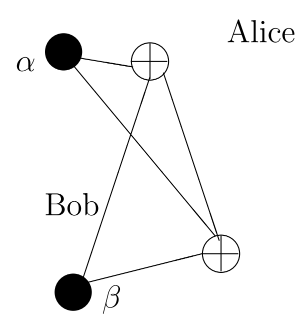
6.3 Bell's inequality for Bayesian networks without metaphysical limitations

Introduction

The standard proofs of Bell's inequality [4] make assumptions based on assuming the underlying reality of hidden variables, which in our language is equivalent to the existence of transition probability function. As we explored in §2.2, this is a metaphysical notion for hidden nodes. For our Bayesian networks, no such limitation is placed, which enlarges the space of possible maps that can be employed. Although, as we already mentioned in §2.2, these additional maps are not necessarily of great interest, we would like to show that Bell's inequality still necessarily holds.

Set-up

Following Clauser, Horne, Shimony, and Holt [11], consider the case of two observers, Alice and Bob, who are at distant locations. Each has a box, connected by a long cable to the other, with a dial and two lights, marked zero and one, which light periodically. They record their observations, then come together later to compare notes. They find that the flashes are independent in time and depend on the settings of the dial, α for Alice and β for Bob, but for each round the lights they each saw were not independent of each other, but had a joint probability distribution given by the four functions of α, β: $\text{Prob}\left((0,0)|\alpha,\beta\right)$, $\text{Prob}\left((0,1)|\alpha,\beta\right)$, $\text{Prob}\left((1,0)|\alpha,\beta\right)$, and $\text{Prob}\left((1,1)|\alpha,\beta\right)$. Presenting the information in this manner corresponds to the graphical model:



(6.8)

There are the additional no-signalling constraints on the probabilities arising from relativity:

$$\begin{aligned}
&\text{Prob}\left((0,0)|\alpha,\beta\right) + \text{Prob}\left((0,1)|\alpha,\beta\right) \text{ is independent of } \beta \\
&\text{Prob}\left((1,0)|\alpha,\beta\right) + \text{Prob}\left((1,1)|\alpha,\beta\right) \text{ is independent of } \beta \\
&\text{Prob}\left((0,0)|\alpha,\beta\right) + \text{Prob}\left((1,0)|\alpha,\beta\right) \text{ is independent of } \alpha \\
&\text{Prob}\left((0,1)|\alpha,\beta\right) + \text{Prob}\left((1,1)|\alpha,\beta\right) \text{ is independent of } \alpha
\end{aligned} \tag{6.9}$$

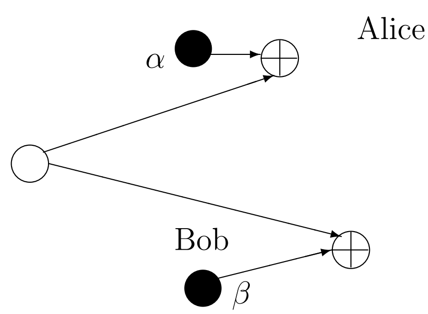
If these constraints were not met, then Alice and Bob could use their dial settings to transmit information superluminally. However, relativity places no additional constraint on the bounds for the value of, for α, α' two settings of Alice's dial and β, β' two settings of Bob's dial,

$$\begin{aligned}
&\text{Prob}\left((0,0)|\alpha,\beta\right) + \text{Prob}\left((1,1)|\alpha,\beta\right) - \text{Prob}\left((0,1)|\alpha,\beta\right) - \text{Prob}\left((1,0)|\alpha,\beta\right) \\
&+\text{Prob}\left((0,0)|\alpha',\beta\right) + \text{Prob}\left((1,1)|\alpha',\beta\right) - \text{Prob}\left((0,1)|\alpha',\beta\right) - \text{Prob}\left((1,0)|\alpha',\beta\right) \\
&+\text{Prob}\left((0,0)|\alpha,\beta'\right) + \text{Prob}\left((1,1)|\alpha,\beta'\right) - \text{Prob}\left((0,1)|\alpha,\beta'\right) - \text{Prob}\left((1,0)|\alpha,\beta'\right) \\
&-\text{Prob}\left((0,0)|\alpha',\beta\right) - \text{Prob}\left((1,1)|\alpha',\beta\right) + \text{Prob}\left((0,1)|\alpha',\beta\right) + \text{Prob}\left((1,0)|\alpha',\beta\right)
\end{aligned} \tag{6.10}$$

which can still achieve its bound in magnitude arising from the rules of probability, namely four.

The standard hidden variable approach

A sufficient way to insure the restrictions from relativity are met is to restrict the possible graphical model to the local model:



(6.11)

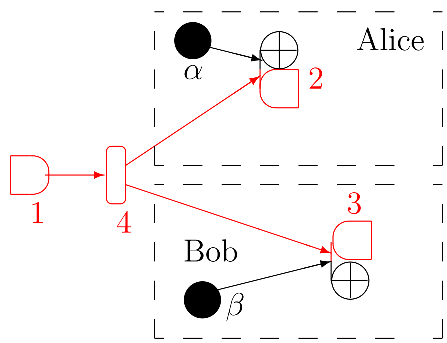
Then, if the space of values for the marginalized node is $\mathcal{X}$ and its probability measure is ρ, the joint probability both Alice and Bob get 1 is

$$\int_{x\in\mathcal{X}} \text{Prob}\left(1 \text{ for Alice}|\alpha,x\right)\text{Prob}\left(1 \text{ for Bob}|\beta,x\right)\, d\rho(x) \tag{6.12}$$

with the other joint probabilities given similarly. Using the standard approach for Bell's inequality [11], it can be shown that for the combination of probabilities in (6.10), the bound on its magnitude has now been reduced to two.

Using classical hidden nodes

A sufficient way to insure the restrictions from relativity are met using classical hidden nodes is to restrict the possible graphical model to to the local model:



(6.13)

The boxes are drawn around the two separate locations to indicate these are potentially universal modules.

Theorem 6.3.1 Employing either option **I** or **II**, Bell's inequality still holds.

Proof Using option **II**, by the Riesz theorem [57], each $\text{Prob}\,((j,k)|\alpha,\beta)$ is necessarily of the form

$$\int_{\mathcal{Y}\times\mathcal{Z}} f_j(y;\alpha)g_k(z;\alpha)\,d\mu(y,z) \tag{6.14}$$

for some positive, continuous functions f_1, f_2, g_1, g_2 with $f_1 + f_2 = 1_{\mathcal{Y}\times[0,2\pi)}$, $g_1 + g_2 = 1_{\mathcal{Z}\times[0,2\pi)}$ and some unit-norm, Radon measure μ on $\mathcal{Y}\times\mathcal{Z}$. This is mathematically the same form as (6.12), if lacking the interpretation in terms of conditional probabilities and probability measures, so the standard argument [11] still applies with bound two for (6.10).

Using option **I**, each $\text{Prob}\,((j,k)|\alpha,\beta)$ is necessarily of the form

$$(L_2 \otimes L_3)\,(\mu \times \rho \times \nu)\,(\{(j,k)\}) \tag{6.15}$$

where μ is some unit-norm measure on $[0,2\pi)$ concentrated about α from a collection of measures $\mathcal{M}_5$, ν is some unit-norm measure on $[0,2\pi)$ concentrated about β from the same collection of measures, ρ is some unit-norm measure on $\mathcal{Y}\times\mathcal{Z}$ from some collection of measures $\mathcal{M}_4$, L_2 is a norm-preserving (on the positive cone) map from measures on $[0,2\pi)\times\mathcal{Y}$ in some collection of measures $\mathcal{M}_2$ to measures on $\{0,1\}$ with set algebra the power set $\{\varnothing,\{0\},\{1\},\{0,1\}\}$, and L_3 is a norm-preserving (on the positive cone) map from measures on $\mathcal{Z}\times[0,2\pi)$ in some collection of measures $\mathcal{M}_3$ to measures on $\{0,1\}$ with set algebra the power set $\{\varnothing,\{0\},\{1\},\{0,1\}\}$. From the rules for option **I**, all the collections of measures are absolutely-continuous-complete and are such that $L_2\otimes L_3$ is well-defined. However, by **A1.2**,

A1.3, **B1.3**, **B1.4**, and **B2.5**, this implies the bound can be established just be considering $\text{Prob}\left((j,k)|\alpha,\beta\right)$ of the form

$$(L_2 \otimes L_3)\left(\sum_l \tau_l \times \sigma_l\right)(\{(j,k)\}) \tag{6.16}$$

where $\sum_l \tau_l \times \sigma_l$ is a unit-norm, finite-tensor-rank measure in $\mathcal{M}_2 \otimes \mathcal{M}_3$. Using this form for $\text{Prob}\left((j,k)|\alpha,\beta\right)$, the usual bound of two readily follows for (6.10). □

Comments

Using the well-known quantum model of Clauser, Horne, Shimony, and Holt [11] (which reuses some of the modules from the double-slit experiment (see §4.5) in a different arrangement–an example of universality), it is possible to violate Bell's inequality and the combination of probabilities in (6.10) can even achieve Tsirelson's bound of $2\sqrt{2}$ [10]. Any classical hidden model duplicating these results must be nonlocal; hence, contextual–certainly the modules we used for the classical hidden model for the double slit experiment (see §4.7) would be of no use.

We then have the question of whether Tsirelson's bound can be broken. This may be possible if there are additional mathematical structures besides measures and density matrices that can be utilized in Bayesian networks while respecting the principles of positivity and potential universality (see §3.8 above).

Part II

Parrondo's paradox and a Parrondo-like paradox

Chapter 7

Parrondo's paradox

7.1 Defining the paradox

Suppose a two-player (or one-player versus the house) repeated game wins with probability greater than one-half for one player as does a second repeated game. Naively, it may seem that the convex combination of the two games using a (possibly biased) coin to determine which rule to employ for each round must also win for that player. However, it is possible that this combined game actually wins with probability greater than one-half for the second player. This paradoxical result is termed Parrondo's paradox [49].

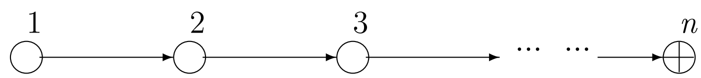
Criteria for winning

While this paradox is displayed for several different winning criteria (see [23] for instance), here the following very simple criteria will be employed for all the following games: in total, n rounds of a game with observation set $\mathcal{R}$ are played. Let $\mathcal{E}$ be the σ-algebra of observable events for each round. For a fixed set $A \in \mathcal{E}$, if the state is in A at the nth round, the first player (call her Alice) wins; otherwise, the second player (call him Bob) wins–regardless of what was observed at the preceding rounds. (For one traditional formulation of Parrondo's paradox, $\mathcal{X} = \mathbb{Z}$, $\mathcal{E}$ is its power set, and $A = \mathbb{Z}^{+}$. [49]) It is easy to see the applicability of this criterion for many different real-life situations–for instance, the difference of scores in sports games or stock options in finance.

7.2 Classical observed Markov chain game

Defining the game

Following the traditional method for Bayesian networks, the graphical model for the Markov chain game is:



(7.1)

Let μ be the initial marginal probability measure. For the subsequent nodes, take all the transition probability functions to be the same $\tau(\cdot|\cdot)$. The joint probability to observe outcomes $(x_1, x_2, \ldots, x_n) \in A_1 \times A_2 \times \cdots \times A_n$, for any $A_1, A_2, \ldots, A_n \in \mathcal{E}$, is then

$$(R_{A_n} \circ L \circ \cdots \circ R_{A_2} \circ L \circ R_{A_1} \mu)(\mathcal{X}) \tag{7.2}$$

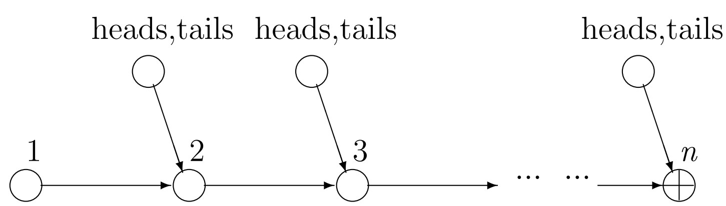
for L the map induced by $\tau(\cdot|\cdot)$ (see 2.14) and R_A the restriction map (see §2.5). Then the probability to observe the system has state in A at the nth round (so Alice wins), denoted $P_{A,n}$, is given by

$$(R_A \circ \underbrace{L \circ \cdots \circ L}_{n-1} \mu)(\mathcal{X}) \tag{7.3}$$

In the common case where the events $\mathcal{E}$ are generated by a finite number of sets, the transition probability functions are stochastic transition matrices, the initial measure is a stochastic column vector, and the maps are given by having the matrices act on the vector in the usual matrix product.

Combining two games

Now suppose we have a second game with the same structure, but a different transition probability function for each round, $\tau'(\cdot|\cdot)$ (there can also be a different initial marginal probability, but that will be of no consequence for the limit we will take). The probability to observe the system has its state in A at the nth round is denoted $P'_{A,n}$. The two games can be combined by flipping a (possibly biased) coin at each round, playing according to the first game if the coin is heads, which occurs with probability p, and according to the second if the coin is tails, which occurs with probability $1-p$. The graphical model for the combined game is:



(7.4)

For the combined game, the transition probability function for each round is

$$\tau(\cdot|\cdot,c) = \begin{cases} \tau(\cdot|\cdot) & \text{if } c = \text{heads} \\ \tau'(\cdot|\cdot) & \text{if } c = \text{tails} \end{cases} \tag{7.5}$$

Let M_c be the map induced by $\tau(\cdot|\cdot,c)$. The marginal probability to then observe the system is in A for the nth round, denoted $P^{\text{comb}}_{A,n}$, is then

$$\sum_{c_2 \in \left\{ \substack{\text{heads} \\ \text{tails}} \right\}} \cdots \sum_{c_n \in \left\{ \substack{\text{heads} \\ \text{tails}} \right\}} \left(\prod_{k=1}^{n} \text{Prob}(c_k) \right) (R_A \circ M_{c_n} \circ \cdots \circ M_{c_2} \mu)(\mathcal{X}) = (R_A \circ \underbrace{L'' \circ \cdots \circ L''}_{n-1} \mu)(\mathcal{X}) \tag{7.6}$$

for L'' the map induced by $\tau''(\cdot|\cdot) = p\tau(\cdot|\cdot) + (1-p)\tau'(\cdot|\cdot)$. Note that we have a convex combination of maps, $L'' = pL + (1-p)L'$.

Bounds on the extent of the paradox

Let p be the probability for the coin to give heads. Then, for $p \in (0,1)$, one may expect the probability for Alice to win the combined game would necessarily be between the probabilities for the individual games, with

$$\min\{P_{A,n}, P'_{A,n}\} < P^{\text{comb}}_{A,n} < \max\{P_{A,n}, P'_{A,n}\} \tag{7.7}$$

if $P_{A,n} \neq P'_{A,n}$ and $P^{\text{comb}}_{A,n} = P_{A,n} = P'_{A,n}$ if $P_{A,n} = P'_{A,n}$. Parrondo's paradox is that not only is that expectation false, but it is possible for Alice to have probability greater than one-half of winning each of the individual games, yet less than one-half for the combined game.

In the common case where the events $\mathcal{E}$ are generated by a finite number of sets, the Perron-Frobenius theorem [84] [46] gives the conditions on L for there to be a stochastic vector ν, termed the Perron-Frobenius eigenvector[1], such that the sequence $\langle L^j \nu_0 \rangle$ converges in norm to ν regardless of the initial stochastic vector ν_0. By generalization, we introduce the following definition:

Definition 7.2.1 A linear map L from $\mathcal{F}$-measures on $\mathcal{X}$ to $\mathcal{F}$-measures on a set $\mathcal{X}$ has the *Perron-Frobenius property*, $\mathcal{PF}(\mathcal{X},\mathcal{E})$, if there is a measure ν, termed the *Perron-Frobenius eigenvector*, such that the sequence $\langle L^j \nu_0 \rangle$ converges in total-variation norm to ν regardless of the initial unit-norm measure ν_0.

Now assume the transition probability functions τ and τ' are such that induced maps L and L' each are in $\mathcal{PF}(\mathcal{X};\mathcal{E})$, as well is their convex combination, $pL + (1-p)L'$. Then let P_A, P'_A, and P^{comb}_A be the limits as $n \to \infty$ of $P_{A,n}$, $P'_{A,n}$ and $P^{\text{comb}}_{A,n}$ respectively. We then have the question of what values of $(P_A, P'_A, P^{\text{comb}}_A) \in [0,1]^{\times 3}$ occur.

[1] Also termed the *invariant probability distribution* [20], the *steady-state probability vector* [39], or the *invariant measure* [47] in the literature.

Definition 7.2.2 The *classically allowed region*, denoted $\mathcal{C}(A,\mathcal{X},\mathcal{E},p)$, is the set of all $(P_A, P'_A, P_A^{\text{comb}}) \in [0,1]^{\times 3}$ that occur, for fixed A, $\mathcal{X}$, $\mathcal{E}$, and p, over all possible transition probability functions τ and τ'.

The existence of the paradox is then determined by the intersection of $\mathcal{C}$ with the cube $\left(\frac{1}{2},1\right] \times \left(\frac{1}{2},1\right] \times \left[0,\frac{1}{2}\right)$. For $p = 1$, $P_A^{\text{comb}} = P_A$ and for $p = 0$, $P_A^{\text{comb}} = P'_A$; clearly no paradox is possible. We have the following theorems for $p \in (0,1)$, which indicate that there are basically only two possibilities, depending on whether we have effectively a two-state system or one with more than two states, as determined by the event σ-algebra $\mathcal{E}$.

Theorem 7.2.3 For $p \in (0,1)$, if there are nonempty, disjoint subsets $A, B, C \in \mathcal{E}$ with $A \cup B \cup C = \mathcal{X}$, then $\mathcal{C} \supset (0,1)^{\times 3}$.

Proof Fix any $P_A, P'_A \in (0,1)$. Then take any $\varepsilon > 0$ sufficiently small such that both the matrices

$$\mathbf{T} = \begin{bmatrix} 1-(1-s)\varepsilon - \left(\frac{1}{P_A} - 1 - \varepsilon\right)\varepsilon^2 & \varepsilon^2 & 1-s \\ \left(\frac{1}{P_A} - 1 - \varepsilon\right)\varepsilon^2 & 1 - \frac{P_A s\varepsilon}{1-(1+\varepsilon)P_A} - \varepsilon^2 & s \\ (1-s)\varepsilon & \frac{P_A s\varepsilon}{1-(1+\varepsilon)P_A} & 0 \end{bmatrix} \tag{7.8}$$

$$\mathbf{T}' = \begin{bmatrix} 1 - \frac{1-P'_A}{P'_A}\varepsilon^2 & \varepsilon^2 & \frac{1}{2} \\ \frac{1-P'_A}{P'_A}\varepsilon^2 & 1-\varepsilon^2 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

are stochastic matrices for all values of $s \in [0,1]$. Take the transition probability functions to be

$$\begin{bmatrix} \tau(A|x \in A) & \tau(A|x \in B) & \tau(A|x \in C) \\ \tau(B|x \in A) & \tau(B|x \in B) & \tau(B|x \in C) \\ \tau(C|x \in A) & \tau(C|x \in B) & \tau(C|x \in C) \end{bmatrix} = \mathbf{T} \tag{7.9}$$

$$\begin{bmatrix} \tau'(A|x \in A) & \tau'(A|x \in B) & \tau'(A|x \in C) \\ \tau'(B|x \in A) & \tau'(B|x \in B) & \tau'(B|x \in C) \\ \tau'(C|x \in A) & \tau'(C|x \in B) & \tau'(C|x \in C) \end{bmatrix} = \mathbf{T}'$$

Then the required calculations reduce to matrix products. Since the maps are assumed to have the Perron-Frobenius property, the initial probabilities chosen for A, B, and C are irrelevant.

By the deliberate construction of $\mathbf{T}$ and $\mathbf{T}'$, the first entry of the Perron-Frobenius eigenvector of $\mathbf{T}$ is indeed P_A for every $s \in [0,1]$ and the first entry of the Perron-Frobenius eigenvector of $\mathbf{T}'$ is indeed P'_A. The first entry of the Perron-Frobenius eigenvector of $p\mathbf{T} +$

$(1-p)\mathbf{T}'$, which gives P_A^{comb}, varies continuously[2] from

$$\frac{\varepsilon}{\frac{p(1-p)}{2} + \varepsilon\left(\frac{p}{P_A} + \frac{1-p}{P'_A}\right)} = \mathcal{O}(\varepsilon) \tag{7.10}$$

to

$$\frac{1}{1 + \frac{2\varepsilon\left(\frac{1}{P_A} - 1 - \varepsilon + \varepsilon p\right)\left(\frac{p}{P_A} + \frac{1-p}{P'_A} - 1 - \varepsilon p\right)}{p(1-p) + 2\varepsilon\left(\frac{1}{P_A} - 1 - \varepsilon\right)}} = 1 - \mathcal{O}(\varepsilon) \tag{7.11}$$

as s goes from zero to one. Since ε may be taken arbitrarily small, every value for $P_A^{\text{comb}} \in (0,1)$ is in the allowed region for the fixed values of P_A and P'_A. □

Theorem 7.2.4 If $\mathcal{E} = \{\varnothing, A, \tilde{A}, \mathcal{X}\}$, then $\mathcal{C}$ is given by (7.7), so the paradox cannot occur.

Proof Define the stochastic matrix $\mathbf{S}$ by

$$\mathbf{S} = \begin{bmatrix} \tau(A\,|\,x \in A) & \tau(A\,|\,x \in \tilde{A}) \\ \tau(\tilde{A}\,|\,x \in A) & \tau(\tilde{A}\,|\,x \in \tilde{A}) \end{bmatrix} \tag{7.12}$$

and the stochastic matrix $\mathbf{S}'$ by

$$\mathbf{S}' = \begin{bmatrix} \tau'(A\,|\,x \in A) & \tau'(A|x \in \tilde{A}) \\ \tau'(\tilde{A}\,|\,x \in A) & \tau'(\tilde{A}\,|\,x \in \tilde{A}) \end{bmatrix} \tag{7.13}$$

For fixed $P_A, P'_A \in (0,1]$, express $\mathbf{S}$ and $\mathbf{S}'$ as

$$\mathbf{S} = \begin{bmatrix} 1 - \left(\frac{1}{P_A} - 1\right)\zeta & \zeta \\ \left(\frac{1}{P_A} - 1\right)\zeta & 1 - \zeta \end{bmatrix}, \mathbf{S}' = \begin{bmatrix} 1 - \left(\frac{1}{P'_A} - 1\right)\xi & \xi \\ \left(\frac{1}{P'_A} - 1\right)\xi & 1 - \xi \end{bmatrix} \tag{7.14}$$

for $\zeta \in \left(0, \min\{\frac{P_A}{1-P_A}, 1\}\right]$ and $\xi \in \left(0, \min\{\frac{P'_A}{1-P'_A}, 1\}\right]$. By the deliberate construction of $\mathbf{S}$ and $\mathbf{S}'$, the first entry of the Perron-Frobenius eigenvector of $\mathbf{S}$ is indeed P_A for every ζ and the first entry of the Perron-Frobenius eigenvector of $\mathbf{S}'$ is indeed P'_A for every ξ. For $P_A \neq P'_A$, the first entry of the Perron-Frobenius eigenvector of $p\mathbf{S} + (1-p)\mathbf{S}'$, which gives P_A^{comb}, varies continuously in the interval $(\min\{P_A, P'_A\}, \max\{P_A, P'_A\})$ as ζ, ξ vary over their allowed values, approaching the extreme values only as either ζ or ξ go to zero since it is a linear rational function in ζ and ξ. For $P_A = P'_A$, P_A^{comb} shares this common value for all ζ, ξ.

For $P_A = 0$, $P'_A \in (0,1]$, express $\mathbf{S}$ as $\begin{bmatrix} 1-\zeta & 0 \\ \zeta & 1 \end{bmatrix}$ for $\zeta \in (0,1]$ and $\mathbf{S}'$ as above. Then the first entry of the Perron-Frobenius eigenvector of $\mathbf{S}$ is $P_A = 0$ for every ζ. The first entry

[2] The eigenvectors of a matrix are continuous functions of the entries of the matrix in any open set where the eigenvalues are distinct.

of the Perron-Frobenius eigenvector of $p\mathbf{S} + (1-p)\mathbf{S}'$, which gives P_A^{comb}, varies continuously in the interval $(0, P'_A)$ as ζ, ξ vary over their allowed values, approaching the bounds as either ζ or ξ go to zero. The symmetric situation holds for $P'_A = 0$, $P_A \in (0,1]$. Finally, for $P_A = P'_A = 0$, express $\mathbf{S}$ as $\begin{bmatrix} 1-\zeta & 0 \\ \zeta & 1 \end{bmatrix}$ for $\zeta \in (0,1]$ and $\mathbf{S}'$ as $\begin{bmatrix} 1-\xi & 0 \\ \xi & 1 \end{bmatrix}$ for $\xi \in (0,1]$. Then $P_A^{\text{comb}} = 0$ for all ζ, ξ. □

Understanding the cause of the paradox

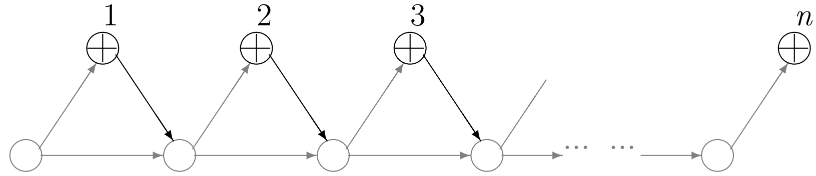
While surprising when approached from the game context, mathematically the cause of the paradox is straightforward: the coefficients of the Perron-Frobenius eigenvector of a convex combination of two matrices, $M_1, M_2 \to tM_1 + (1-t)M_2$, need not be in the convex hull of the coefficients of the Perron-Frobenius eigenvectors of each matrix. Large discrepancies are possible for certain slivers of parameter values near where eigenvalues coalesce and the resulting eigenvectors are discontinuous functions of the parameters.

A similar cause is at play in the following analogous hidden classic and quantum versions of the preceding game. In each case, there is a convex space of matrices of spectral radius one determining the behavior of the system. A coin flip determining the rule to employ for the next round has the effect of forming a convex combination of the matrices for each rule. In the limit as the number of rounds goes to infinity, only the eigenvector of the matrix with eigenvalue one is of importance, but the coefficients for this eigenvector are not in the convex hull of the coefficients of the eigenvectors of each matrix, leading to the paradox.

7.3 Classical hidden-Markov chain game

Defining the combined game

As will be proved below, it is possible to have Parrondo's paradox with a game for a system with two states if the system is hidden. The game arises from a generalization of the of the commonly utilized hidden-Markov model [29] to include imagined back-reaction on the hypothesized hidden state. The standard graphical model (with hidden nodes in gray) is:



(7.15)

However, as has been mentioned in §2.2, this requires interpreting probabilities and/or conditional probabilities concerning unobservable events, which are meaningless in both Bayesian and frequentist interpretations of probability. Also, this approach has no analogy in the quantum case.

Therefore, following the program we have laid out in the first part of this dissertation, the information for hidden nodes can simply be regarded as linear maps on measures, which is preferable in both avoiding metaphysical notions and in having analogous meaning in the quantum case. (As has been mentioned in §2.2, this may allow additional maps which do not arise from transition probability functions, but these are not of great interest and are not an important motivation for making the change.) Using the pince-nez construction (see §4.1), this becomes:

(7.16)

Employing option **I'** (the results can be readily recast for option **II** or more general forms of option **I**–see §3.2), we have the following definition:

Definition 7.3.1 A *hidden-Markov-with-back-reaction-model* ($\mathcal{R}$, τ, $\mathcal{H}$, μ, L, ν_0) is composed of: an observation set $\mathcal{R}$ with base measure η, a hidden set $\mathcal{H}$ with base measure μ, a pince-nez map $L \in \mathcal{B}\left(L^1(\mathcal{H};\mu), L^1(\mathcal{R}\times\mathcal{H};\eta\times\mu)\right)$, and an initial, unit-norm measure ν_0 on $\mathcal{H}$ that is absolutely continuous with respect to μ.

Then the joint probability to observe $(x_1, x_2, \ldots, x_n) \in A_1 \times A_2 \times \cdots \times A_n$, for any η-measurable $A_1, A_2, \ldots, A_n$, is

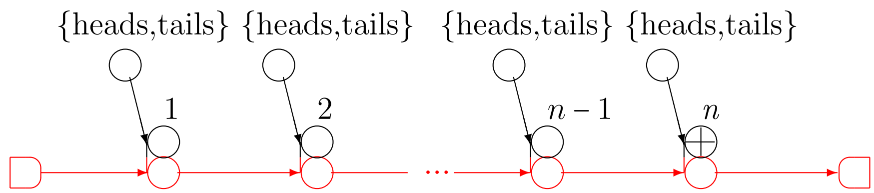
$$\int_{\mathcal{H}} \left(S_{A_n} \circ L \circ \cdots \circ S_{A_1} \circ L \frac{d\nu_0}{d\mu} \right) d\mu \tag{7.17}$$

for S_A the linear map from $L^1(\mathcal{R}\times\mathcal{H};\eta\times\mu)$ to $L^1(\mathcal{H};\mu)$ given by the partial integration $S_A f = \int_A f \, d\eta$. Employing this model for the game, then the marginal probability for the observation to be in A at the nth round (so Alice wins) is

$$P_{A,n} = \int_{\mathcal{H}} \left(S_A \circ L \circ \underbrace{N \circ \cdots \circ N}_{n-1} \frac{d\nu_0}{d\mu} \right) d\mu \tag{7.18}$$

where $N \in \mathcal{B}\left(L^1(\mathcal{H};\mu)\right)$ is defined as the map $N = S_{\mathcal{R}} \circ L$. Because of the definition of the game, it is possible to reduce to the case where the event space for $\mathcal{R}$ is simply $\mathcal{E} = \{\varnothing, A, \tilde{A}, \mathcal{X}\}$ and its base measure η is the measure with $\eta(A) = \eta(\tilde{A}) = 1$.

Now suppose there is a second model with the same data except for a different pince-nez L' for each round (it is also possible to have a different initial measure, but since the limit as $n \to \infty$ will be taken, this is of no consequence). Once again, the two models can be combined by flipping a (possibly biased) coin at each round. The graphical model is:



(7.19)

Let γ be the previously given probability measure for the coin of p for heads and $1-p$ for tails, with base measure the counting measure. For simplicity of notation, we will identify γ with its Radon-Nikodým derivative $\frac{d\gamma}{d\text{count}}$. For the combined game, the map for each round is

$$M(f \otimes \gamma) = p\,Lf + (1-p)L'f = L''f \tag{7.20}$$

for $L'' = p\,L + (1-p)L'$. Let $N', N'' \in \mathcal{B}(L^1(\mathcal{H};\mu))$ be defined using L', L'' as N was following (7.18), $N' = S_{\mathcal{R}} \circ L'$ and $N'' = S_{\mathcal{R}} \circ L''$. Then the marginal probability for the observation to be in A for the nth round (so Alice wins) is

$$\begin{aligned} P_n^{\text{comb}} &= \int_{\mathcal{H}} S_A \circ M\left(\cdots\left(S_{\mathcal{R}} \circ M\left(\left(S_{\mathcal{R}} \circ M\frac{d\nu_0}{d\mu}\right)\otimes\gamma\right)\right)\otimes\gamma\cdots\right)d\mu \\ &= \int_{\mathcal{H}}\left(S_A \circ L'' \circ \underbrace{N'' \circ \cdots \circ N''}_{n-1}\frac{d\nu_0}{d\mu}\right)d\mu \end{aligned} \tag{7.21}$$

Bounds on the extent of the paradox

One may expect the paradox to be present to an equal or greater degree for this more general class of games. This is indeed the case, as is seen by comparing the statements of theorem **7.2.3** and **7.2.4** with the following theorem **7.3.3** and **7.3.4**. Restrict the space of pince-nez maps to those where $N, N', N'' = pN + (1-p)N'$ all have the Perron-Frobenius property, where $N, N', N'' \in \mathcal{B}(L^1(\mathcal{H};\mu))$ are defined from L, L', L'' as above. Then, once again, let P_A, P'_A, and P_A^{comb} be the limits as $n \to \infty$ of $P_{A,n}$, $P'_{A,n}$ and $P_{A,n}^{\text{comb}}$ respectively.

Definition 7.3.2 The *classical-hidden allowed region*, denoted $\mathcal{CH}(A, \mathcal{R}, \eta, \mathcal{H}, \mu, p)$, is the set of all $(P_A, P'_A, P_A^{\text{comb}}) \in [0,1]^{\times 3}$ that occur, for fixed A, $\mathcal{R}$, τ, $\mathcal{H}$, μ, and p, over all allowed pince-nez maps L and L'.

As in the classical observed case, the existence of the paradox is then determined by the intersection of $\mathcal{CH}$ with the cube $\left(\frac{1}{2}, 1\right] \times \left(\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right)$. For $p = 1$, $P_A^{\text{comb}} = P_A$ and for $p = 0$, $P_A^{\text{comb}} = P'_A$; clearly no paradox is possible. We have the following theorems for $p \in (0,1)$, which indicate that, once again, there are basically only two possibilities, depending on whether we have effectively a two-state hidden system or one with more than two states, as determined by the hidden base measure μ.

Theorem 7.3.3 For $p \in (0,1)$, if there are disjoint, μ-measurable subsets B, C, D with $B \cup C \cup D = \mathcal{H}$ and $\mu(B)$, $\mu(C)$, and $\mu(D)$ all strictly positive, then $\mathcal{CH} \supset (0,1)^{\times 3}$.

Proof Pick any $\varepsilon > 0$ and let κ be the map from $L^1(\mathcal{H};\mu)$ to $\mathbb{R}^3$ given by

$$\kappa(f) = \begin{bmatrix} \int_B f\, d\mu \\ \int_C f\, d\mu \\ \int_D f\, d\mu \end{bmatrix} \tag{7.22}$$

Let h_B, h_C, and h_D be unit-norm functions in $L^1(B;\mu)^+$, $L^1(C;\mu)^+$, and $L^1(D;\mu)^+$ respectively, each extended to all of $\mathcal{H}$ by zero. Such functions certainly exist by the σ-finiteness of μ. Then define the pince-nez maps L and L' by

$$\begin{aligned} Lf &= 1_A \otimes \left(\begin{bmatrix} h_B & 0 & 0 \end{bmatrix} \mathbf{T}\kappa(f) \right) + 1_{\tilde{A}} \otimes \left(\begin{bmatrix} 0 & h_C & h_D \end{bmatrix} \mathbf{T}\kappa(f) \right) \\ L'f &= 1_A \otimes \left(\begin{bmatrix} h_B & 0 & 0 \end{bmatrix} \mathbf{T}'\kappa(f) \right) + 1_{\tilde{A}} \otimes \left(\begin{bmatrix} 0 & h_C & h_D \end{bmatrix} \mathbf{T}'\kappa(f) \right) \end{aligned} \tag{7.23}$$

for $\mathbf{T}, \mathbf{T}'$ as in theorem **7.2.3**. Because of the properties of B, C, D, h_B, h_C, h_D, and κ, the required calculation reduces to matrix products.

Then, by the deliberate construction of the maps, the row vector

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \mathbf{T} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{7.24}$$

acting on the Perron-Frobenius eigenvector of $\mathbf{T}$ is its the first entry, which is indeed P_A for every $s \in [0,1]$. Similarly, $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \mathbf{T}'$ acting on the Perron-Frobenius eigenvector of $\mathbf{T}'$ is its first entry, which is indeed P'_A. For the combined game, P_A^{comb} is given by the row vector

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} (p\mathbf{T} + (1-p)\mathbf{T}') \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{7.25}$$

acting on the Perron-Frobenius eigenvector of $p\mathbf{T} + (1-p)\mathbf{T}'$, which is once again its first entry. As in theorem **7.2.3**, this varies continuously from $0 + \mathcal{O}(\varepsilon)$ to $1 - \mathcal{O}(\varepsilon)$ as ζ goes from zero to one. Since ε may be taken arbitrarily small, every value for $P_A^{\text{comb}} \in (0,1)$ is in the allowed region for the fixed values of P_A and P'_A. □

Theorem 7.3.4 For $p \in (0,1)$,if the base measure μ is such that there is a μ-measurable subset B with $\mu(B)$ and $\mu(\tilde{B})$ both strictly positive, but the conditions of the preceding theorem do not hold, then $\mathcal{CH}(p)$ is given by the relatively open region

$$\min\{pP_A, (1-p)P'_A\} < P_A^{\text{comb}} < \max\{1 - p + pP_A, p + (1-p)P'_A\}$$

Proof Represent measures ν on $\mathcal{H}$ absolutely continuous with respect to μ by column vectors with column sum one, $\begin{bmatrix} \nu(B) \\ \nu(\tilde{B}) \end{bmatrix}$. Then all the linear maps from $L^1(\mathcal{H};\mu)$ to $L^1(\mathcal{H};\mu)$ are represented by 2×2-matrices acting by the usual matrix product.

Firstly, we will show that $\mathcal{CH}(p)$ contains the given region. Fix any $P_A, P'_A \in (0,1]$. Then take any $\varepsilon > 0$ sufficiently small such that the matrix

$$N = \begin{bmatrix} 1-\left(\frac{1}{P_A}-1\right)\varepsilon & \varepsilon \\ \left(\frac{1}{P_A}-1\right)\varepsilon & 1-\varepsilon \end{bmatrix} \tag{7.26}$$

is stochastic. Let

$$L = 1_A \otimes \begin{bmatrix} 1-\left(\frac{1}{P_A}-1\right)\varepsilon & 0 \\ \left(\frac{1}{P_A}-1\right)\varepsilon & 0 \end{bmatrix} + 1_{\tilde{A}} \otimes \begin{bmatrix} 0 & \varepsilon \\ 0 & 1-\varepsilon \end{bmatrix} \tag{7.27}$$

and, for $s \in [0,1]$,

$$N' = \begin{bmatrix} \frac{1}{2}(1+s) & \frac{1}{2}s \\ \frac{1}{2}(1-s) & 1-\frac{1}{2}s \end{bmatrix}, L' = P'_A 1_A \otimes N' + (1-P'_A)1_{\tilde{A}} \otimes N' \tag{7.28}$$

Then $\begin{bmatrix} 1 & 1 \end{bmatrix} S_A \circ L = \begin{bmatrix} 1 & 0 \end{bmatrix}$ acting on the Perron-Frobenius eigenvector of N is its first entry, which is indeed P_A by the deliberate construction of the maps. Similarly, $\begin{bmatrix} 1 & 1 \end{bmatrix} S_A \circ L' = P'_A \begin{bmatrix} 1 & 1 \end{bmatrix}$ acting on the Perron-Frobenius eigenvector of N' is indeed P'_A for all s. For the combined game, P_A^{comb} is given by

$$\begin{bmatrix} 1 & 1 \end{bmatrix} S_A \circ (pL + (1-p)L') = p\begin{bmatrix} 1 & 0 \end{bmatrix} + (1-p)P'_A \begin{bmatrix} 1 & 1 \end{bmatrix} \tag{7.29}$$

acting on the Perron-Frobenius eigenvector of $pN + (1-p)N'$. This varies continuously from $(1-p)P'_A + \mathcal{O}(\varepsilon)$ to $p + (1-p)P'_A - \mathcal{O}(\varepsilon)$ as s goes from zero to one. Since ε may be taken arbitrarily small, every value in the interval $((1-p)P'_A, p+(1-p)P'_A)$ can be taken by P_A^{comb} for the fixed values of P_A and P'_A. Repeating the process with primed↔unprimed and $p \leftrightarrow 1-p$, every value in the interval $(pP_A, 1-p+pP_A)$ can be taken by P_A^{comb} for the fixed values of P_A and P'_A. A similar argument using appropriately modified maps holds for either P_A or P'_A or both of them being zero.

Now, we will show that $\mathcal{CH}(p)$ is contained in the given region. A key observation is that instead of varying over all maps $S_A \circ L$ and $S_A \circ L$ with positive entries subject to the constraint that $N = S_{\mathcal{R}} \circ L$ is stochastic, one may equivalently vary over all stochastic maps N and row vectors $\mathbf{v}$ with entries in $[0,1]$ (with $\mathbf{v} = \begin{bmatrix} 1 & 1 \end{bmatrix} S \circ L$). The same situation holds for the primed maps. Write $\mathbf{v} = \begin{bmatrix} a & b \end{bmatrix}$, $N = \begin{bmatrix} 1-c & d \\ c & 1-d \end{bmatrix}$, $\mathbf{v}' = \begin{bmatrix} a' & b' \end{bmatrix}$, and $N' = \begin{bmatrix} 1-c' & d' \\ c' & 1-d' \end{bmatrix}$, with $a,b,c,d,a',b',c',d' \in [0,1]$ except $c = d = 0$, $c = d = 1$, $c' = d' = 0$, and $c' = d' = 1$ are not allowed since the maps must have the Perron-Frobenius property.

Fixing the values of P_A and P'_A and using explicit expressions for the Perron-Frobenius eigenvector gives the optimization problem:

$$\text{constraints: } ad + bc = P_A(c+d), a'd' + b'c' = P'_A(c'+d') \tag{7.30}$$
$$\text{extremize: } P_A^{\text{comb}} = \frac{(pa + (1-p)a')(pd + (1-p)d') + (pb + (1-p)b')(pc + (1-p)c')}{p(c+d) + (1-p)(c'+d')}$$

The quantity to be extremized is a rational linear function of c, d, c', d'; furthermore, the constraints are homogeneous in the sense that they are unaffected by $c \to \xi c$, $d \to \xi d$ for any positive ξ and $c' \to \xi' c'$, $d' \to \xi' d'$ for any positive ξ'. Therefore, the extreme values can only be attained as either $c, d \to 0$ or $c', d' \to 0$. In the first case, using the remaining condition, the quantity to be extremized is $(1-p)P'_A + \dfrac{p(ad' + bc')}{(c'+d')}$. The term on the right is between 0 and p. In the second case, using the remaining condition, the quantity to be extremized is $pP_A + \dfrac{(1-p)(a'd + b'c)}{(c+d)}$. The term on the right is between 0 and $1-p$. The actual extreme values cannot be attained because the maps corresponding to those parameter values are not allowed. □

7.4 Defining a quantum analogue of the Markov model

An analogous game

As we showed in §3.3, one may consider the analogous game to the above one employing maps on density matrices rather than maps on measures. This leads to the following Bayesian network model:

Definition 7.4.1 A *Markov-Bayesian-quantum model* ($\mathcal{R}$, η, H, L, ρ_0) is composed of: an observation set $\mathcal{R}$ with base measure η, a Hilbert space H, a pince-nez map $L \in \mathcal{B}(\mathcal{D}(\mathsf{H}), L^1(\mathcal{R}; \eta; \mathcal{D}(\mathsf{H})))$ that is norm-preserving (on the positive cone) and completely-positive, and an initial density-matrix ρ_0 on H with tr $\rho_0 = 1$.

The use of the word *quantum* will be justified below. The accompanying graphical model is (note it is the same as (7.16), illustrating the analogy graphically):

(7.31)

Then the joint probability to observe $(x_1, x_2, \ldots, x_n) \in A_1 \times A_2 \times \cdots \times A_n$, for any η-measurable $A_1, A_2, \ldots, A_n \in \mathcal{R}$, is (compare to (7.17))

$$\text{tr } S_{A_n} \circ L \circ \cdots \circ S_{A_1} \circ L\rho_0 \tag{7.32}$$

for $S_A \in \mathcal{B}(L^1(\mathcal{R};\eta;\mathcal{D}(\mathsf{H})),\mathcal{D}(\mathsf{H}))$ the linear map given by the integration $S_A\varphi = \int_A \varphi\, d\eta$.

Employing this model for the game, then the marginal probability for the observation to be in A at the nth round (so Alice wins) is (compare to (7.18))

$$P_{A,n} = \text{tr } S_A \circ L \circ \underbrace{N \circ \cdots \circ N}_{n-1} \rho_0 \tag{7.33}$$

where $N \in \mathcal{B}(\mathcal{D}(\mathsf{H}))$ is defined as the map $S_{\mathcal{R}} \circ L$. Because of the definition of the game, it is possible to reduce to the case where the event space for $\mathcal{R}$ is simply $\mathcal{E} = \{\varnothing, A, \tilde{A}, \mathcal{X}\}$ and its base measure η is the measure with $\eta(A) = \eta(\tilde{A}) = 1$.

An alternate game using textbook quantum mechanics

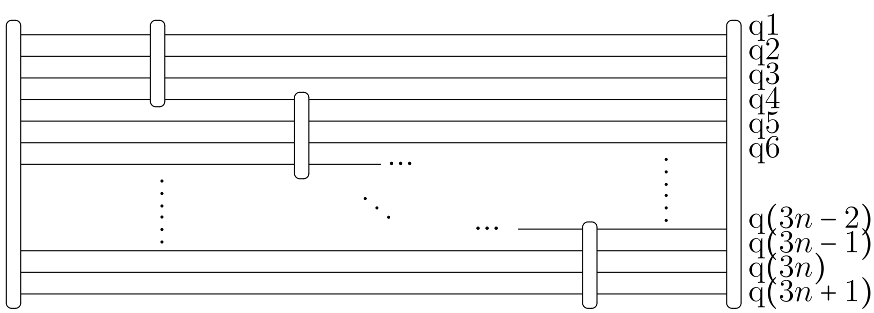
Consider the following quantum game that is a generalization of that employed by [23], but using our simpler winning criterion.

Definition 7.4.2 A *Markov-quantum process* (H, q, U) with: Hilbert space H, number of *quantum channels* q, and unitary operator $U : \mathsf{H}^{\otimes q} \to \mathsf{H}^{\otimes q}$–is the operator

$$\psi \to (I_{\mathsf{H}^{\otimes(q-1)(n-1)}} \otimes U) \cdots (I_{\mathsf{H}^{\otimes(q-1)}} \otimes U \otimes I_{\mathsf{H}^{\otimes(q-1)(n-2)}}) (U \otimes I_{\mathsf{H}^{\otimes(q-1)(n-1)}}) \psi$$

for any $\psi \in \mathsf{H}^{\otimes((q-1)n+1)}$.

The corresponding quantum circuit for the case of $q = 4$ is:



(7.34)

This may be compared to the quantum process with accompanying quantum circuit employed in [23].

Let $\{P_{A_1\times\cdots\times A_n}\}$ over all τ-measurable $A_1, \ldots, A_n$ be a complete set of mutually commuting, orthogonal projectors[3] on $\mathsf{H}^{\otimes((q-1)n+1)}$.

A textbook quantum model is then provided by the following:

[3] A complete set of mutually commuting, orthogonal projectors relative to a base measure μ is defined similarly with that relative to a σ-algebra except the condition $E_\varnothing = 0$ is replaced by the condition that $\mu(B) = 0 \Rightarrow E_B = 0$.

Definition 7.4.3 A *Markov-quantum-model* ($\mathcal{R}$, τ, H, q, U, $\{P_{A_1\times\cdots\times A_n}\}$, ψ_i) is composed of: an observation set $\mathcal{R}$ with base measure τ, a Markov-quantum process (H, q, U), a collection of orthogonal projectors $\{P_{A_1\times\cdots\times A_n}\}$ as described above, and an initial length-one vector $\psi_i \in \mathsf{H}^{\otimes((q-1)n+1)}$.

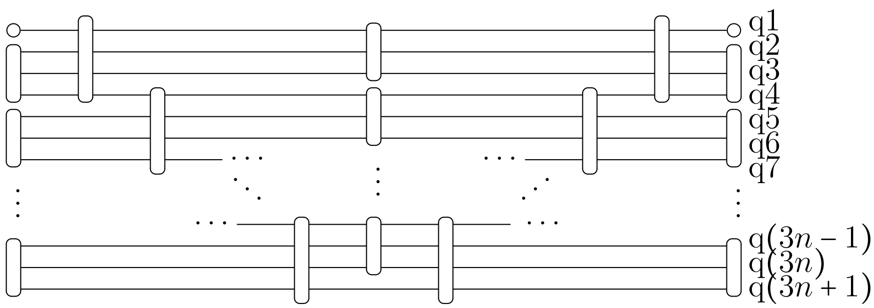
The joint probability to observe $(x_1, x_2, \ldots, x_n) \in A_1\times A_2\times\cdots\times A_n$ is then $\|P_{A_1\times\cdots\times A_n}\psi_f\|^2$, where ψ_f is given by applying the Markov-quantum process to ψ_i. In particular, the marginal probability for the observation to be in A at the nth round (so Alice wins) is $P_{A,n} = \|P_{\mathcal{X}\times\cdots\times\mathcal{X}\times A}\psi_f\|^2$.

Equivalence of the two models

Start with any Markov-quantum model. The joint probability to observe $(x_1, x_2, \ldots, x_n)$ in $A_1 \times\cdots\times A_n$ is given by

$$\|P_{A_1\times\cdots\times A_n}\psi_f\|^2 = \langle P_{A_1\times\cdots\times A_n}\psi_f, \psi_f\rangle = \text{tr } P_{A_1\times\cdots\times A_n}\psi_f\psi_f^* \tag{7.35}$$

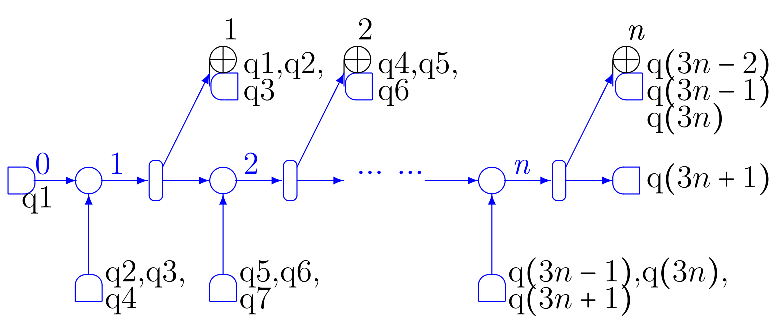
This has a graphical representation as the following tensor network (illustrated for $q = 4$):



(7.36)

The initial wavefunction $\psi_i = \psi_0 \otimes (\psi_1)^{\otimes n}$ is on the left, with its conjugate on the right. The unitary operators U are on the left diagonal, with the adjoints U^* on the right diagonal. The projectors are down the center.

Rather than performing the calculation in the difficult manner (left-to-right), the calculation can be done the easy way (top-to-bottom). This reordering gives rise to the following graphical model for a Bayesian network, where the labels indicate where the quantum channels from the quantum circuit and tensor network models go in the latter:



(7.37)

By replacing each repeated module by a pince-nez, this becomes (7.31). (Note the efficiency of the graphical model for the Bayesian network compared to the tensor network–there is no need to place every operator and initial condition twice in the graph.) The preceding graphical manipulation inspires the following:

Theorem 7.4.4 If a Markov-quantum model is of the form

$$(\mathcal{R}, \tau, \mathsf{H}, q, U, \{Q_{A_1} \otimes \cdots \otimes Q_{A_n} \otimes I_{\mathsf{H}}\}, \psi_0 \otimes (\psi_1^{\otimes n}))$$

for some unit-length vectors $\psi_0 \in \mathsf{H}$ and $\psi_1 \in \mathsf{H}^{\otimes(q-1)}$ and collection of orthogonal projectors $\{Q_C\}_{C \in \mathcal{E}}$ on $\mathsf{H}^{\otimes(q-1)}$, then there is an equivalent (in the sense of giving the same joint probabilities) Markov-Bayesian-quantum model ($\mathcal{R}$, τ, H, L, ρ_0).

Proof Start with any such Markov-quantum model. The joint probability to observe $(x_1, x_2, \ldots, x_n)$ in $A_1 \times \cdots \times A_n$ is given by

$$\|P_{A_1 \times \cdots \times A_n} \psi_f\|^2 = \langle P_{A_1 \times \cdots \times A_n} \psi_f, \psi_f \rangle = \text{tr } P_{A_1 \times \cdots \times A_n} \psi_f \psi_f^* \tag{7.38}$$

Write the trace as the tensor product of reduced traces,

$$\text{tr}_{\mathsf{H}^{\otimes((q-1)n+1)}} = (\text{tr}_{\mathsf{H}^{\otimes(q-1)}})^{\otimes n} \otimes \text{tr}_{\mathsf{H}} \tag{7.39}$$

This tensor product of maps is well-defined using **A1.3**, **B4.1**, **B5.5**, and **B5.8**. Using the given forms, $P_{A_1 \times \cdots \times A_n} = Q_{A_1} \otimes \cdots \otimes Q_{A_n} \otimes I_{\mathsf{H}}$ and $\psi_i = \psi_0 \otimes (\psi_1^{\otimes n})$, then "pushing" the partial traces into the expression as far as possible gives the joint probability as

$$\begin{gathered}(K(\cdot; A_n) \otimes \text{tr}_{\mathsf{H}}) \big(U \big(\big(\big(K(\cdot; A_{n-1}) \otimes I_{\mathcal{B}(\mathsf{H})}\big) (\cdots \\ \big(K(\cdot; A_2) \otimes I_{\mathcal{B}(\mathsf{H})}\big)\big(U\big(\big(\big(K(\cdot; A_1) \otimes I_{\mathcal{B}(\mathsf{H})}\big) (U(\psi_0\psi_0^* \otimes (\psi_1\psi_1^*))U^*)\big) \otimes (\psi_1\psi_1^*)\big)U^*\big) \\ \cdots)) \otimes (\psi_1\psi_1^*))\, U^*)\end{gathered} \tag{7.40}$$

where $I_{\mathcal{B}(\mathsf{H})}$ is the identity map on operators in $\mathcal{B}(\mathsf{H})$ and

$$K : \mathcal{D}\left(\mathsf{H}^{\otimes(q-1)}\right) \times \begin{pmatrix} \tau - \text{measurable} \\ \text{subsets of } \mathcal{R} \end{pmatrix} \to \mathbb{R} \tag{7.41}$$

is given by $K(\rho; C) = \text{tr } Q_C \rho$. The various tensor products of maps in the expression are still well-defined by the above propositions. K is countably-additive in its second argument by the countable-additivity of the projectors $\{Q_C\}$ (with convergence in the weak* topology). $K(\cdot; C)$ is manifestly completely-positive for any $C \in \mathcal{E}$. Then, by **B5.5**, and **B5.8**, $\|K(\cdot; C) \otimes I_{\mathcal{B}(\mathsf{H})}\|_{\text{op}} = \|K(\cdot; C)\|_{\text{op}}$. Therefore, for any fixed $\rho \in \mathcal{D}\left(\mathsf{H}^{\otimes(q-1)}\right)^+$, $K(\rho; \cdot)$ is a measure on $\mathcal{R}$. Furthermore, this measure is clearly absolutely continuous with respect to τ since $Q_C = 0$ if $\tau(C) = 0$.

Let $N : \mathcal{D}(\mathsf{H}) \times \mathcal{E} \to \mathcal{D}(\mathsf{H})$ be the map given by

$$MN\rho; C) = \left(K(\cdot; C) \otimes I_{\mathcal{B}(\mathsf{H})}\right)\left(U(\rho \otimes (\psi_1\psi_1^*))U^*\right) \tag{7.42}$$

By the above properties of K, for any fixed $\rho \in \mathcal{D}(\mathsf{H})^+$, $N(\rho;\cdot)$ is a $\mathcal{D}(\mathsf{H})^+$-valued vector measure on $\mathcal{R}$ that is absolutely continuous with respect to τ. $\mathcal{D}(\mathsf{H})$ has the Radon-Nikodým property, so $N(\rho;\cdot)$ has Radon-Nikodým derivative $\frac{dN(\rho;\cdot)}{d\tau} \in L^1(\mathcal{R};\tau;\mathcal{D}(\mathsf{H})^+)$ with the property that $N(\rho;C) = \int_C \frac{dN(\rho;\cdot)}{d\tau}\,d\tau$ for any τ-measurable subset C. Then define the pince-nez map L by $L\rho = \frac{dN(\rho;\cdot)}{d\tau}$. The required completely-positive and norm-preserving (on the positive cone) properties of L are readily demonstrated. □

More remarkable (and requiring theorem **5.1.1** to prove) is the following result:

Theorem 7.4.5 If a Markov-Bayesian-quantum model is of the form ($\mathcal{R}$, τ, H, L, $\psi_0\psi_0^*$), then there is an equivalent (in the sense of giving the same joint probabilities) Markov-quantum-model

$$(\mathcal{R}', \tau', \mathsf{H}, q, U, \{P_{A_1\times\cdots\times A_n}\}, \psi_i)$$

Proof Start with any such Markov-Bayesian-quantum model. From theorem **5.1.1** and the Radon-Nikodým property of $\mathcal{D}(\mathsf{H})$, L can be written in terms of the Radon-Nikodým derivative, $L\rho = \frac{dN(\rho;\cdot)}{d\tau}$, with

$$N(\rho; B) = \left(\mathrm{tr}_{\mathsf{K}} \otimes P(B) \otimes I_{\mathcal{B}(\mathsf{H})}\right)(V\rho V^*) \tag{7.43}$$

for some Hilbert space K, some partial isometry $V : L^2(\mathsf{H} \to \mathsf{K} \otimes L^2(\mathcal{R};\tau) \otimes \mathsf{H}$, and where $P(B)$ is the functional on $\mathcal{D}(L^2(\mathcal{R};\tau))$ given by $P(B)\tau = \int_{x\in B} \tau(x,x)\,d\nu(x)$. Furthermore, by the theorem, if all the Hilbert spaces have finite dimension, the dimension of K less than or equal to $(\dim \mathsf{H})^2$.

The dimension of $L^2(\mathcal{R};\tau)$ then determines the minimum number of quantum channels, q. If $\dim L^2(\mathcal{R};\tau) \le (\dim \mathsf{H})$, take $q = 4$. If $\dim \mathsf{H} < \dim L^2(\mathcal{R};\tau) \le (\dim \mathsf{H})^2$, take $q = 5$. If $(\dim \mathsf{H})^2 < \dim L^2(\mathcal{R};\tau) \le (\dim \mathsf{H})^3$, take $q = 6$. For any of these three cases, take Q to be any partial isometry from $L^2(\mathcal{R};\tau)$ to $\mathsf{H}^{\otimes(q-3)}$. For $(\dim \mathsf{H})^3 < \dim L^2(\mathcal{R};\tau)$, let G be the subspace of $L^2(\mathcal{R};\tau)$ which is relevant, given by the norm-closure of the span of $\{(\psi^* \otimes \pi(1) \otimes \xi^*)\,V\phi \,|\, \phi,\psi \in \mathsf{H}, \xi \in \mathsf{K}\}$. Then $\dim \mathsf{G} \le (\dim \mathsf{H})^2 \dim \mathsf{K} \le (\dim \mathsf{H})^4$, so take $q = 7$ and Q to be any partial isometry from G to $\mathsf{H}^{\otimes 4}$.

Now select any unit-length $\psi_1 \in \mathsf{H}^{\otimes(q-1)}$ and any partial isometry $R : \mathsf{K} \to \mathsf{H} \otimes \mathsf{H}$. Then there is some unitary $U : \mathsf{H}^{\otimes q} \to \mathsf{H}^{\otimes q}$ such that

$$U(\phi \otimes \psi_1) = (R \otimes Q \otimes I_{\mathsf{H}})V\phi \text{ for every } \phi \in \mathsf{H} \tag{7.44}$$

Hence, by the spectral theorem for compact operators,

$$U(\rho \otimes \psi_1\psi_1^*)U^* = (R \otimes Q \otimes I_{\mathsf{H}})V\rho V^*(R^* \otimes Q^* \otimes I_{\mathsf{H}}) \text{ for every } \rho \in \mathcal{D}(\mathsf{H})^+ \tag{7.45}$$

Take $\psi_i = \psi_0 \otimes (\psi_1^{\otimes n})$. Finally, if $\dim L^2(\mathcal{X};\mu) = (\dim \mathsf{H})^{q-3}$ or $\dim \mathsf{G} = (\dim \mathsf{H})^4$, take $\mathcal{X}' = \mathcal{X}$, $\tau' = \tau$, and $P_{A_1\times\cdots\times A_n}$ to be

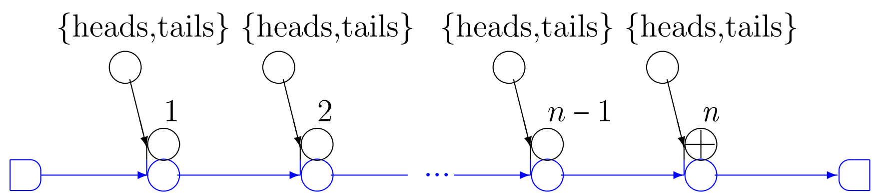
$$(I_{\mathsf{H}^{\otimes 2}} \otimes Q\pi(1_{A_1})Q^*) \otimes \cdots \otimes (I_{\mathsf{H}^{\otimes 2}} \otimes Q\pi(1_{A_n})Q^*) \otimes I_{\mathsf{H}} \tag{7.46}$$

Otherwise, add one point x_0 to $\mathcal{X}$ to form $\mathcal{X}'$. Add an atom of weight one at x_0 to τ to form τ'. For $A_1, \ldots, A_n \in \mathcal{E}$, let $P_{A_1\times\cdots\times A_n}$ be defined as above. Replacing any of the A_j's with $\{x_0\}$, replace $Q\pi(1_{A_j})Q^*$ with $I_{\mathsf{H}^{\otimes(q-3)}} - Q^*Q$. (Note that with the given initial wavefunction and time-evolution, the probabilities for observing $\{x_0\}$ are all zero–it is only added so the projectors form a complete set with $P_{\mathcal{X}'^{\times n}} = I_{\mathsf{H}^{\otimes((q-1)n+1}}$.) □

7.5 Parrondo's paradox for the quantum game

Defining the combined game

As a reward for our efforts in part **I** and in the preceding section, we can now ignore the complicated Markov-quantum model and instead utilize the far simpler Markov-Bayesian-quantum model. Furthermore, since the latter model uses maps that live on a convex space, it is possible to utilize coin flips to combine two games instead of the recourse to more complicated alternating rules for rounds previously used by [23] [41]. This combined game has graphical model (note it is the same as (7.19)):



(7.47)

Again, let γ be the previously given probability measure for the coin of p for heads and $1-p$ for tails, and, for simplicity of notation, identify γ with its Radon-Nikodým derivative with respect to the counting measure. For the combined game, the map for each round is

$$M(\rho \otimes \gamma) = p\,L\rho + (1-p)L'\rho = L''\rho \tag{7.48}$$

for $L'' = p\,L + (1-p)L'$. Let $N', N'' \in \mathcal{B}(\mathcal{D}(\mathsf{H}))$ be defined using L', L'' as N was following (7.33),as $N' = S_{\mathcal{R}} \circ L'$ and $N'' = S_{\mathcal{R}} \circ L''$. Then the marginal probability for the observation to be in A for the nth round (so Alice wins) is (compare to (7.21))

$$\begin{aligned} P_{A,n}^{\text{comb}} &= \operatorname{tr} S_A \circ M\left(\cdots\left(S_{\mathcal{R}} \circ M\left(\left(S_{\mathcal{R}} \circ M\rho_0\right) \otimes \gamma\right)\right) \otimes \gamma \cdots\right) d\mu \\ &= \operatorname{tr} S_A \circ L'' \circ \underbrace{N'' \circ \cdots \circ N''}_{n-1} \rho_0 \end{aligned} \tag{7.49}$$

Because of the definition of the game, it is possible to reduce to the case where $\mathcal{E} = \{\varnothing, A, \tilde{A}, \mathcal{X}\}$.

Bounds on the paradox for the quantum game

Since it is possible to embed a classical hidden system within a quantum one (see §3.6), one may expect the paradox to be present to an equal or greater degree for these quantum games. This is indeed the case, as is seen by comparing the statements of theorems **7.3.3** and **7.3.4** with the following theorem **7.5.3** and claim **7.5.5**. By analogy to definition **7.2.1**, we have:

Definition 7.5.1 A linear map L from $\mathcal{D}(\mathsf{H})^+$ to $\mathcal{D}(\mathsf{H})^+$ has the *quantum-Perron-Frobenius property*, $\mathcal{QPF}(\mathsf{H})$, if there is a density matrix ρ, termed the *quantum-Perron-Frobenius eigenvector*, such that the sequence $\langle L^j \rho_0 \rangle$ converges in total-variation norm to ρ regardless of the initial unit-norm density matrix ρ_0.

Restrict the space of pince-nez maps L, L' to those where N ,N', and their convex combination, $pN + (1-p)N'$, are all in $\mathcal{QPF}(\mathsf{H})$. Then, once again, let P_A, P'_A, and P_A^{comb} be the limits as $n \to \infty$ of $P_{A,n}$, $P'_{A,n}$ and $P_{A,n}^{\text{comb}}$ respectively.

Definition 7.5.2 The *quantum allowed region*, denoted $\mathcal{Q}(A, \mathcal{X}, \mathcal{E}, \mathsf{H}, p)$, is the set of all $(P_A, P'_A, P_A^{\text{comb}}) \in [0,1]^{\times 3}$ that occur, for fixed A, $\mathcal{X}$, $\mathcal{E}$, H, and p, over all allowed pince-nez maps L and L'.

As in the classical case, the existence of the paradox is then determined by the intersection of $\mathcal{Q}$ with the cube $\left(\frac{1}{2}, 1\right] \times \left(\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right)$. For $p = 1$, $P_A^{\text{comb}} = P_A$ and for $p = 0$, $P_A^{\text{comb}} = P'_A$; clearly no paradox is possible. We have the following theorem and claims for $p \in (0,1)$, which indicate that, once again, there are basically only two possibilities, depending on whether we have $\dim \mathsf{H} = 2$ or $\dim \mathsf{H} > 2$.

Theorem 7.5.3 For $p \in (0,1)$, if $\dim \mathsf{H} > 2$, then $\mathcal{Q} \supset (0,1)^{\times 3}$.

Proof Pick any $\varepsilon > 0$ Let $\mathsf{H}_1, \mathsf{H}_2, \mathsf{H}_3$ be three non-trivial subspaces of H such that $\mathsf{H} = \mathsf{H}_1 \oplus \mathsf{H}_2 \oplus \mathsf{H}_3$. Choose any density matrices $\nu_1 \in \mathcal{D}(\mathsf{H}_1)^+$, $\nu_2 \in \mathcal{D}(\mathsf{H}_2)^+$, $\nu_3 \in \mathcal{D}(\mathsf{H}_3)^+$ with unit trace. Let $\kappa : \mathcal{D}(\mathsf{H})^+ \to \mathbb{R}^3$ be the map

$$\kappa(\rho) = \begin{bmatrix} \text{tr } \rho|_{\mathsf{H}_1} \\ \text{tr } \rho|_{\mathsf{H}_2} \\ \text{tr } \rho|_{\mathsf{H}_3} \end{bmatrix} \tag{7.50}$$

and define the pince-nez maps L and L' by

$$\begin{aligned} L\rho &= 1_A \otimes \begin{bmatrix} \nu_1 & 0 & 0 \end{bmatrix} \mathbf{T}\kappa(\rho) + 1_{\tilde{A}} \otimes \begin{bmatrix} 0 & \nu_2 & \nu_3 \end{bmatrix} \mathbf{T}\kappa(\rho) \\ L'\rho &= 1_A \otimes \begin{bmatrix} \nu_1 & 0 & 0 \end{bmatrix} \mathbf{T}'\kappa(\rho) + 1_{\tilde{A}} \otimes \begin{bmatrix} 0 & \nu_2 & \nu_3 \end{bmatrix} \mathbf{T}'\kappa(\rho) \end{aligned} \tag{7.51}$$

for $\mathbf{T}, \mathbf{T}'$ as in theorem **7.2.3**. Because of the properties of ν_1, ν_2, ν_3, and κ, the required calculation reduces to matrix products, which then proceeds as in the proof for theorem

7.3.3. □

Employing a similar construction, for $\dim \mathsf{H} = 2$ we have $\mathcal{Q}(p) \supset \mathcal{CH}(p)$. We have the following two claims, backed by extensive numerical calculation:

Claim 7.5.4 For any fixed $p, P_A, P'_A \in (0,1)$, the minimum and maximum values of P_A^{comb} are achieved for pince-nez maps L and L' each of the form $\rho \to 1_A \otimes J\rho J^* + 1_{\tilde{A}} \otimes K\rho K^*$.

Claim 7.5.5 For any fixed $p, P_A, P'_A \in (0,1)$, the minimum and maximum values of P_A^{comb} lie outside the bounds given in theorem **7.3.3**.

Assuming the validity of the first claim, the extremizing maps can be expressed as

$$L\rho = 1_A \otimes \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix} \rho \begin{bmatrix} \overline{u_1} & \overline{u_2} \\ \overline{v_1} & \overline{v_2} \end{bmatrix} + 1_{\tilde{A}} \otimes \begin{bmatrix} u_3 & v_3 \\ u_4 & v_4 \end{bmatrix} \rho \begin{bmatrix} \overline{u_3} & \overline{u_4} \\ \overline{v_3} & \overline{v_4} \end{bmatrix} \tag{7.52}$$
$$L'\rho = 1_A \otimes \begin{bmatrix} u'_1 & v'_1 \\ u'_2 & v'_2 \end{bmatrix} \rho \begin{bmatrix} \overline{u'_1} & \overline{u'_2} \\ \overline{v'_1} & \overline{v'_2} \end{bmatrix} + 1_{\tilde{A}} \otimes \begin{bmatrix} u'_3 & v'_3 \\ u'_4 & v'_4 \end{bmatrix} \rho \begin{bmatrix} \overline{u'_3} & \overline{u'_4} \\ \overline{v'_3} & \overline{v'_4} \end{bmatrix}$$

for $\|\vec{u}\|^2 = \|\vec{v}\|^2 = 1$, $\vec{u} \perp \vec{v}$ except for points excluded by the requirement of the quantum Perron-Frobenius property. We are then faced by the following optimization problem, which may be treated numerically:

$$\text{constraints: } \vec{w}\mathbf{M}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = P_A, \vec{w}'\mathbf{M}'^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = P'_A \tag{7.53}$$

$$\text{extremize: } P_A^{\text{comb}} = (p\vec{w} + (1-p)\vec{w}')\,(p\mathbf{M} + (1-p)\mathbf{M}')^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

for

$$\vec{w} = \begin{bmatrix} |u_1|^2 + |u_2|^2 & \overline{u_1}v_1 + \overline{u_2}v_2 & u_1\overline{v_1} + u_2\overline{v_2} & |v_1|^2 + |v_2|^2 \end{bmatrix} \tag{7.54}$$
$$\vec{w}' = \begin{bmatrix} |u'_1|^2 + |u'_2|^2 & \overline{u'_1}v'_1 + \overline{u'_2}v'_2 & u'_1\overline{v'_1} + u'_2\overline{v'_2} & |v'_1|^2 + |v'_2|^2 \end{bmatrix}$$
$$\mathbf{M} = \begin{bmatrix} |u_1|^2 + |u_3|^2 - 1 & \overline{u_1}v_1 + \overline{u_3}v_3 & \overline{v_1}u_1 + \overline{v_3}u_3 & |v_1|^2 + |v_3|^2 \\ \overline{u_1}u_2 + \overline{u_3}u_4 & \overline{u_1}v_2 + \overline{u_3}v_4 - 1 & \overline{v_1}u_2 + \overline{v_3}u_4 & \overline{v_1}v_2 + \overline{v_3}v_4 \\ \overline{u_2}u_1 + \overline{u_4}u_3 & \overline{u_2}v_1 + \overline{u_4}v_3 & \overline{v_2}u_1 + \overline{v_4}u_3 - 1 & \overline{v_2}v_1 + \overline{v_4}v_3 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$
$$\mathbf{M}' = \begin{bmatrix} |u'_1|^2 + |u'_3|^2 - 1 & \overline{u'_1}v'_1 + \overline{u'_3}v'_3 & \overline{v'_1}u'_1 + \overline{v'_3}u'_3 & |v'_1|^2 + |v'_3|^2 \\ \overline{u'_1}u'_2 + \overline{u'_3}u'_4 & \overline{u'_1}v'_2 + \overline{u'_3}v'_4 - 1 & \overline{v'_1}u'_2 + \overline{v'_3}u'_4 & \overline{v'_1}v'_2 + \overline{v_3}v'_4 \\ \overline{u'_2}u'_1 + \overline{u'_4}u'_3 & \overline{u'_2}v'_1 + \overline{u'_4}v'_3 & \overline{v'_2}u'_1 + \overline{v'_4}u'_3 - 1 & \overline{v'_2}v'_1 + \overline{v'_4}v'_3 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Graphs, for certain representative values of p, of the minimum achievable P_A^{comb} given P_A, P'_A, are given in the following section.

7.6 Summary of results

In summary, for the considered classical game whose event space $\mathcal{E}$ is effectively three or more states, the analogous classical hidden game whose σ-algebra $\mathcal{F}$ for its hidden set is effectively three or more states, and the analogous quantum game whose Hilbert space is three or more dimensions, we have proven there is no limitation on the probability for Alice to win the combined game, P_A^{comb}, given her probabilities to win the individual games, P_A and P'_A. For the classical game with two-states, $\mathcal{E} = \{\varnothing, A, \tilde{A}, \mathcal{O}\}$, we have proven the paradox cannot occur. For the analogous classical hidden game with two-states, $\mathcal{F} = \{\varnothing, B, \tilde{B}, \mathcal{H}\}$, we have

$$\min\{pP_A, (1-p)P'_A\} < P_A^{\text{comb}} < \max\{1-p+pP_A, p+(1-p)P'_A\} \tag{7.55}$$

where $p \in (0,1)$ is the probability for the coin to land heads.

The following contour plots illustrate the lower bounds on P_A^{comb} for the two representative values of $p = 0.5$ (the fair coin) and $p = 0.1$. The paradox is displayed by minimum achievable values of P_A^{comb} being less than one-half in the region $(P_A, P'_A) \in \left(\frac{1}{2}, 1\right] \times \left(\frac{1}{2}, 1\right]$. For comparison, plots of the minimum achieved P_A^{comb} for the analogous quantum game whose Hilbert space H has dimension two, which were numerically calculated from the optimization problem posed in (7.53), are given as well.

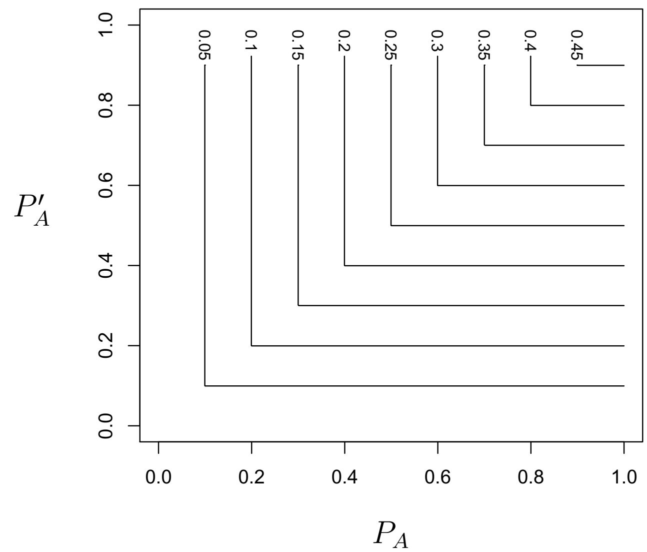


Figure 7.1: Minimum achievable values of P_A^{comb} for $p = 0.5$ for the two-state, classical hidden game.

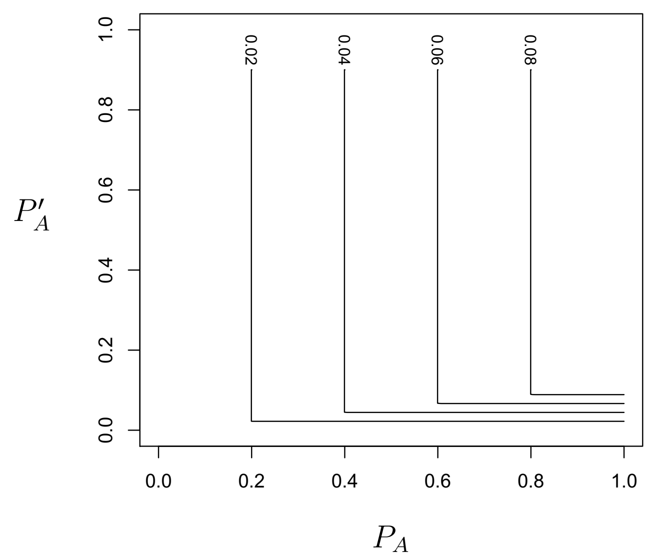


Figure 7.2: Minimum achievable values of P_A^{comb} for $p = 0.1$ for the two-state, classical hidden game.

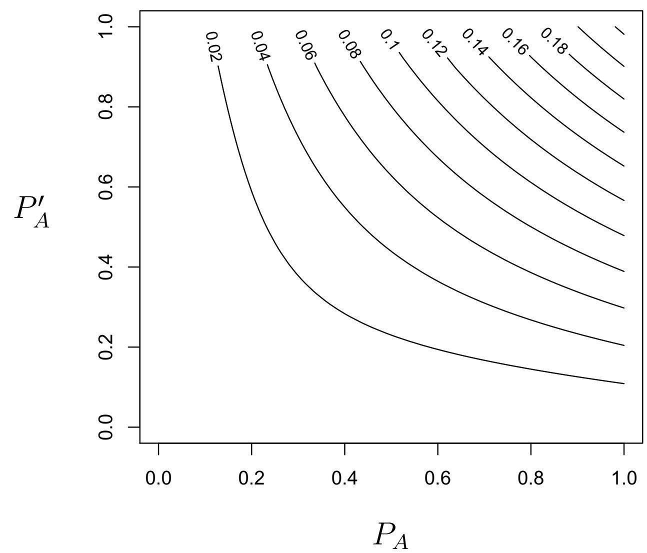


Figure 7.3: Minimum achievable values of P_A^{comb} for $p = 0.5$ for the two-dimensional, quantum game.

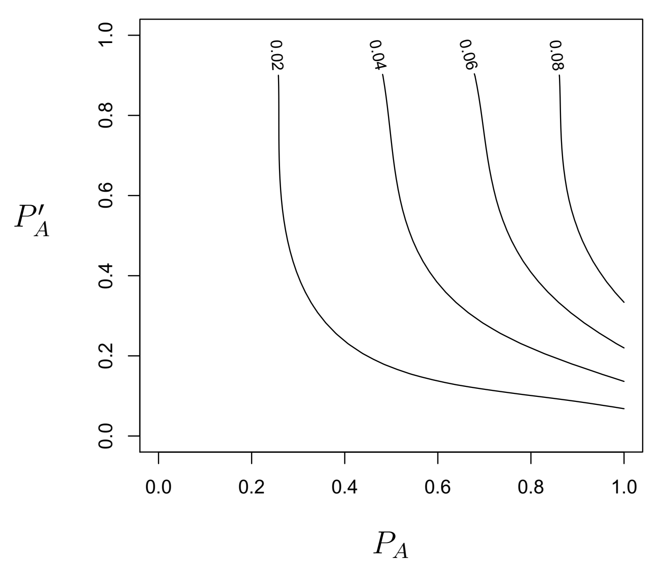


Figure 7.4: Minimum achievable values of P_A^{comb} for $p = 0.1$ for the two-dimensional, quantum game.

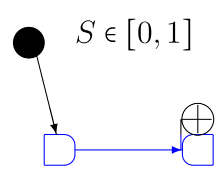
Chapter 8

A Parrondo-like paradox for an one-round game

8.1 Defining the game

Instead of the multi-round game considered in the preceding chapter, now consider a single round quantum game where the winning criteria is still taken to be that the observation is in some specified set A. Instead of having two maps which are combined by a coin flip, the map is now fixed and only the initial states are varied. If they were combined in a convex combination by flipping a coin, there would clearly be no Parrondo's paradox since the probabilities depend linearly on the initial state. However, if the initial states were constrained to be rank one (hence, described by a wavefunction), instead of a convex combination, one may consider the minimizing geodesic joining them. If two initial states both give probability greater than one-half for the first player, Alice, to win, but there is somewhere on the minimizing geodesic where the second player, Bob, has probability greater than one-half to win, we will term that a Parrondo-like paradox.

Suppose there is a continuous control, say a slider with continuous values from zero to one, which varies the initial state along a geodesic in the space of rank-one density matrices. Using the Bayesian network model, this initial state is input into a fixed pince-nez map and the outcome observed. The graphical model is:



(8.1)

Employing option **I'** (see §3.2), the data for the initial node is $([0,1], \text{Lebesgue}; \mathsf{H}; K)$ for some given Hilbert space H. The map K will be determined below in §8.2. The data for the pince-nez is $(\mathsf{H}; \mathcal{R}, \tau; ; L)$ for some given observable set $\mathcal{R}$ with base measure τ. By the rules of the game, it is possible to reduce to the case where the σ-algebra is $\{\varnothing, A, \tilde{A}, \mathcal{R}\}$ with base measure τ assigning both A and $\tilde{A}$ the value one. Suppose L is also given.

The probability for the observation to be in set A (so Alice wins) is then

$$\int_A (L \circ K)\nu \, d\tau = ((L \circ K)\nu)|_A \tag{8.2}$$

8.2 Defining geodesics on the space of wavefunctions

Let $\mathbb{S}_{\mathsf{H}} \subset \mathsf{H}$ be the unit ball. For any $\zeta, \eta \in \mathcal{S}$, let $\sim$ be the equivalence relation $\zeta \sim \eta$ if $\zeta = \omega\eta$ for some phase $\omega \in \mathbb{S}^1 \subset \mathbb{C}$ and let $[\cdot]$ denote the equivalence classes. Consider a wavefunction $\psi \in \mathbb{S}_{\mathsf{H}}$. Since only the rank-one density matrix $\psi\psi^* = \psi\langle\cdot, \psi\rangle$ is meaningful, ψ is only defined up to an overall phase, so what should actually be considered is the equivalence class $[\psi]$ in the quotient space $\mathbb{S}_{\mathsf{H}}/\sim$. The question is then what is the correct topology and metric structure to place on this quotient space.

Choice one–using the trace norm

Using the topology and metric structure inherited from the placement of rank-one density matrices within all density matrices, consider the choice of metric $\mathrm{dist}_{\mathrm{trace}}$ on $\mathbb{S}_{\mathsf{H}}/\sim$ given by, for any $[\xi], [\psi] \in \mathcal{S}/\sim$,

$$\mathrm{dist}_{\mathrm{trace}}([\xi], [\psi]) = \mathrm{tr}\ |\xi\xi^* - \psi\psi^*| = 2\sqrt{1 - |\langle\psi, \xi\rangle|^2} \tag{8.3}$$

The last equality holds since $\mathrm{span}_{\mathbb{C}}\{\xi, \psi\}$ is an invariant subspace of the operator

$$(\xi\xi^* - \psi\psi^*)^* (\xi\xi^* - \psi\psi^*) = (\xi\xi^* - \psi\psi^*)^2 \tag{8.4}$$

with eigenvalue $1 - |\langle\psi, \xi\rangle|^2$ while the orthogonal subspace is the kernel of the operator.

Choice two–using the round metric on $\mathbb{S}_{\mathsf{H}}$

Let $\mathrm{dist}_{\mathrm{round}}$ be the round metric on $\mathbb{S}_{\mathsf{H}}$, which is the standard metric induced by its embedding in H equipped with its norm, so, for any $\zeta, \eta \in \mathbb{S}_{\mathsf{H}}$,

$$\mathrm{dist}_{\mathrm{round}}(\zeta, \eta) = \arccos \mathfrak{R}\langle\zeta, \eta\rangle \tag{8.5}$$

Then let the metric $\mathrm{dist}_{\sim\mathrm{round}}$ on $\mathbb{S}_{\mathsf{H}}/\sim$ be given by the usual prescription for quotient spaces, so, for any $[\xi], [\psi] \in \mathbb{S}_{\mathsf{H}}/\sim$,

$$\mathrm{dist}_{\sim\mathrm{round}}([\xi], [\psi]) = \min_{w \in \mathbb{S}^1 \subset \mathbb{C}} \mathrm{dist}_{\mathrm{round}}(\xi, w\psi) = \arccos \max_{w \in \mathbb{S}^1 \subset \mathbb{C}} \mathfrak{R}\langle\psi, w\xi\rangle = \arccos|\langle\psi, \xi\rangle| \tag{8.6}$$

Then

$$\mathrm{dist}_{\mathrm{trace}}([\xi], [\psi]) = 2 \sin \mathrm{dist}_{\sim\mathrm{round}}([\xi], [\psi]) \tag{8.7}$$
$$\Leftrightarrow \mathrm{dist}_{\sim\mathrm{round}}([\xi], [\psi]) = \arcsin \frac{1}{2} \mathrm{dist}_{\mathrm{trace}}([\xi], [\psi])$$

In particular, the metrics are equivalent and give rise to the same topology on $\mathbb{S}_{\mathsf{H}}/\sim$.

Geodesics on the quotient space

Fix any $\xi, \psi \in \mathbb{S}_{\mathsf{H}}$ with corresponding equivalence classes $[\xi], [\psi] \in \mathbb{S}_{\mathsf{H}}/\sim$. Let $\hat{\xi}$ be an element of $[\xi]$ such that $\mathrm{dist}_{\mathrm{round}}(\hat{\xi}, \psi) = \mathrm{dist}_{\sim\mathrm{round}}([\xi], [\psi]) \Leftrightarrow \langle \psi, \hat{\xi} \rangle = |\langle \psi, \xi \rangle|$, so $\hat{\xi} = \dfrac{\langle \psi, \xi \rangle}{|\langle \psi, \xi \rangle|}\xi$ if $\langle \psi, \xi \rangle \neq 0$ while $\hat{\xi}$ can be any element of $[\xi]$ if $\langle \psi, \xi \rangle = 0$. Now take the arc $\gamma : \left[0, \mathrm{dist}_{\mathrm{round}}(\hat{\xi}, \psi)\right] \to \mathbb{S}_{\mathsf{H}}$ of the great circle in $\mathbb{S}_{\mathsf{H}}$ connecting ψ and $\hat{\xi}$ with the standard parametrization, given by

$$\gamma(\theta) = (\cos\theta)\,\psi + (\sin\theta)\,\frac{\hat{\xi} - \langle \hat{\xi}, \psi \rangle \psi}{\|\hat{\xi} - \langle \hat{\xi}, \psi \rangle \psi\|} \tag{8.8}$$

where $\|\hat{\xi} - \langle \hat{\xi}, \psi \rangle \psi\| = \sqrt{\langle \hat{\xi} - \langle \hat{\xi}, \psi \rangle \psi, \hat{\xi} - \langle \hat{\xi}, \psi \rangle \psi \rangle} = \sqrt{1 - \langle \hat{\xi}, \psi \rangle^2}$.

Proposition 8.2.1 The curve $[\gamma]$ in $\mathbb{S}_{\mathsf{H}}/\sim$ is an unit speed geodesic with respect to the metric $\mathrm{dist}_{\sim\mathrm{round}}$.

Proof For any $\theta \in \left(0, \mathrm{dist}_{\mathrm{round}}(\hat{\xi}, \psi)\right)$,

$$\begin{aligned} \mathrm{dist}_{\sim\mathrm{round}}([\gamma(\theta)], [\psi]) + \mathrm{dist}_{\sim\mathrm{round}}([\xi], [\gamma(\theta)]) &= \mathrm{dist}_{\mathrm{round}}(\gamma(\theta), \psi) + \mathrm{dist}_{\mathrm{round}}(\hat{\xi}, \gamma(\theta)) \\ &= \mathrm{dist}_{\mathrm{round}}(\hat{\xi}, \psi) \\ &= \mathrm{dist}_{\sim\mathrm{round}}([\xi], [\psi]) \end{aligned} \tag{8.9}$$

so not only is the curve $[\gamma]$ a geodesic, but by further partitioning, the metric is seen to be precisely the arclength of the geodesic. Also, for any $\theta \in \left(0, \mathrm{dist}_{\mathrm{round}}(\hat{\xi}, \psi)\right)$,

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\,\mathrm{dist}_{\sim\mathrm{round}}([\gamma(\theta + \varepsilon)], [\gamma(\theta)]) = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon}\,\arccos\left(\cos(\varepsilon)\right) = \lim_{\varepsilon \to 0^+} \frac{|\varepsilon|}{\varepsilon} = 1 \tag{8.10}$$

so the curve is unit speed. □

Proposition 8.2.2 The curve $[\gamma]$ in $\mathbb{S}_{\mathsf{H}}/\sim$ is a geodesic with respect to the metric $\mathrm{dist}_{\mathrm{trace}}$.

Proof Since $2\sin z = 2z + \mathcal{O}(z^3)$ as $z \to 0$, the arclength of a curve with respect to the metric $\mathrm{dist}_{\mathrm{trace}}$ is simply twice the arclength with respect to the metric $\mathrm{dist}_{\sim\mathrm{round}}$. Hence, the two metrics have the same geodesics.□

Completing the definition of the game.

We still need to define the initialization map K from §8.1 which uses the slider position to determine a point on the geodesic. From the preceding results on geodesics, K is determined by first fixing the endpoints $[\xi], [\psi] \in \mathbb{S}_{\mathsf{H}}/\sim$. Let $\delta = \mathrm{dist}_{\sim\mathrm{round}}([\xi],[\psi])$. Then, for ν a measure on $[0,1]$ absolutely continuous with respect to Lebesgue measure λ,

$$K\frac{d\nu}{d\lambda} = \int_{x\in[0,1]} \gamma(x\delta)\gamma(x\delta)^*\, d\nu(x) \tag{8.11}$$

If ν is sufficiently concentrated, then, as desired, the image of K is approximately rank-one. The complication of dealing with a concentrated measure ν for the slider position is another instance of the previously mentioned problem (see §3.2 and §4.5) of inputting parameters encountered when employing option **I'**–the value of the slider cannot be read in directly. For purposes of simplicity and clarity, in the remainder we will instead simply suppose we can directly select the desired point on the geodesic and work with strictly rank-one density matrices.

8.3 Bounds on the extent of the Parrondo-like paradox

To quantify the extent of the Parrondo-like paradox, by analogy to §7 we have P_A, P'_A, as the probability Alice wins with initial wavefunction $[\xi]$ and $[\psi]$ respectively. Define P_A^{geo} as the probability Alice wins with initial wavefunction for a specified point on the geodesic joining $[\xi]$ and $[\psi]$. Analogously to definition **7.5.2**, we have the following:

Definition 8.3.1 The *quantum allowed region*, denoted $\mathcal{Q}(A,\mathcal{R},\tau,\mathsf{H})$, is the set of all $(P_A, P'_A, P_A^{\mathrm{geo}}) \in [0,1]^{\times 3}$ that occur–for fixed A, $\mathcal{R}$, τ, and H–over all allowed pince-nez maps L, initial wavefunctions $[\xi]$ and $[\psi]$, and points on the geodesic joining them.

The paradox can occur if $\mathcal{Q}$ intersects the cube $\left(\frac{1}{2},1\right]\times\left(\frac{1}{2},1\right]\times\left[0,\frac{1}{2}\right)$. We have the following theorem giving $\mathcal{Q}$ precisely:

Theorem 8.3.2 If the Hilbert space H is nontrivial, $\mathcal{Q}$ is the closed region

$$\max\{0, P_A + P'_A - 1\} \le P_A^{\mathrm{geo}} \le \min\{P_A + P'_A, 1\}$$

Proof Since the bounded operators are dual to the trace-class ones, there is some self-adjoint operator $\eta \in \mathcal{B}(\mathsf{H})$ such that $L(\rho)|_A = \mathrm{tr}\ \eta\rho$ for every $\rho \in \mathcal{D}(\mathsf{H})^+$ and, by **B5.15** and **B5.16**, every $\rho \in \mathcal{S}_1(\mathsf{H})$. The condition on η is that it is in the order interval $0 \le \eta \le I_{\mathsf{H}}$. Let the 2×2-, Hermitian matrix $B = [b_{jk}]$ be given by

$$\begin{bmatrix} L(\psi\psi^*)|_A & L(\psi\hat{\xi}^*)|_A \\ L(\hat{\xi}\psi^*)|_A & L(\xi\xi^*)|_A \end{bmatrix} = \begin{bmatrix} \langle\eta\psi,\psi\rangle & \langle\eta\psi,\hat{\xi}\rangle \\ \langle\eta\hat{\xi},\psi\rangle & \langle\eta\xi,\xi\rangle \end{bmatrix} \tag{8.12}$$

Let $\mathcal{B}$ be the set of all such matrices over $0 \le \eta \le I_{\mathsf{H}}$ and $\psi, \xi \in \mathbb{S}_{\mathsf{H}}$.

To get a simpler characterization of $\mathcal{B}$, let $\mathcal{C}$ be the union of order intervals of 2×2-, Hermitian matrices given by

$$\bigcup_{\delta\in[0,\frac{\pi}{2}]}\left\{0\le C\le\begin{bmatrix}1 & \cos\delta\\ \cos\delta & 1\end{bmatrix}\right\} \tag{8.13}$$

The claim is that $\mathcal{B}=\mathcal{C}$. To see this is true, take any $B\in\mathcal{B}$. Since $\eta\ge 0$, for any $a,b\in\mathcal{C}$,

$$\begin{bmatrix}\bar{a} & \bar{b}\end{bmatrix}B\begin{bmatrix}a\\ b\end{bmatrix}=\begin{bmatrix}\bar{a} & \bar{b}\end{bmatrix}\begin{bmatrix}\langle\eta\psi,\psi\rangle & \langle\eta\psi,\hat{\xi}\rangle\\ \langle\eta\hat{\xi},\psi\rangle & \langle\eta\xi,\xi\rangle\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}=\left\langle\eta(a\psi+b\hat{\xi}),a\psi+b\hat{\xi}\right\rangle\ge 0 \tag{8.14}$$

Therefore, $B\ge 0$. Similarly, since $I_{\mathsf{H}}-\eta\ge 0$, for $\cos\delta=\langle\psi,\hat{\xi}\rangle$,

$$\begin{aligned}\begin{bmatrix}\bar{a} & \bar{b}\end{bmatrix}\left(\begin{bmatrix}1 & \cos\delta\\ \cos\delta & 1\end{bmatrix}-B\right)\begin{bmatrix}a\\ b\end{bmatrix}&=\begin{bmatrix}\bar{a} & \bar{b}\end{bmatrix}\begin{bmatrix}\langle(I_{\mathsf{H}}-\eta)\psi,\psi\rangle & \langle(I_{\mathsf{H}}-\eta)\psi,\hat{\xi}\rangle\\ \langle(I_{\mathsf{H}}-\eta)\hat{\xi},\psi\rangle & \langle(I_{\mathsf{H}}-\eta)\xi,\xi\rangle\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}\\ &=\left\langle(I_{\mathsf{H}}-\eta)(a\psi+b\hat{\xi}),a\psi+b\hat{\xi}\right\rangle\end{aligned} \tag{8.15}$$

which is always greater than or equal to zero, so $B\le\begin{bmatrix}1 & \cos\delta\\ \cos\delta & 1\end{bmatrix}$ for $\delta=\arccos\langle\psi,\hat{\xi}\rangle\in\left[0,\frac{\pi}{2}\right]$. Hence, $\mathcal{B}\subset\mathcal{C}$.

Now take any $C\in\mathcal{C}$, so there is some $\delta\in\left[0,\frac{\pi}{2}\right]$ such that $C\le\begin{bmatrix}1 & \cos\delta\\ \cos\delta & 1\end{bmatrix}$. Since H is nontrivial, it has a pair of orthonormal vectors, $\{\mathbf{e}_1,\mathbf{e}_2\}$. Take $\psi=\mathbf{e}_1$ and $\xi=\hat{\xi}=\cos\delta\mathbf{e}_1+\sin\delta\mathbf{e}_2$, which are both clearly of unit norm. Take the operator η to be zero on the complement of the span of $\{\mathbf{e}_1,\mathbf{e}_2\}$. On the span, using $\{\mathbf{e}_1,\mathbf{e}_2\}$ as the basis, let η be given by

$$\begin{aligned}C=\begin{bmatrix}\langle\eta\psi,\psi\rangle & \langle\eta\psi,\hat{\xi}\rangle\\ \langle\eta\hat{\xi},\psi\rangle & \langle\eta\xi,\xi\rangle\end{bmatrix}&=\begin{bmatrix}1 & 0\\ \cos\delta & \sin\delta\end{bmatrix}\eta\begin{bmatrix}1 & \cos\delta\\ 0 & \sin\delta\end{bmatrix}\\ \Leftrightarrow\eta&=\begin{bmatrix}1 & 0\\ \cos\delta & \sin\delta\end{bmatrix}^{-1}C\begin{bmatrix}1 & \cos\delta\\ 0 & \sin\delta\end{bmatrix}^{-1}\end{aligned} \tag{8.16}$$

Then, since $C\ge 0$, clearly $\eta\ge 0$. Since $C\le\begin{bmatrix}1 & \cos\delta\\ \cos\delta & 1\end{bmatrix}$, for $\mathbf{v}=\begin{bmatrix}1 & \cos\delta\\ 0 & \sin\delta\end{bmatrix}^{-1}\begin{bmatrix}a\\ b\end{bmatrix}$,

$$\begin{bmatrix}\bar{a} & \bar{b}\end{bmatrix}(I_{\mathsf{H}}-\eta)\begin{bmatrix}a\\ b\end{bmatrix}=\mathbf{v}^*\left(\begin{bmatrix}1 & \cos\delta\\ \cos\delta & 1\end{bmatrix}-C\right)\mathbf{v}\ge 0 \tag{8.17}$$

Hence, $\eta\le I_{\mathsf{H}}$, so $\mathcal{C}\subset\mathcal{B}$ and $\mathcal{B}=\mathcal{C}$.

Therefore, using the above expression (8.8) for the geodesic and $\|\hat{\xi}-\langle\hat{\xi},\psi\rangle\psi\|=\sin\delta$, we wish to extremize

$$P_A^{\text{geo}}=\frac{1}{\sin^2\delta}\begin{bmatrix}\sin(\delta-\theta) & \sin\theta\end{bmatrix}B\begin{bmatrix}\sin(\delta-\theta)\\ \sin\theta\end{bmatrix} \tag{8.18}$$

over all $0 \le B \le \begin{bmatrix} 1 & \cos\delta \\ \cos\delta & 1 \end{bmatrix}$, $\delta \in [0, \frac{\pi}{2}]$, and $\theta \in [0, \delta]$ for fixed $P_A = b_{11}$ and $P'_A = b_{22}$. Setting the imaginary parts of the off-diagonal entries of B to zero keeps B in the allowed order interval and does not change the value of P_A, P'_A, or P_A^{geo}, so B can be taken real.

For fixed δ and θ, the expression in (8.18) is linear in B. Since the allowed B form a convex set, the extrema are achieved on the set of extreme points, so B can be restricted to either *(i)* being rank one or *(ii)* having $\begin{bmatrix} 1 & \cos\delta \\ \cos\delta & 1 \end{bmatrix} - B$ be rank one. For case *(i)*, we have two subcases for the choice of either + or − in

$$B = \begin{bmatrix} b_{11} & \pm\sqrt{b_{11}b_{22}} \\ \pm\sqrt{b_{11}b_{22}} & b_{22} \end{bmatrix} \tag{8.19}$$

Choosing the +, then $b_{11}, b_{22} \in [0,1]$ with

$$\max\{0, \sqrt{b_{11}b_{22}} - \sqrt{(1-b_{11})(1-b_{22})}\} \le \cos\delta \le \min\{1, \sqrt{b_{11}b_{22}} + \sqrt{(1-b_{11})(1-b_{22})}\} \tag{8.20}$$

and $\theta \in [0, \delta]$. Let $f_1^{\max}(b_{11}, b_{22})$ be the maximum of (8.18) over all allowed δ, θ for the given b_{11}, b_{22} and $f_1^{\min}(b_{11}, b_{22})$ be the minimum. Then $f_1^{\max}(b_{11}, b_{22}) = \min\{b_{11} + b_{22}, 1\}$, with the maximizing δ_0, θ_0 given by $\delta_0 = \frac{\pi}{2}$, $\sin\theta_0 = \sqrt{\frac{b_{22}}{b_{11}+b_{22}}}$ if $b_{11} + b_{22} \le 1$ and by $\cos\delta_0 = \sqrt{b_{11}b_{22}} - \sqrt{(1-b_{11})(1-b_{22})}$, $\sin\theta_0 = \sqrt{1-b_{11}}$ if $b_{11} + b_{22} > 1$. The minimum bound is $f_1^{\min}(b_{11}, b_{22}) = \min\{b_{11}, b_{22}\}$, with the minimizing δ_0, θ_0 given by $\theta_0 = 0$ if $b_{11} \le b_{22}$ and $\theta_0 = \delta_0$ if $b_{11} > b_{22}$, with δ_0 arbitrary. Choosing the −, then $(b_{11}, b_{22}) \in [0,1]^{\times 2} \cap \{b_{11} + b_{22} \le 1\}$ with

$$0 \le \cos\delta \le \min\{1, \sqrt{(1-b_{11})(1-b_{22})} - \sqrt{b_{11}b_{22}}\} \tag{8.21}$$

and $\theta \in [0, \delta]$. Let $f_2^{\max}(b_{11}, b_{22})$ be the maximum of (8.18) over all allowed δ, θ for the given b_{11}, b_{22} and $f_2^{\min}(b_{11}, b_{22})$ be the minimum. Then $f_2^{\max}(b_{11}, b_{22}) = \max\{b_{11}, b_{22}\}$, with the maximizing δ_0, θ_0 given complementary to that for the preceding $f_1^{\min}$. The minimum bound is $f_2^{\min}(b_{11}, b_{22}) = 0$, with the minimizing δ_0, θ_0 given by $\delta_0 = \frac{\pi}{2}$, $\sin\theta_0 = \sqrt{\frac{b_{11}}{b_{11}+b_{22}}}$.

For case *(ii)*, once again we have two subcases for the choice of either + or − in

$$B = \begin{bmatrix} b_{11} & \cos\delta \pm \sqrt{(1-b_{11})(1-b_{22})} \\ \cos\delta \pm \sqrt{(1-b_{11})(1-b_{22})} & b_{22} \end{bmatrix} \tag{8.22}$$

Choosing the +, then $(b_{11}, b_{22}) \in [0,1]^{\times 2} \cap \{b_{11} + b_{22} \ge 1\}$ with

$$0 \le \cos\delta \le \min\{1, \sqrt{b_{11}b_{22}} - \sqrt{(1-b_{11})(1-b_{22})}\} \tag{8.23}$$

and $\theta \in [0, \delta]$. Let $f_3^{\max}(b_{11}, b_{22})$ be the maximum of (8.18) over all allowed δ, θ for the given b_{11}, b_{22} and $f_3^{\min}(b_{11}, b_{22})$ be the minimum. Then $f_3^{\max}(b_{11}, b_{22}) = 1$, with the minimizing δ_0, θ_0 given by $\delta_0 = \frac{\pi}{2}$, $\sin\theta_0 = \sqrt{\frac{1-b_{11}}{2-b_{11}-b_{22}}}$. The minimum bound is $f_3^{\min}(b_{11}, b_{22}) = \min\{b_{11}, b_{22}\}$ with the minimizing δ_0, θ_0 given similarly to that for the preceding $f_1^{\min}$. Choosing the $-$, then $b_{11}, b_{22} \in [0,1]$ with

$$\max\{0, \sqrt{(1-b_{11})(1-b_{22})} - \sqrt{b_{11}b_{22}}\} \le \cos\delta \le \min\{1, \sqrt{(1-b_{11})(1-b_{22})} + \sqrt{b_{11}b_{22}}\} \quad (8.24)$$

and $\theta \in [0,\delta]$. Let $f_4^{\max}(b_{11}, b_{22})$ be the maximum of (8.18) over all allowed δ, θ for the given b_{11}, b_{22} and $f_4^{\min}(b_{11}, b_{22})$ be the minimum. Then $f_4^{\max}(b_{11}, b_{22}) = \min\{b_{11}, b_{22}\}$ with the maximizing δ_0, θ_0 given similarly to that for the preceding $f_1^{\min}$. The minimum bound is $f_4^{\min}(b_{11}, b_{22}) = \max\{0, b_{11} + b_{22} - 1\}$, with the minimizing δ_0, θ_0 given by $\cos\delta_0 = \sqrt{1-b_{11}}\sqrt{1-b_{22}} - \sqrt{b_{11}b_{22}}$, $\sin\theta_0 = \sqrt{b_{11}}$ if $b_{11} + b_{22} \le 1$ and $\delta_0 = \frac{\pi}{2}$, $\sin\theta_0 = \sqrt{\frac{1-b_{22}}{2-b_{11}-b_{22}}}$ if $b_{11} + b_{22} > 1$.

Putting the cases and subcases together, the minimum value of P_A^{geo} for fixed P_A, P'_A is given by

$$\left(f_1^{\min} \wedge f_2^{\min} \wedge f_3^{\min} \wedge f_4^{\min}\right)(P_A, P'_A) = f_4^{\min}(P_A, P'_A) = \max\{0, P_A + P'_A - 1\} \quad (8.25)$$

The maximum value of P_A^{geo} for fixed P_A, P'_A is given by

$$(f_1^{\max} \vee f_2^{\max} \vee f_3^{\max} \vee f_4^{\max})(P_A, P'_A) = f_1^{\max}(P_A, P'_A) = \min\{P_A + P'_A, 1\} \quad (8.26)$$

With fixed values for P_A and P'_A, within each case and subcase P_A^{geo} is a continuous function of δ and θ, so all intermediate values for P_A^{geo} are achieved. □

Note that, by the result of the preceding theorem, the paradox can only occur for values of (P_A, P'_A) in the triangle bounded by $P_A > \frac{1}{2}$, $P'_A > \frac{1}{2}$, and $P_A + P'_A < \frac{3}{2}$.

8.4 Conditions for the occurrence of the Parrondo-like paradox

The matrix B defined in the proof of the preceding theorem (8.12) puts restrictions on the occurrence of the paradox. If $B \le \frac{1}{2} I_2$ then both P_A and P'_A are less than or equal to one-half, so the paradox cannot occur for any choices of $[\psi]$, $[\xi]$, or point on the geodesic joining them. Similarly, if the trace of B is less than or equal to one, then either P_A or P'_A is less than one-half, so the paradox also cannot occur. From these obvious statements, we then have the following nontrivial result:

Theorem 8.4.1 If, for any particular orthonormal $\mathbf{e}_1, \mathbf{e}_2 \in \mathsf{H}$,

$$\operatorname{tr}\begin{bmatrix} L(\mathbf{e}_1\mathbf{e}_1^*)|_A & L(\mathbf{e}_1\mathbf{e}_2^*)|_A \\ L(\mathbf{e}_2\mathbf{e}_1^*)|_A & L(\mathbf{e}_2\mathbf{e}_2^*)|_A \end{bmatrix} \le 1$$

then, for any normalized $\psi, \xi \in \operatorname{span}_{\mathbb{C}}\{\mathbf{e}_1, \mathbf{e}_2\}$ and point on the geodesic joining $[\psi]$ and $[\xi]$, the paradox cannot occur.

Proof Let $\psi = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$ and $\hat{\xi} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}$ with $|a|^2 + |b|^2 = |c|^2 + |d|^2 = 1$ and $a\overline{c} + b\overline{d} = \cos\delta$ real and greater than or equal to zero. Let

$$C = \begin{bmatrix} L(\mathbf{e}_1\mathbf{e}_1^*)|_A & L(\mathbf{e}_1\mathbf{e}_2^*)|_A \\ L(\mathbf{e}_2\mathbf{e}_1^*)|_A & L(\mathbf{e}_2\mathbf{e}_2^*)|_A \end{bmatrix} \tag{8.27}$$

If $C \le \frac{1}{2} I_2$, then

$$\text{tr } B = \text{tr} \begin{bmatrix} \overline{a} & \overline{b} \\ \overline{c} & \overline{d} \end{bmatrix} C \begin{bmatrix} a & c \\ b & d \end{bmatrix} \le 1 \tag{8.28}$$

so, by the comment preceding the theorem, the paradox cannot occur. Therefore, the only remaining case is where C, whose eigenvalues are necessarily real, has one eigenvalue, $\lambda_1 > \frac{1}{2}$, and one eigenvalue, $\lambda_2 < \frac{1}{2}$, with $\lambda_1 + \lambda_2 \le 1$. There are corresponding normalized eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$, necessarily orthogonal. Writing $\begin{bmatrix} a \\ c \end{bmatrix} = f_1\mathbf{v}_1 + f_2\mathbf{v}_2$ and $\begin{bmatrix} b \\ d \end{bmatrix} = g_1\mathbf{v}_1 + g_2\mathbf{v}_2$, the above conditions on a, b, c, d become the following conditions on f_1, f_2, g_1, g_2: $|f_1|^2 + |f_2|^2 = 1$, $|g_1|^2 + |g_2|^2 = 1$, and $f_1\overline{g_1} + f_2\overline{g_2} = \cos\delta$ is a positive real or zero.

For the paradox to occur, it must be that both $|f_1|^2\lambda_1 + |f_2|^2\lambda_2 = b_{11} = P_A > \frac{1}{2}$ and $|g_1|^2\lambda_1 + |g_2|^2\lambda_2 = b_{22} = P'_A > \frac{1}{2}$; hence,

$$|f_1| > \sqrt{\frac{\frac{1}{2} - \lambda_2}{\lambda_1 - \lambda_2}} \text{ and } |f_2| < \sqrt{\frac{\lambda_1 - \frac{1}{2}}{\lambda_1 - \lambda_2}} \Rightarrow |f_1| > \sqrt{\frac{\frac{1}{2} - \lambda_2}{\lambda_1 - \frac{1}{2}}}\, |f_2| \tag{8.29}$$

$$|g_1| > \sqrt{\frac{\frac{1}{2} - \lambda_2}{\lambda_1 - \lambda_2}} \text{ and } |g_2| < \sqrt{\frac{\lambda_1 - \frac{1}{2}}{\lambda_1 - \lambda_2}} \Rightarrow |g_1| > \sqrt{\frac{\frac{1}{2} - \lambda_2}{\lambda_1 - \frac{1}{2}}}\, |g_2| \tag{8.30}$$

Then, since

$$|f_1\overline{g_1}| = |f_1||g_1| > \frac{\frac{1}{2} - \lambda_2}{\lambda_1 - \frac{1}{2}} |f_2||g_2| > |f_2||g_2| = |f_2\overline{g_2}| \tag{8.31}$$

it must be that $f_1\overline{g_1} + f_2\overline{g_2}$ is actually strictly positive and that $\Re f_1\overline{g_1} > 0$. Furthermore, since the imaginary parts of $f_1\overline{g_1}$ and $f_2\overline{g_2}$ are equal in magnitude, it must be that the real part of $f_1\overline{g_1}$ is greater than $\frac{\frac{1}{2}-\lambda_2}{\lambda_1-\frac{1}{2}}$ times the magnitude of the real part of $f_2\overline{g_2}$ and, therefore, is greater than $\frac{\frac{1}{2}-\lambda_2}{\lambda_1-\frac{1}{2}}$ times the real part of $f_2\overline{g_2}$. Hence, rearranging terms,

$$\Re\left(\lambda_1 f_1\overline{g_1} + \lambda_2 f_2\overline{g_2}\right) > \frac{1}{2}\left(f_1\overline{g_1} + f_2\overline{g_2}\right) = \frac{1}{2}\cos\delta \tag{8.32}$$

However, then

$$
\begin{aligned}
P_A^{\text{geo}} &= \frac{1}{\sin^2\delta}\left[\begin{array}{cc} \sin(\delta-\theta) & \sin\theta \end{array}\right] B \left[\begin{array}{c} \sin(\delta-\theta) \\ \sin\theta \end{array}\right] \\
&= \frac{1}{\sin^2\delta}\Big(\big(|f_1|^2\lambda_1+|f_2|^2\lambda_2\big)\sin^2(\delta-\theta)+\big(|g_1|^2\lambda_1+|g_2|^2\lambda_2\big)\sin^2\theta \\
&\quad +2\Re\left(\lambda_1 f_1\overline{g_1}+\lambda_2 f_2\overline{g_2}\right)\sin(\delta-\theta)\sin\theta\Big) \\
&> \frac{1}{\sin^2\delta}\left(\frac{1}{2}\sin^2(\delta-\theta)+\frac{1}{2}\sin^2\theta+\cos\delta\sin(\delta-\theta)\sin\theta\right) \\
&= \frac{1}{2\sin^2\delta}\left(\sin(\delta-\theta)\left(\sin(\delta-\theta)+\cos\delta\sin\theta\right)\right. \\
&\quad \left.+\sin\theta\left(\sin(\delta-(\delta-\theta))+\cos\delta\sin(\delta-\theta)\right)\right) \\
&= \frac{1}{2\sin^2\delta}\left(\sin(\delta-\theta)\sin\delta\cos\theta+\sin\theta\sin\delta\cos(\delta-\theta)\right) \\
&= \frac{\sin^2\delta}{2\sin^2\delta}=\frac{1}{2}
\end{aligned}
\tag{8.33}
$$

Therefore, if both $b_{11} = P_A > \frac{1}{2}$ and $b_{22} = P'_A > \frac{1}{2}$, then $P_A^{\text{geo}} > \frac{1}{2}$ everywhere on the geodesic. □

Chapter 9

Quantum walks and the Parrondo-like paradox

9.1 Classical random and classical hidden walks

Definitions of classical random and classical hidden walks

A classical random walk is a special case of the observable Markov chain earlier discussed, with graphical model given in figure (7.1). It models a walker who is in one of a finite number of internal states and occupies one of a countable number of positions at one time, so the space $\mathcal{X}$ is either $\mathcal{J} \times \mathbb{Z}^{+}$ for a walk on the half-line, or $\mathcal{J} \times \mathbb{Z}$ for a walk on the full line. After each time-step, the walker is at the same or a neighboring location and its internal state can change as well. The process is random, with the transition probability functions possibly dependent on the internal state as well as on the location. This constrains the transition probability functions further than the constraints already imposed by the Markov conditions, but these further constraints are not indicated in the graphical model. If the transition probability functions are spatially translation invariant, the walk is termed homogeneous.

Now we may consider a classical hidden walk. This is in some ways a special case of the classical hidden-Markov chain earlier discussed (7.16), while in other ways it is a generalization. The hidden set $\mathcal{H}$ is required to be either $\mathcal{J} \times \mathbb{Z}^{+}$ for a walk on the half-line, or $\mathcal{J} \times \mathbb{Z}$ for a walk on the full line. The base measure for the hidden set is required to be the counting measure, and $L^1(\mathcal{H}; \text{counting measure}) \cong \ell^1$, so all integrals can be taken to just be sums. The space ℓ^1 has a natural basis of sequences that have a single nonzero entry with value one. There is a dual "basis" of sequences in ℓ^∞ of the same sequences, which is not a basis in the norm topology, but is one in the weak* topology. This basis and dual "basis" can be used to assign matrix elements to any operator in $\mathcal{B}(\ell^1)$. Then any operator, A, in $\mathcal{B}(\ell^1)$ is in one-to-one correspondence to a certain sequence of matrices, $\{A_n\}$, which may be considered the truncations of the infinite matrix corresponding to the operator. By the triangle inequality, these sequences converge to the operator in the strong-operator topology:

$$\begin{aligned}\lim_{n\to\infty} \|A\mathbf{x} - A_n\mathbf{x}\| = \lim_{n\to\infty} \|A\mathbf{x} - P_nAP_n\mathbf{x}\| &\le \lim_{n\to\infty} \left(\|A\mathbf{x} - P_nA\mathbf{x}\| + \|P_nA\mathbf{x} - P_nAP_n\mathbf{x}\|\right) \\ &\le \lim_{n\to\infty} \left(\|(A\mathbf{x}) - P_n(A\mathbf{x})\| + \|P_nA\|_{\mathrm{op}}\|\mathbf{x} - P_n\mathbf{x}\|\right) = 0\end{aligned} \tag{9.1}$$

for any $\mathbf{x} \in \ell^1$, where $\{P_n\}$ are the usual, diagonal projectors onto the span of the first n basis elements.

Conversely, given a sequence of matrices, each of which is the truncation of the following, the condition on the sequence so that it actually corresponds to a bounded operator is that the induced operator norm of all the matrices is bounded. However, the operator norm induced by the ℓ^1 norm is simply the supremum over all columns of the column sum of the magnitudes of the entries, so it is readily calculated. With this form for the maps, the condition that this is a walk (rather than some other sort of hidden-Markov process) is that matrix entries connecting spatial locations that are not neighboring are all zero.

The generalization from the classical hidden-Markov chain is that the last pince-nez map in the chain is no longer required to be the same as the preceding ones; in particular, all the preceding pince-nez can be taken to be simply nodes, so the graphical model is

(9.2)

Note there is nothing graphically that distinguishes this from a more general Markov process– the constraints that make it a walk are not represented graphically.

Connection to orthogonal polynomials and measures on $\mathbb{R}$

Orthogonal polynomials[1] $\{p_j\}$ result from the Gram-Schmidt algorithm applied to $\{1, x, x^2, \ldots\}$ on the real line with inner-product given with respect to some Borel measure μ, $\langle f, g\rangle = \int_{\mathbb{R}} f g \, d\mu$. These polynomials all have the maximal number of real roots, which are all simple, else they would not change signs enough times to be orthogonal. For the same reason, the roots are all within the convex hull of the support of μ and they interlace as j increases by one. By orthogonality, the polynomials necessarily obey a three-term recurrence relation, which may be written in matrix form as

$$x \begin{bmatrix} p_0 & p_1 & \cdots \end{bmatrix} = \begin{bmatrix} p_0 & p_1 & \cdots \end{bmatrix} \begin{bmatrix} b_0 & c_1 & & \\ a_0 & b_1 & c_2 & \\ & a_1 & \ddots & \ddots \\ & & \ddots & \ddots \end{bmatrix} \tag{9.3}$$

[1]The following results on orthogonal polynomials are well known and included for comparison to the results for quantum walks given below. For details, see[89].

If the polynomials are normalized to have value one at $x = 1$, the tridiagonal, infinite matrix on the right has column sum one for each of its columns. If μ is such that all the entries in that matrix are positive, then it is a stochastic matrix and can be used for the map per time step for a classical random or classical hidden walk on the half-line where the internal states are trivial.

Conversely, given such a walk, then there is an infinite, stochastic matrix

$$\begin{bmatrix} b_0 & c_1 & & \\ a_0 & b_1 & c_2 & \\ & a_1 & \ddots & \ddots \\ & & \ddots & \ddots \end{bmatrix} \tag{9.4}$$

giving the map per time step. If all the a's and c's are strictly positive, then each $n \times n$-truncation A_n of the matrix is similar to a Hermitian matrix via

$$\begin{bmatrix} 1 & & & \\ & \frac{1}{d_1} & & \\ & & \ddots & \\ & & & \frac{1}{d_{n-1}} \end{bmatrix} \begin{bmatrix} b_0 & c_1 & & & \\ a_0 & b_1 & c_2 & & \\ & a_1 & \ddots & \ddots & \\ & & \ddots & \ddots & c_{n-1} \\ & & & a_{n-2} & b_{n-1} \end{bmatrix} \begin{bmatrix} 1 & & & \\ & d_1 & & \\ & & \ddots & \\ & & & d_{n-1} \end{bmatrix} \tag{9.5}$$

where $d_j^2 = \frac{c_1 \cdots c_j}{a_0 \cdots a_{j-1}}$. Therefore, for each such truncation, the eigenvalues $\{x_1^{(n)}, \ldots, x_n^{(n)}\}$ are all real. These eigenvalues are all necessarily less than or equal to one in magnitude since the spectral radius of A_n is less than or equal to its operator norm induced by the ℓ^1-norm, which is one. By the Courant-Fischer minimax theorem [12] the eigenvalues are all simple and they interlace as n increases, so they are actually all less than one in magnitude. Define the polynomials $\{p_n\}$ by

$$p_n(x) = \frac{\det(xI_n - A_n)}{\det(I_n - A_n)} = \frac{(x - x_1^{(n)}) \cdots (x - x_n^{(n)})}{(1 - x_1^{(n)}) \cdots (1 - x_n^{(n)})} \tag{9.6}$$

By adding the first row to the second, the second to the third, and so on, it is easy to see that $\det(I_n - A_n) = a_0 a_1 \cdots a_{n-1}$. Then, by expanding $\det(xI_n - A_n)$ by minors along its last column and evaluation at the n values $x \in \{x_1^{(n)}, \ldots, x_n^{(n)}\}$ (which is enough to determine a degree-n polynomial), the polynomial $x\, p_{n-1}(x)$ obeys the recurrence relation in (9.3).

Let μ_1 be the single atom measure δ_{b_0} and, for $n > 1$, let μ_n be the atomic measure $\sum_{j=1}^{n} w_j^{(n)} \delta_{x_j^{(n)}}$, where

$$w_j^{(n)} = \sum_{l=1}^{k} w_l^{(k)} \frac{p_n\left(x_l^{(k)}\right)}{\left(x_l^{(k)} - x_j^{(n)}\right) p'\left(x_j^{(n)}\right)} \tag{9.7}$$

for any $k \in \left\{\left\lceil\frac{n}{2}\right\rceil, \ldots, n-1\right\}$ (they all give the same result). Furthermore, the sequence of measures $\langle\mu_n\rangle$ stabilizes for any fixed polynomial in the sense that for any degree-n polynomial q, $\int_{\mathbb{R}} q\, d\mu_k$ is the same for all $k \geq \left\lfloor\frac{n}{2}+1\right\rfloor$. Therefore, the μ_n are indeed measures (and not just signed measures) since

$$w_j^{(n)} = \sum_{l=1}^{n-1} w_l^{(n-1)} \left(\frac{p_n\left(x_l^{(n-1)}\right)}{\left(x_l^{(n-1)} - x_j^{(n)}\right) p'\left(x_j^{(n)}\right)} \right)^2 \tag{9.8}$$

so all the w's are positive.

Hence, $\mu_n(\mathbb{R}) = \int_{\mathbb{R}} 1\, d\mu_n = 1$ is the total-variation norm of μ_n for each n. Since the μ_n are all Radon measures and the interval $[-1,1]$ is compact, by the Riesz-Markov theorem [57] and Alaoglu's theorem [64], the sequence $\langle\mu_n\rangle$ has a weak* limit point. Since, by the Weierstrass theorem [82], polynomials are dense in the supremum norm among continuous functions on the compact interval $[-1,1]$, by the above stabilizing property of the sequence, the limit point is unique and the entire sequence converges to it in the weak* topology. Let this limit be denoted μ. Then $\{p_n\}$ are the orthogonal polynomials corresponding to the measure μ on $\mathbb{R}$. The measure μ is unique among Radon measures since any other measure with this property agrees with μ on polynomials, but, as stated above, they are dense in norm among continuous functions on $[-1,1]$, which separate Radon measures; hence, polynomials separate these measures as well.

Furthermore, one may ask if, starting with ν such that the tridiagonal, infinite matrix has all positive entries, then forming the measure μ following the procedure outlined, it is necessarily the case that $\nu \propto \mu$ (there may be a scale factor since μ necessarily has total-variation norm one). This is true since μ and ν agree (up to the scale factor) when integrated with polynomials on $\mathbb{R}$; however, μ is supported on $[-1,1]$ so its moments (and, hence, those of ν) are bounded. Therefore, by [18], the moment problem on $\mathbb{R}$ has a unique solution in this case.

Finally, there is the question of whether every Radon measure on $[-1,1]$ corresponds to a classical random walk. The answer is no, as is seen by the Jacobi polynomials [1], normalized[2] to have the value one at $x=1$, $\left\{\frac{n!}{(1+\alpha)_n} P_n^{(\alpha,\beta)}\right\}$. These have measure with support on $[-1,1]$, given there by the measure absolutely continuous with respect to Lebesgue measure and with Radon-Nikodým derivative $(1+x)^\beta(1-x)^\alpha$. For the recurrence relation, the entries in the tridiagonal, infinite matrix for $\{a_n\}$ and $\{c_n\}$ are always strictly positive (as they must be for any measure with its support on $(-\infty,1]$), but the $\{b_n\}$ are given by

$$b_n = \frac{(\beta^2-\alpha^2)(2n+\alpha+\beta+1)}{(2n+\alpha+\beta)_3} \tag{9.9}$$

which are negative for $\alpha > \beta$. (However, for the Jacobi polynomials shifted to live on the interval $[0,1]$ with Radon-Nikodým derivative $x^\alpha(1-x)^\beta$, $\left\{\frac{(-1)^n n!}{(1+\beta)_n} P_n^{(\alpha,\beta)}(1-2x)\right\}$, there is an associated walk–see [25]).

[2]Using Pochhammer's symbol, $(a)_0 = 1$ and $(a)_n = a(a+1)\cdots(a+n-1)$.

9.2 Quantum walks

Definition of a quantum walk

Similarly to the classical hidden walk, the quantum walk is in some ways a special case of the quantum Markov chain earlier discussed (7.31), while in other ways it is a generalization. The Hilbert space H is required to be either $L^2(\mathcal{J}\times\mathbb{Z}^+;\text{counting measure})$ for a walk on the half-line, or $L^2(\mathcal{J}\times\mathbb{Z};\text{counting measure})$ for a walk on the full line; both are clearly isometrically isomorphic to ℓ^2, so all integrals can be taken to just be sums. The space ℓ^2 has a natural basis of sequences that each have a single nonzero entry with value one. This basis and the inner-product can be used to assign matrix elements to any operator $\mathcal{B}(\ell^2)$. Any operator A in $\mathcal{B}(\ell^2)$ is in one-to-one correspondence to a certain sequence of matrices, $\{A_n\}$, which may be considered the truncations of the infinite matrix corresponding to the operator. By the triangle inequality, these sequences converge to the operator in the strong-operator topology:

$$\begin{aligned}\lim_{n\to\infty}\|A\psi - A_n\psi\| = \lim_{n\to\infty}\|A\psi - P_nAP_n\psi\| &\le \lim_{n\to\infty}\left(\|A\psi - P_nA\psi\| + \|P_nA\psi - P_nAP_n\psi\|\right) \\ &\le \lim_{n\to\infty}\left(\|(A\psi) - P_n(A\psi)\| + \|P_nA\|_{\text{op}}\|\psi - P_n\psi\|\right) = 0\end{aligned} \tag{9.10}$$

for any $\psi\in\ell^2$, where $\{P_n\}$ are the orthogonal projectors onto the span of the first n basis elements.

Conversely, given a sequence of matrices, each of which is the truncation of the following, the condition on the sequence so that it actually corresponds to a bounded operator is that the induced operator norm of all the matrices is bounded. The operator norm induced by the ℓ^2 norm is the largest singular value, which, unfortunately, is not generally readily calculated. However, for this sequence to correspond to a partial isometry, it is only necessary to show that for all fixed, finite collections of columns, those columns of the $\{A_n\}$ are mutually orthonormal in the limit as $n\to\infty$. It is readily shown that this condition implies the induced operator norm of each A_n is less than or equal to one. With this form for the maps, the condition that this is a walk (rather than some other sort of quantum Markov process) is that matrix entries connecting spatial locations that are not neighboring are all zero.

The generalization from the quantum Markov chain is that the last pince-nez in the chain is no longer required to be the same as the preceding. In particular, all the preceding pince-nez can be taken to be simply nodes, so the graphical model is

(9.11)

Note, as for the classical random walks, there is nothing graphically that distinguishes this from a more general quantum Markov process–the constraints that make it a walk are not represented in the graphical model.

Orthogonal trigonometric polynomials and measures on $\mathbb{S}^1$

Trigonometric orthogonal polynomials[3] $\{q_j\}$ result from the Gram-Schmidt algorithm applied to $\{1, z, z^{-1}, z^2, z^{-2}, \ldots\}$ on the unit circle $\mathbb{S}^1$ within $\mathbb{C}$ with sesquilinear inner-product given with respect to some Borel measure μ, $\langle f, g\rangle = \int_{\mathbb{S}^1} f\overline{g}\, d\mu$. To form these, it is useful to start with the monic orthogonal polynomials on the unit circle, the Szegö polynomials, $\{s_n\}$. For any polynomial, define the *reciprocal* polynomial to be the polynomial with its coefficients conjugated and flipped in order, so, if p is a nth order polynomial,

$$p^{\text{reciprocal}}(z) = z^n \overline{p}\left(\frac{1}{z}\right) = z^n \overline{p\left(\frac{1}{\overline{z}}\right)} \tag{9.12}$$

where the overline only over the function means to conjugate its coefficients. On $\mathbb{S}^1$, $p^{\text{reciprocal}}(z) = z^n \overline{p(z)}$. In particular, if z is a root of p, then $\frac{1}{\overline{z}}$ is a root of $p^{\text{reciprocal}}$.

Define the *Verblunsky coefficients* by the values of the Szegö polynomials at zero, $\alpha_n = -\overline{s_{n+1}(0)}$. Then, by orthogonality, we have the following Szegö recurrence identities for all $z \in \mathbb{C}$:

$$\begin{aligned} z\, s_n(z) &= s_{n+1}(z) + \overline{\alpha_n}\, s_n^{\text{reciprocal}}(z) \\ \Leftrightarrow s_{n+1}^{\text{reciprocal}}(z) &= s_n^{\text{reciprocal}}(z) - \alpha_n z\, s_n(z) \end{aligned} \tag{9.13}$$

and

$$\begin{aligned} s_n(z) &= -\overline{\alpha_{n-1}}\, s_n^{\text{reciprocal}}(z) + \left(1 - |\alpha_{n-1}|^2\right) z\, s_{n-1}(z) \\ \Leftrightarrow s_n^{\text{reciprocal}}(z) &= -\alpha_{n-1}\, s_n(z) + \left(1 - |\alpha_{n-1}|^2\right) s_{n-1}^{\text{reciprocal}}(z) \end{aligned} \tag{9.14}$$

From the first identity, it follows that

$$\|s_n\|^2 = \int_{z\in\mathbb{S}^1} |s_n(z)|^2\, d\mu(z) = \left(1 - |\alpha_{n-1}|^2\right) \cdots \left(1 - |\alpha_0|^2\right) \mu(\mathbb{S}^1) \tag{9.15}$$

By Verblunsky's theorem, the measure μ, the moments of the measure $\{m_j\}$, and the Verblunsky coefficients $\{\alpha_j\}$ all determine each other. The only condition on the Verblunsky coefficients that they do indeed correspond to some measure is that $|\alpha_j| \le 1$ for all j.

The monic orthogonal trigonometric polynomials are then given by $q_0 = 1$, and, for $j \in \{1, 2, \ldots\}$,

$$q_j(z) = z^{-(j-1)} s_{2j-1}(z), q_{-j}(z) = z^{-j} s_{2j}^{\text{reciprocal}}(z) \tag{9.16}$$

[3] Only a few basic results that are most applicable to quantum walks of this rich topic are presented here. See [87] for details and elaboration.

The orthogonal trigonometric polynomials necessary satisfy a pentadiagonal recurrence relation, with first $z\,q_0(z) = q_1(z) + \overline{\alpha_0} q_1(z)$, then, using (9.13) and (9.14) repeatedly,

$$\begin{aligned} zq_j(z) &= z^{-j} z^2 s_{2j-1}(z) = z^{-j} z \left(s_{2j}(z) + \overline{\alpha_{2j-1}} s_{2j-1}^{\text{reciprocal}}(z) \right) \\ &= z^{-j} \left(s_{2j+1}(z) + \overline{\alpha_{2j}} s_{2j}^{\text{reciprocal}}(z) + \overline{\alpha_{2j-1}} z\, s_{2j-1}^{\text{reciprocal}}(z) \right) \\ &= z^{-j} \Big(s_{2j+1}(z) + \overline{\alpha_{2j}} s_{2j}^{\text{reciprocal}}(z) \\ &\quad + z\overline{\alpha_{2j-1}} \left(-\alpha_{2j-2}\, s_{2j-1}(z) + \left(1 - |\alpha_{2j-2}|^2\right) s_{2j-2}^{\text{reciprocal}}(z) \right) \Big) \\ &= q_{j+1}(z) + \overline{\alpha_{2j}} q_{-j}(z) - \overline{\alpha_{2j-1}} \alpha_{2j-2} q_j(z) + \overline{\alpha_{2j-1}} \left(1 - |\alpha_{2j-2}|^2\right) q_{-(j-1)}(z) \end{aligned} \tag{9.17}$$

and

$$\begin{aligned} zq_{-j}(z) &= z^{-j} z\, s_{2j}^{\text{reciprocal}}(z) = z^{-j} z \left(-\alpha_{2j-1}\, s_{2j} + \left(1 - |\alpha_{2j-1}|^2\right) s_{2j-1}^{\text{reciprocal}} \right) \\ &= z^{-j} \Big(-\alpha_{2j-1} \left(s_{2j+1} + \overline{\alpha_{2j}} s_{2j}^{\text{reciprocal}} \right) \\ &\quad + z \left(1 - |\alpha_{2j-1}|^2\right) \left(-\alpha_{2j-2}\, s_{2j-1} + \left(1 - |\alpha_{2j-2}|^2\right) s_{2j-2}^{\text{reciprocal}} \right) \Big) \\ &= -\alpha_{2j-1} q_{j+1}(z) - \overline{\alpha_{2j}} \alpha_{2j-1} q_{-j}(z) - \alpha_{2j-2} \left(1 - |\alpha_{2j-1}|^2\right) q_j \\ &\quad + \left(1 - |\alpha_{2j-1}|^2\right) \left(1 - |\alpha_{2j-2}|^2\right) q_{-(j-1)}(z) \end{aligned} \tag{9.18}$$

Following Cantero, Moral, and Velázquez [6] [7], writing this in matrix form as

$$\begin{bmatrix} zq_0(z) & zq_1(z) & zq_{-1}(z) & \cdots \end{bmatrix} = \begin{bmatrix} q_0(z) & q_1(z) & q_{-1}(z) & \cdots \end{bmatrix} Z \tag{9.19}$$

gives the CMV-matrix Z for the monic orthogonal trigonometric polynomials,

$$Z = \begin{bmatrix} \overline{\alpha_0} & \overline{\alpha_1}\left(1 - |\alpha_0|^2\right) & \left(1 - |\alpha_1|^2\right)\left(1 - |\alpha_0|^2\right) & 0 & \cdots \\ 1 & -\overline{\alpha_1}\alpha_0 & -\left(1 - |\alpha_1|^2\right)\alpha_0 & 0 & \cdots \\ 0 & \overline{\alpha_2} & -\alpha_1\overline{\alpha_2} & \overline{\alpha_3}\left(1 - |\alpha_2|^2\right) & \cdots \\ 0 & 1 & -\alpha_1 & -\overline{\alpha_3}\alpha_2 & \cdots \\ 0 & 0 & 0 & \overline{\alpha_4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{9.20}$$

The matrix Z can be written as the product of two block-diagonal matrices (the columns within each 2×2-block correspond to the identities in (9.13) and (9.14))

$$\begin{bmatrix} \overline{\alpha_0} & 1 - |\alpha_0|^2 & & & \\ 1 & -\alpha_0 & & & \\ & & \overline{\alpha_2} & 1 - |\alpha_2|^2 & \\ & & 1 & -\alpha_2 & \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} 1 & & & & \\ & \overline{\alpha_1} & 1 - |\alpha_1|^2 & & \\ & 1 & -\alpha_1 & & \\ & & & \overline{\alpha_3} & \cdots \\ & & & \vdots & \ddots \end{bmatrix} \tag{9.21}$$

Using (9.15), the CMV-matrix U for the orthonormal trigonometric polynomials is then given by

$$\begin{bmatrix} \overline{\alpha_0} & \rho_0 & & & \\ \rho_0 & -\alpha_0 & & & \\ & & \overline{\alpha_2} & \rho_2 & \\ & & \rho_2 & -\alpha_2 & \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} 1 & & & & \\ & \overline{\alpha_1} & \rho_1 & & \\ & \rho_1 & -\alpha_1 & & \\ & & & \overline{\alpha_3} & \cdots \\ & & & \vdots & \ddots \end{bmatrix} \tag{9.22}$$

for $\rho_j = \sqrt{1 - |\alpha_j|^2}$.

Connection to quantum walks–the CMV-matrix

Consider the quantum walk on the half-line $\mathbb{Z}^+ \cup \{0\}$ with internal state set $\mathcal{J} = \{\uparrow, \downarrow\}$, so the Hilbert space H is $\ell^2(\mathbb{Z} \times \{\uparrow, \downarrow\})$. Then, taking the indices as $0\uparrow, 0\downarrow, 1\uparrow, 1\downarrow, 2\uparrow, \ldots$, following a proposal of Grünbaum in [8], the CMV-matrix U above can be used as the map for a quantum walk. With an abuse of notation, for a quantum walk on the line $\mathbb{Z}$ with the same internal state set and with indices $\ldots, -2\downarrow, -1\uparrow, -1\downarrow, 0\uparrow, 0\downarrow, 1\uparrow, 1\downarrow, 2\uparrow, \ldots$, the matrix (underlines indicate indices corresponding spatial location zero) given by the product:

$$\begin{bmatrix} \ddots & & & & & & & \\ & \overline{\alpha_{-2}} & \rho_{-2} & & & & & \\ & \rho_{-2} & -\alpha_{-2} & & & & & \\ & & & \underline{\overline{\alpha_0}} & \underline{\rho_0} & & & \\ & & & \underline{\rho_0} & \underline{-\alpha_0} & & & \\ & & & & & \overline{\alpha_2} & \rho_2 & \\ & & & & & \rho_2 & -\alpha_2 & \\ & & & & & & & \ddots \end{bmatrix} \tag{9.23}$$

$$\begin{bmatrix} \ddots & \vdots & & & & & & \\ \cdots & \alpha_{-3} & & & & & & \\ & & \overline{\alpha_{-1}} & \rho_{-1} & & & & \\ & & \rho_{-1} & \underline{-\alpha_{-1}} & \underline{0} & & & \\ & & & \underline{0} & \underline{\overline{\alpha_1}} & \rho_1 & & \\ & & & & \rho_1 & -\alpha_1 & & \\ & & & & & & \overline{\alpha_3} & \cdots \\ & & & & & & \vdots & \ddots \end{bmatrix}$$

will also be called a CMV-matrix and the $\{\alpha_j\}$ called the Verblunsky coefficients, although there is no longer a connection to the orthogonal trigonometric polynomials.

Coined quantum walks

Once again, consider the quantum walk on the half-line $\mathbb{Z}^+ \cup \{0\}$ with internal state set $\mathcal{J} = \{\uparrow, \downarrow\}$, so the Hilbert space H is $\ell^2(\mathbb{Z} \times \{\uparrow, \downarrow\})$. Then, taking the indices as $0\uparrow, 0\downarrow, 1\uparrow, 1\downarrow, 2\uparrow, \ldots$, if it is either of the form

$$\begin{bmatrix} a_0 & b_0 & & & \\ c_0 & d_0 & & & \\ & & a_2 & b_2 & \\ & & c_2 & d_2 & \\ & & & & \ddots \end{bmatrix} \begin{bmatrix} 1 & & & & \\ & 0 & 1 & & \\ & 1 & 0 & & \\ & & & 0 & \cdots \\ & & & \vdots & \ddots \end{bmatrix} \tag{9.24}$$

or

$$\begin{bmatrix} 1 & & & & \\ & 0 & 1 & & \\ & 1 & 0 & & \\ & & & 0 & \cdots \\ & & & \vdots & \ddots \end{bmatrix} \begin{bmatrix} a_0 & b_0 & & & \\ c_0 & d_0 & & & \\ & & a_2 & b_2 & \\ & & c_2 & d_2 & \\ & & & & \ddots \end{bmatrix} \tag{9.25}$$

then it is termed a *coined walk* with *coins* $\left\{\begin{bmatrix} a_{2j} & b_{2j} \\ c_{2j} & d_{2j} \end{bmatrix}\right\}$, which are unitary matrices. The first form is a CMV-matrix if $b_{2j} = c_{2j}$ is a positive real or zero and if $\overline{a_{2j}} = -d_{2j}$; then all the Verblunsky coefficients with odd index are zero and the Verblunsky coefficients with even index are given by $\alpha_{2j} = -d_{2j}$. The second form is the adjoint of a CMV-matrix if the same conditions hold. The difference between the two forms for a quantum walk is clearly just a matter of transforming the initial state by the unitary matrix

$$\begin{bmatrix} 1 & & & & \\ & 0 & 1 & & \\ & 1 & 0 & & \\ & & & 0 & \cdots \\ & & & \vdots & \ddots \end{bmatrix} \tag{9.26}$$

Therefore, which form is adopted is largely a matter of convention, except for possible restrictions on the initial state.

If all the coins are the same, the quantum walk is termed a coined walk with *constant coin.* Since the overall phase of the wavefunction is irrelevant, the phase of the determinant of the coin is arbitrary. Choosing it to be -1, the coin is necessarily of the form $\begin{bmatrix} \overline{\alpha} & \beta \\ \overline{\beta} & -\alpha \end{bmatrix}$ for some $\alpha, \beta \in \mathbb{C}$ with $|\alpha|^2 + |\beta|^2 = 1$. Then the condition for the unitary matrix for the walk to be a CMV-matrix (or the adjoint of one) is that β is positive real or zero. Similar terminology may be employed for walks on the full line.

9.3 The Parrondo-like paradox for quantum walks

Set-up

Consider the quantum walk on the line $\mathbb{Z}$ with internal state set $\mathcal{J} = \{\uparrow, \downarrow\}$, so the Hilbert space H is $\ell^2(\mathbb{Z} \times \{\uparrow, \downarrow\})$. Let P_+ be the orthogonal projector onto spatial locations with positive index, P_- be the orthogonal projector onto spatial locations with negative index, and P_0 be the orthogonal projector onto spatial location zero, so $P_- + P_0 + P_+$ is the identity. Let U be the unitary operator that gives time evolution for one time step, $\rho \to U\rho U^*$. Then, after n time steps, an observation is made with an observation set $\mathcal{R}$ with σ-algebra $\{\varnothing, A, \tilde{A}, \mathcal{R}\}$ and base measure τ given by $\tau(A) = \tau(\tilde{A}) = 1$. The pince-nez map L is given by

$$L\rho = 1_A \text{tr } \rho P_+ + 1_{\tilde{A}} \text{tr } \rho(P_0 + P_-) \tag{9.27}$$

Then, by considering two different initial states, $\psi\psi^*$ and $\xi\xi^*$, and the geodesic joining them, we can analyze the occurrence and extent of the Parrondo-like paradox.

Showing the paradox is impossible for certain classes of quantum walks

Let $\eta_{0\uparrow}$ be the wavefunction with one for $\uparrow$ at location zero and all other amplitudes zero and $\eta_{0\downarrow}$ be the wavefunction with one for $\downarrow$ at location zero and all other amplitudes zero. Consider the six following cases for the Verblunsky coefficients determining the CMV-matrix U: *(i)* $\alpha_j = \omega^j \alpha_{-j}$ for some $\omega \in \mathbb{S}^1 \subset \mathbb{C}$ and all $j \in \mathbb{Z}$; *(ii)* $\alpha_{2j} = -\omega^{2j}\alpha_{-2j}, \alpha_{2j+1} = \omega^{2j+1}\alpha_{-2j-1}$ for some $\omega \in \mathbb{S}^1 \subset \mathbb{C}$ and all $j \in \mathbb{Z}$; *(iii)* $\alpha_j = \overline{\alpha_{-j}}$ for all $j \in \mathbb{Z}$; *(iv)* $\alpha_j = -\overline{\alpha_{-j}}$ for all $j \in \mathbb{Z}$; *(v)* $\alpha_{2j} = \overline{\alpha_{-2j}}, \alpha_{2j+1} = -\overline{\alpha_{-2j-1}}$ for all $j \in \mathbb{Z}$; and *(vi)* $\alpha_{2j} = -\overline{\alpha_{-2j}}, \alpha_{2j+1} = \overline{\alpha_{-2j-1}}$ for all $j \in \mathbb{Z}$. With the preceding set-up, we have the following result:

Theorem 9.3.1 If the Verblunsky coefficients are in any of the preceding six cases, then for any initial wavefunctions ψ, ξ in the subspace for spatial location zero, $\text{span}_{\mathbb{C}}\{\eta_{0\uparrow}, \eta_{0\downarrow}\}$, the Parrondo-like paradox cannot occur.

Proof Let $\psi^{(n)} = U^n \eta_{0\uparrow}$ and $\xi^{(n)} = U^n \eta_{0\downarrow}$. Then $\xi^{(n)}$ is related to $\psi^{(n)}$ by the following, for each $n \in \{1, \dots\}$ and $j \in \mathbb{Z}$, depending on the case:

$$\begin{array}{lll}
i) & \xi^{(n)}_{j,\uparrow} = -\omega^{j-n}\overline{\psi^{(n)}_{-j,\downarrow}} & \xi^{(n)}_{j,\downarrow} = \omega^{-j-n}\overline{\psi^{(n)}_{-j,\uparrow}} \\
ii) & \xi^{(n)}_{j,\uparrow} = (-\omega)^{j-n}\overline{\psi^{(n)}_{-j,\downarrow}} & \xi^{(n)}_{j,\downarrow} = (-\omega)^{-j-n}\overline{\psi^{(n)}_{-j,\uparrow}} \\
iii) & \xi^{(n)}_{j,\uparrow} = -\psi^{(n)}_{-j,\downarrow} & \xi^{(n)}_{j,\downarrow} = \psi^{(n)}_{-j,\uparrow} \\
iv) & \xi^{(n)}_{j,\uparrow} = \psi^{(n)}_{-j,\downarrow} & \xi^{(n)}_{j,\downarrow} = \psi^{(n)}_{-j,\uparrow} \\
v) & \xi^{(n)}_{j,\uparrow} = (-1)^{j+n+1}\psi^{(n)}_{-j,\downarrow} & \xi^{(n)}_{j,\downarrow} = (-1)^{j+n}\psi^{(n)}_{-j,\uparrow} \\
vi) & \xi^{(n)}_{j,\uparrow} = (-1)^{j+n}\psi^{(n)}_{-j,\downarrow} & \xi^{(n)}_{j,\downarrow} = (-1)^{j+n}\psi^{(n)}_{-j,\uparrow}
\end{array} \tag{9.28}$$

Therefore, in any of the six cases,

$$\operatorname{tr}\begin{bmatrix} L(U^n\eta_{0\uparrow}\eta_{0\uparrow}^*U^{*n})|_A & L(U^n\eta_{0\uparrow}\eta_{0\downarrow}^*U^{*n})|_A \\ L(U^n\eta_{0\downarrow}\eta_{0\uparrow}^*U^{*n})|_A & L(U^n\eta_{0\downarrow}\eta_{0\downarrow}^*U^{*n})|_A \end{bmatrix} = \operatorname{tr}\begin{bmatrix} \operatorname{tr} P_+\psi^{(n)}\psi^{(n)*} & \operatorname{tr} P_+\psi^{(n)}\xi^{(n)*} \\ \operatorname{tr} P_+\xi^{(n)}\psi^{(n)*} & \operatorname{tr} P_+\xi^{(n)}\xi^{(n)*} \end{bmatrix} \tag{9.29}$$

$$= \left\langle (P_+ + P_-)\psi^{(n)}, \psi^{(n)} \right\rangle \le 1$$

Hence, by theorem **8.4.1**, for any initial wavefunctions ψ, ξ in the subspace for spatial location zero and any point on the geodesic joining $[\psi]$ and $[\xi]$, the paradox cannot occur. □

In particular, note this rules out the paradox for the case of constant coin walks of the first form with initial wavefunctions ψ, ξ in the subspace for spatial location zero where the coin is of the form $\begin{bmatrix} \overline{\alpha} & \sqrt{1-|\alpha|^2} \\ \sqrt{1-|\alpha|^2} & -\alpha \end{bmatrix}$ for some $\alpha \in \mathbb{C}$ with $|\alpha|^2 \le 1$. Using a limit theorem by Konno [33] [32], we have the following limiting result for more general constant coin walks (not necessarily in the CMV-matrix form) adopting the second form of coined quantum walks:

Theorem 9.3.2 If a quantum walk has constant coin presented in the second form, then for any initial wavefunctions ψ, ξ in the subspace for spatial location zero, $\operatorname{span}_{\mathbb{C}}\{\eta_{0\uparrow}, \eta_{0\downarrow}\}$, the Parrondo-like paradox cannot occur in the limit as $n \to \infty$.

Proof Adapting the terminology of Konno to our notation, let the coin be given by $\begin{bmatrix} b & a \\ d & c \end{bmatrix}$. The wavefunctions $\eta_{0\uparrow}$ and $\eta_{0\downarrow}$ are given by $\beta = 1, \alpha = 0$ and $\beta = 0, \alpha = 1$ respectively. Then, employing Konno's limit theorem, the limit of the sum of probabilities,

$$\lim_{n\to\infty} \left(\langle P_+U^n\tau_{0\uparrow}, U^n\tau_{0\uparrow}\rangle + \langle P_+U^n\tau_{0\downarrow}, U^n\tau_{0\downarrow}\rangle \right) \tag{9.30}$$

is given by

$$\int_0^{|a|} \frac{2\sqrt{1-|a|^2}}{\pi(1-x^2)\sqrt{|a|^2-x^2}}dx = 1 \tag{9.31}$$

Hence, by theorem **8.4.1**, for any initial wavefunctions ψ, ξ in the subspace for spatial location zero and any point on the geodesic joining $[\psi]$ and $[\xi]$, the paradox cannot occur in the limit as $n \to \infty$. □

Examples of quantum walks displaying the paradox to the maximal extent

Example 9.3.3 With the above set-up, take initial wavefunctions $\psi = \frac{1}{\sqrt{2}}(\eta_{0\uparrow} + \eta_{0\downarrow})$ and $\xi = \frac{1}{\sqrt{2}}(\eta_{0\uparrow} - \eta_{0\downarrow})$. Then, halfway on the minimizing geodesic between them, the initial wavefunction is $\eta_{0\uparrow}$. Take all the Verblunsky coefficients to be zero except for α_{-1}, which has value $\frac{1}{\sqrt{3}}$. Let U be the corresponding CMV-matrix.

Let $\psi^{(n)} = U^n\psi$, $\xi^{(n)} = U^n\xi$, and $\chi^{(n)} = U^n\eta_{0\uparrow}$ for $n \geq 1$. Then $\psi^{(n)}$ has all amplitudes zero except for $\frac{1}{\sqrt{2}}$ for $\downarrow$ at location n, $-\frac{1}{\sqrt{6}}$ for $\downarrow$ at location $n-1$, and $\frac{1}{\sqrt{3}}$ for $\uparrow$ at location $-n$; $\xi^{(n)}$ has all amplitudes zero except for $-\frac{1}{\sqrt{2}}$ for $\downarrow$ at location n, $-\frac{1}{\sqrt{6}}$ for $\downarrow$ at location $n-1$, and $\frac{1}{\sqrt{3}}$ for $\uparrow$ at location $-n$; and $\chi^{(n)}$ has all amplitudes zero except for $\sqrt{\frac{2}{3}}$ for $\uparrow$ at location $-n$ and $-\frac{1}{\sqrt{3}}$ for $\downarrow$ at location $n-1$. Consequently, $P_{A,n} = P'_{A,n} = \frac{2}{3}$ for all $n > 1$, yet $P^{\text{geo}}_{A,n} = \frac{1}{3}$ for the initial wavefuntion halfway on the minimizing geodesic. By theorem **8.3.2**, this example is on the boundary of allowed values of $\left(P_{A,n}, P'_{A,n}, P^{\text{geo}}_{A,n}\right)$ for the paradox.

Example 9.3.4 Again with the above set-up, let U be the matrix for the constant coin walk in the second form with coin $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. Take $\varepsilon > 0$ small. Take $\sigma_1 > 0$ sufficiently small relative to ε such that the normal distribution with mean ε and variance σ_1^2 has neglectfully small measure for $(-\infty, 0)$. Take $a \in \left(-\frac{\pi}{2}, 0\right)$ and $\sigma_2 > 0$ such that: *(i)* the normal distribution with mean a and variance σ_2^2 has neglectfully small measure outside $\left(-\frac{\pi}{2}, 0\right)$ and *(ii)* we have

$$\int_{k\in(-\frac{\pi}{2},0)} \frac{-\sin k}{\sqrt{1+\cos^2 k}}\, d\text{Normal}(a,\sigma_2)(k) = \frac{1}{3} \tag{9.32}$$

Let $\varphi, \zeta : \mathbb{Z} \to \mathbb{C}$ be given by

$$\varphi_j = \sqrt[4]{\frac{2\sigma_1^2}{\pi}} e^{-\sigma_1^2 j^2 + \imath\left(\frac{\pi}{2}-\varepsilon\right)j} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp\left(-\frac{\left(k-\left(\frac{\pi}{2}-\varepsilon\right)\right)^2}{4\sigma_1^2} + \imath jk\right)}{\sqrt[4]{2\pi\sigma_1^2}} dk \tag{9.33}$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \frac{\exp\left(-\frac{\left(k-\left(\frac{\pi}{2}-\varepsilon\right)\right)^2}{4\sigma_1^2} + \imath jk\right)}{\sqrt[4]{2\pi\sigma_1^2}} dk$$

$$\zeta_j = \sqrt[4]{\frac{2\sigma_2^2}{\pi}} e^{-\sigma_2^2 j^2 + \imath a j} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\exp\left(-\frac{(k-a)^2}{4\sigma_2^2} + \imath jk\right)}{\sqrt[4]{2\pi\sigma_2^2}} dk \tag{9.34}$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} \frac{\exp\left(-\frac{(k-a)^2}{4\sigma_2^2} + \imath jk\right)}{\sqrt[4]{2\pi\sigma_2^2}} dk$$

Then, by the inversion of Fourier series,

$$\sum_{j\in\mathbb{Z}}|\varphi_j|^2 = \int_{-\infty}^{\infty}\frac{\exp\left(-\frac{\left(k-\left(\frac{\pi}{2}-\varepsilon\right)\right)^2}{2\sigma_1^2}\right)}{\sqrt{2\pi\sigma_1^2}}dk = 1 \tag{9.35}$$

$$\sum_{j\in\mathbb{Z}}|\zeta_j|^2 = \int_{-\infty}^{\infty}\frac{\exp\left(-\frac{(k-a)^2}{2\sigma_2^2}\right)}{\sqrt{2\pi\sigma_2^2}}dk = 1$$

$$\sum_{j\in\mathbb{Z}}\varphi_j\overline{\zeta_j} = \int_{-\infty}^{\infty}\frac{\exp\left(-\frac{\left(k-\left(\frac{\pi}{2}-\varepsilon\right)\right)^2}{2\sigma_1^2}-\frac{(k-a)^2}{4\sigma_2^2}\right)}{\sqrt[4]{2\pi\sigma_1^2}\sqrt[4]{2\pi\sigma_2^2}}dk \approx 0$$

Hence, if the initial wavefunctions are taken to be $\psi = \frac{1}{2}(\zeta+\varphi)\otimes\begin{bmatrix} \imath \\ 1 \end{bmatrix}$ and $\xi = \frac{1}{2}(\zeta-\varphi)\otimes\begin{bmatrix} \imath \\ 1 \end{bmatrix}$, they will be properly normalized and approximately orthogonal. Midway on the geodesic joining them, we have $\chi = \frac{1}{\sqrt{2}}\zeta\otimes\begin{bmatrix} \imath \\ 1 \end{bmatrix}$.

Now suppose for the observations we always lump the internal states for any spatial sites, so the σ-algebra for the observations is that generated by only the spatial locations. Let $\{Q_B\}$ be the complete set of mutually commuting, orthogonal projectors for this σ-algebra. Define $\kappa:\mathbb{R}\to[-\pi,\pi]$ by

$$\kappa(x) = \begin{cases} \arccos\frac{x}{\sqrt{1-x^2}} & \text{if } x\in\left[-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}}\right] \\ 0 & \text{if } x\in\left[\frac{1}{\sqrt{2}},\infty\right) \\ \pi & \text{if } x\in\left(-\infty,-\frac{1}{\sqrt{2}}\right] \end{cases} \tag{9.36}$$

Then, for each n, let the complete set of commuting projectors $\{R_C^{(n)}\}$ for the Borel σ-algebra on $[0,\pi]$ be given by $R_C^{(n)} = Q_{(n\cdot\kappa^{-1}(C))\cap\mathbb{Z}}$. By Machida's limit theorem [37] [38], the weak*-limit of the measures $\mu^{(n)} = \langle R^{(n)}U^n\psi, U^n\psi\rangle$ exists and is given by

$$\begin{aligned}\mu^{(\infty)} \approx &\frac{1}{2}\text{Normal}\left(\frac{\pi}{2}-\varepsilon,\sigma_2^2\right)+\frac{1}{4}\left(1-\frac{\sin k}{\sqrt{1+\cos^2 k}}\right)\text{Normal}(-a,\sigma_2) \\ &+\frac{1}{4}\left(1+\frac{\sin k}{\sqrt{1+\cos^2 k}}\right)\text{Normal}(\pi+a,\sigma_2)\end{aligned} \tag{9.37}$$

Similarly, the weak*-limit of the measures $\nu^{(n)} = \langle Q^{(n)}U^n\xi, U^n\xi\rangle$ exists and is given by the same expression. For both of these, there is approximately $\frac{1}{2}+\frac{1}{4}\left(1-\frac{1}{3}\right)=\frac{2}{3}$ probability to have $k\in\left[0,\frac{\pi}{2}\right)$, which corresponds to spatial locations with positive index j. The weak*-limit of the measures $\tau^{(n)} = \langle Q^{(n)}U^n\chi, U^n\chi\rangle$ also exists and is given by

$$\tau^{(\infty)} \approx \frac{1}{2}\left(1-\frac{\sin k}{\sqrt{1+\cos^2 k}}\right)\text{Normal}(-a,\sigma_2)+\frac{1}{2}\left(1+\frac{\sin k}{\sqrt{1+\cos^2 k}}\right)\text{Normal}(\pi+a,\sigma_2) \tag{9.38}$$

However, now there is only approximately $\frac{1}{2}\left(1-\frac{1}{3}\right)=\frac{1}{3}$ probability to have $k \in \left[0, \frac{\pi}{2}\right)$. These values of two-thirds for P_A, two-thirds for P'_A, and one-third for P_A^{geo} can be arbitrarily closely approached by taking ε, σ_1, and σ_2 sufficiently small. By theorem **8.3.2**, this example can approach arbitrarily closely to the boundary of allowed values of $(P_A, P'_A, P_A^{\text{geo}})$ for the paradox.

Remark The results of the preceding example also hold for the commonly employed Hadamard coin $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Then the matrix U is the adjoint of a CMV-matrix with all Verblunsky coefficients with even index equal to $\frac{1}{\sqrt{2}}$. For the Hadamard coin, we take initial wavefunctions to be $\psi = \frac{1}{2}(\zeta + \varphi) \otimes \begin{bmatrix} \imath \\ -1 \end{bmatrix}$ and $\xi = \frac{1}{2}(\zeta - \varphi) \otimes \begin{bmatrix} \imath \\ -1 \end{bmatrix}$, so we have $\chi = \frac{1}{\sqrt{2}}\zeta \otimes \begin{bmatrix} \imath \\ -1 \end{bmatrix}$, where

$$\varphi_j = \sqrt[4]{\frac{2\sigma_1^2}{\pi}} e^{-\sigma_1^2 j^2 + \imath\varepsilon j} \tag{9.39}$$

$$\zeta_j = \sqrt[4]{\frac{2\sigma_2^2}{\pi}} e^{-\sigma_2^2 j^2 - \imath(\frac{\pi}{2}+a)j} \tag{9.40}$$

for a, ε, σ_1, and σ_2 as in the given example. The coin utilized in the example was chosen to agree with that used by Machida [37] [38].

Bibliography

[1] M. Abramowitz and I.A. Stegun, eds. *Handbook of Mathematical Functions*. National Bureau of Standards, Applied Mathematics Series 55. Washington: U.S. Government Printing Office, 1964, pp. 773–774.

[2] Y. Aharonov, S. Popescu, and J. Tollaksen. "A time-symmetric formulation of quantum mechanics". In: *Physics Today* (2010).

[3] E. Beckenstein and L. Narici. *Topological Vector Spaces*. Monographs and Textbooks in Pure and Applied Mathematics 95. New York: Marcel Dekker, Inc., 1985, pp. 87–88.

[4] J.S. Bell. "On the Einstein-Podolsky-Rosen Paradox". In: *Physics* 1 (1964).

[5] B.H. Bennet et al. "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels". In: *Physical Review Letters* 70.13 (1993).

[6] M.J. Cantero, L. Moral, and L. Velázquez. "Five-dimensional matrices and zeros of orthogonal polynomials on the unit circle". In: *Linear Algebra Applications* 362 (2003), pp. 29–56.

[7] M.J. Cantero, L. Moral, and L. Velázquez. "Minimal representations of unitary operators and orthogonal polynomials on the unit circle". In: *Linear Algebra Applications* 408 (2005), pp. 40–65.

[8] M.J. Cantero et al. "Matrix valued Szegö polynomials and quantum random walks". In: *Communications on Pure and Applied Mathematics* 63 (2010), pp. 464–507.

[9] A. Cho. "Furtive approach rolls back the limits of quantum uncertainty". In: *Science* (2011), pp. 690–693.

[10] B. Cirel'son. "Quantum Generalizations of Bell's Inequality". In: *Letters un Mathematical Physics* 4 (1980).

[11] J.F. Clauser et al. "Proposed experiment to test local hidden-variable theories". In: *Physical Review Letters* 23 (1969).

[12] J.W. Demmel. *Applied Numerical Linear Algebra*. Society for Industrial and Applied Mathematics 199-200. Philadelphia: American Mathematical Society, 1997.

[13] J. Diestel, J.H. Fourie, and J. Swart. *The Metric Theory of Tensor Products: Grothendieck's Resume Revisited*. Providence: American Mathematical Society, 2008, p. 7.

[14] J. Diestel and J.J. Uhl. *Vector Measures*. Mathematical surveys 15. Providence: American Mathematical Society, 1977, pp. 217–219.

[15] P.A.M. Dirac. *The Principles of Quantum Mechanics*. Oxford: Claredon Press, 1930, p. 35.

[16] P.A.M. Dirac. *The Principles of Quantum Mechanics*. Oxford: Claredon Press, 1930, p. 36.

[17] H.R. Dowson. *Spectral Theory of Linear Operators*. London Mathematics Society Monograms 12. London: Academic Press, Inc., 1978, pp. 50–52.

[18] R. Durrett. *Probability Theory and Examples*. 3rd ed. Belmont: Brooks/Cole, 2005, p. 108.

[19] E.G. Effros and Z.J. Ruan. *Operator Spaces*. London Mathematical Society Monographs–New Series 23. Oxford: Clarendon Press, 2000, p. 25.

[20] W. Feller. *An Introduction to Probability Theory and its Applications*. New York: John Wiley & Sons, Inc., 1950, pp. 430–432.

[21] F.Hausdorff. *Set Theory*. 2nd ed. New York: Chelsea Publishing Company, 1962, pp. 96–108.

[22] F.Hausdorff. *Set Theory*. 2nd ed. New York: Chelsea Publishing Company, 1962, p. 72.

[23] A.P. Flitney, J. Ng, and D. Abbott. "Quantum Parrondo's Games". In: *Physica A* 314 (2002), pp. 35–42.

[24] A.M. Gleason. "Measures on the closed subspaces of a Hilbert space". In: *Journal of Mathematics and Mechanics* 6.6 (1957), pp. 885–893.

[25] F.A. Grünbaum. "An urn model associated with Jacobi polynomials". In: *Communications in Applied Mathematics and Computational Science* 5.1 (2010), pp. 55–63.

[26] F.A. Grünbaum and M. Pejic. "Maximal Parrondo's Paradox for Classical and Quantum Markov Chains". In: *Communications in Mathematical Physics* (2015). submitted.

[27] Jasper Hopkins. *Nicholas of Cusa On Learned Ignorance: A Translation and Appraisal of De Docta Ignorantia*. Minneapolis: The Arthur J. Banning Press, 1981, p. 51.

[28] F.V. Jensen. *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science. New York: Springer-Verlang, 2001, pp. 35–44.

[29] F.V. Jensen. *Bayesian Networks and Decision Graphs*. Statistics for Engineering and Information Science. New York: Springer-Verlang, 2001, pp. 37,65.

[30] S. Kochen and E.P. Specker. "The Problem of Hidden Variables in Quantum Mechanics". In: *Journal of Mathematics and Mechanics* 17.1 (1967), pp. 75–81.

[31] S. Kochen and E.P. Specker. "The Problem of Hidden Variables in Quantum Mechanics". In: *Journal of Mathematics and Mechanics* 17.1 (1967).

[32] N. Konno. “A new type of limit theorem for the one-dimensional quantum random walk”. In: *Journal of the Mathematical Society of Japan* 57 (2005), pp. 1179–1195.

[33] N. Konno. “Limit theorems and absorption problems for quantum random walks in one dimension”. In: *Quantum Information and Computation* 2 (2002). Special issue, pp. 578–595.

[34] K. Kuratowski. *Topology*. Vol. 1.

[35] K. Kuratowski. *Topology*. Vol. 1.

[36] A.D. Lorenzo. “Reassessment of Leggett Inequality”. In: *Foundations of Physics* 43 (2011), pp. 685–698.

[37] T. Machida. “A quantum walk with a delocalized initial state: contribution from a coin-flip operator”. In: *International Journal of Quantum Information* ().

[38] T. Machida. “Realization of the probability laws in the quantum central limit theorems by a quantum walk”. In: *Quantum Information and Computation* 13 (2013), pp. 430–438.

[39] J.J. Martin. *Bayesian Decision Problem and Markov Chains*. New York: John Wiley & Sons, Inc., 1967, p. 62.

[40] N.D. Mermin. *Quantum Computer Science*. Cambridge: Cambridge University Press, 2007, pp. 21–23.

[41] D.A. Meyer. “Noisy quantum Parrondo game”.

[42] J.R. Munkres. *Topology*. 2nd ed. Upper Saddle River: Prentice Hall, 2000, p. 167.

[43] J.R. Munkres. *Topology*. 2nd ed. Upper Saddle River: Prentice Hall, 2000, pp. 182–185.

[44] J.R. Munkres. *Topology*. 2nd ed. Upper Saddle River: Prentice Hall, 2000, pp. 62–74.

[45] J.R. Munkres. *Topology*. 2nd ed. Upper Saddle River: Prentice Hall, 2000, p. 225.

[46] J.R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1997, pp. 40–46.

[47] N.R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 1997, pp. 33–39.

[48] J.C. Oxtoby. “Invariant measures in groups which are not locally compact”. In: *Transactions of the American Mathematical Society* 60.2 (1946), pp. 215–237.

[49] J. Parrondo and P. Espagnol. “Criticism of Feynman’ analysis of the ratchet as an engine”. In: *American Journal of Physics* 64 (1996).

[50] G.K. Pedersen. *C*-algebras and their Automorphism Groups*. London: Academic Press, Inc., 1979, p. 52.

[51] A.L. Peressini. *Ordered Topological Vector Spaces*. Harper’s Series in Modern Mathematics. New York: Harper & Row, 1967, p. 7.

[52] M.M. Rao. *Measure Theory and Integration*. New York: John Wiley & Sons, Inc., 1987, pp. 339–344.

[53] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 276–278.

[54] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 307–309.

[55] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 278.

[56] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 268–270.

[57] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 357.

[58] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 331–335.

[59] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 197–198.

[60] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 201.

[61] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 274.

[62] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 284.

[63] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 248–249.

[64] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 237–238.

[65] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 240.

[66] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 223.

[67] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 25.

[68] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 267.

[69] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 259.

[70] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 53.

[71] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 274.

[72] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, pp. 303–304.

[73] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 200.

[74] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 309.

[75] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 179.

[76] H.L. Royden. *Real Analysis*. 3rd ed. Upper Saddle River: Prentice Hall, 1988, p. 53.

[77] R.A. Ryan. *Introduction to Tensor Products of Banach Spaces*. Springer Monographs in Mathematics. London: Springer-Verlag, 2002.

[78] R.A. Ryan. *Introduction to Tensor Products of Banach Spaces*. Springer Monographs in Mathematics. London: Springer-Verlag, 2002, p. 94.

[79] Shôichirô Sakai. *Markov Chains*. Ergebnisse der Mathematik und ihrer Grenzgebiete 60. New York: Springer-Verlang, 1971, pp. 102–103.

[80] Shôichirô Sakai. *Markov Chains*. Ergebnisse der Mathematik und ihrer Grenzgebiete 60. New York: Springer-Verlang, 1971, p. 48.

[81] Z. Semadeni. *Banach Spaces of Continuous Functions*. Vol. 1.

[82] Z. Semadeni. *Banach Spaces of Continuous Functions*. Vol. 1.

[83] Z. Semadeni. *Banach Spaces of Continuous Functions*. Vol. 1.

[84] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer series in Statistics. New York: Springer-Verlang, 2002, pp. 35–44.

[85] C.E. Shannon and W. Weaver. *The Mathematical Theory of Communication*. Urbana: University of Illinois Press, 1949.

[86] W. Sierpińnski. *Cardinal and Ordinal Numbers*. Monografie Matematyczne 34. Warsaw: Państwowe Wydawnictwo Naukowe.

[87] B. Simon. *Orthogonal Polynomials on the Unit Circle*. Vol. 2. American Mathematical Society Colloquium Publications 54. Providence: American Mathematical Society, 2005.

[88] W.F. Stinespring. "Positive functions on C^*-algebras". In: *Proceedings of the American Mathematical Society* 6.2 (1955), pp. 211–216.

[89] G. Szegö. *Orthogonal Polynomials*. American Mathematical Society Colloquium Publications 23. New York: American Mathematical Society, 1939.

[90] A. Wilansky. *Modern Methods in Topological Vector Spaces*. New York: McGraw-Hill, Inc., 1978, p. 237.

[91] A. Wilansky. *Modern Methods in Topological Vector Spaces*. New York: McGraw-Hill, Inc., 1978, p. 17.

[92] W.K. Wooters and W.H. Zureck. "A single quantum cannot be cloned". In: *Nature* 299 (1982), pp. 802–803.

Appendix A

General propositions

A.1 Banach space propositions

Notation Let A, B,... denote Banach spaces.

Proposition A1.1 Maps in $\mathcal{B}(\mathsf{A},\mathsf{B})$ are continuous in the weak topologies on A and B.

Proof Take any such map L. For any weak neighborhood

$$\mathcal{N}(L\mathsf{a};\phi_1,\ldots,\phi_n;\varepsilon) = \{\mathbf{b}\in\mathsf{B}|\,|\phi_j(\mathbf{b}-L\mathbf{a})|<\varepsilon \text{ for } j\in\{1,\ldots,n\}\} \tag{A.1}$$

with $\mathsf{a}\in\mathsf{A}$, $\phi_1,\ldots,\phi_n\in\mathsf{B}^*$, and $\varepsilon>0$ we have

$$L\left(\mathcal{N}(\mathsf{a};L^*\phi_1,\ldots,L^*\phi_n;\varepsilon)\right) = \mathcal{N}(L\mathbf{a};\phi_1,\ldots,\phi_n;\varepsilon) \qquad \square \tag{A.2}$$

Corollary A1.2 A map in $L\in\mathcal{B}(\mathsf{A},\mathsf{B})$ is uniquely determined by its values on a weakly dense subset.

Comment Starting with a weakly dense subset that is a vector space, its norm closure is necessarily a linear subspace by the properties of Cauchy sequences. Since this subspace is convex, by the separating theorem [65], it cannot be weakly dense unless it is the entire space; hence, a weakly dense subset is necessarily also norm-dense. This argument is not entirely satisfactory since the separating theorem uses the Hahn-Banach theorem [66], which depends on the axiom of choice [67]. This deficiency will be rectified for the situation of interest by proposition **B1.3** below.

The following important proposition is assigned as an exercise in [91].

Proposition A1.3 Let A and B be Banach spaces. A subset V of A that is a vector space uniquely determines a map $L\in\mathcal{B}(\mathsf{A},\mathsf{B})$ if V is dense in the norm topology on A and the operator norm of L restricted to V is bounded. The operator norm of L shares the bound to the operator norm of its restriction.

Proof Since a continuous function is uniquely determined by its values on a weakly dense subset, and since the norm topology is finer than the weak topology, if L exists, it is unique by **A1.2**. For any $\mathsf{a} \in \mathsf{A}$, define $L\mathsf{a}$ as the limit of the Cauchy sequence $\langle L\mathsf{a}_j \rangle_{j=1}^{\infty}$, where $\langle \mathsf{a}_j \rangle_{j=1}^{\infty}$ is any Cauchy sequence converging to **a** composed of elements of V. The choice of Cauchy sequence does not matter because, given another Cauchy sequence $\langle \mathsf{b}_j \rangle_{j=1}^{\infty}$ converging to a composed of elements of V, then $\langle \mathsf{a}_j - \mathsf{b}_j \rangle_{j=1}^{\infty}$ is a Cauchy sequence converging to zero, so $\langle L(\mathsf{a}_j - \mathsf{b}_j) \rangle_{j=1}^{\infty}$ is a Cauchy sequence converging to zero. Since the product of a Cauchy sequence with a scalar is a Cauchy sequence and the sum of two Cauchy sequences is a Cauchy sequence, L is linear. □

A.2 Hilbert space propositions

Notation Let H be a Hilbert space.

Proposition A2.1 Convergence of a countable sum of disjoint ($P_j P_k = 0$ if $j \neq k$), orthogonal projectors, $\sum_j P_j$, to an orthogonal projector P is equivalent in the following four topologies on $\mathcal{B}(\mathsf{H})$: ultrastrong-operator, strong-operator, ultraweak-operator[1], and weak-operator.

Proof The weak-operator topology is courser than the other three, so convergence in any of the others implies convergence in it. Suppose convergence occurs in the weak-operator topology, so for any $\psi, \psi' \in \mathsf{H}$,

$$\lim_{k\to\infty} \left\langle \left(P - \sum_{j=1}^{k} P_j \right) \psi, \psi' \right\rangle = 0 \tag{A.3}$$

Then, for any fixed k,

$$\left\langle P \sum_{j=1}^{k} P_j \psi, \psi' \right\rangle = \lim_{m\to\infty} \left\langle \sum_{l=1}^{m} P_j \sum_{j=1}^{k} P_j \psi, \psi' \right\rangle = \left\langle \sum_{j=1}^{k} P_j \psi, \psi' \right\rangle \tag{A.4}$$

Since ψ, ψ' were arbitrary, $P \sum_{j=1}^{k} P_j = \sum_{j=1}^{k} P_j$. Similarly, $\left(\sum_{j=1}^{k} P_j\right) P = \sum_{j=1}^{k} P_j$. Therefore, $\left(P - \sum_{j=1}^{k} P_j\right)^2 = P - \sum_{j=1}^{k} P_j$, so $P - \sum_{j=1}^{k} P_j$ is itself an orthogonal projector (it is clearly self-adjoint). Then, by taking $\psi = \psi'$, this implies

$$\lim_{k\to\infty} \left\| \left(P - \sum_{j=1}^{k} P_j \right) \psi \right\| = 0 \tag{A.5}$$

so convergence necessarily also occurs in the strong-operator topology. Since the sequence of operators $\left\langle P - \sum_{j=1}^{k} P_j \right\rangle_{k=1}^{\infty}$ is bounded in operator norm (all being projectors), the convergence necessarily occurs in the ultrastrong-operator topology as well. However, the

[1] Same as the weak* topology.

ultrastrong-operator topology is finer than the others, so convergence in it implies convergence in the other three. □

A.3 Transition function propositions

Notation For the following, sets will be denoted $\mathcal{X}$, $\mathcal{Y}$,… and σ-algebras by $\mathcal{E}$, $\mathcal{F}$,…. The spaces of finite, signed-measures on the given set with the given σ-algebra will be denoted $\mathcal{M}(\mathcal{X};\mathcal{E})$, $\mathcal{M}(\mathcal{Y};\mathcal{F})$,…. These are Banach spaces using the total variation norm. Note–by convention product measures are written using $\times$ although they are actually tensor products and should be written using $\otimes$ (see [81] for a use of the latter notation).

Following [52], we have the following:

Definition A3.1 For σ-algebras $\mathcal{E}$ on $\mathcal{X}$ and $\mathcal{F}$ on $\mathcal{Y}$, a function $\tau(\cdot|\cdot) : \mathcal{F} \times \mathcal{X} \to \mathbb{R}$ is a *transition function* if: *(i)* for each $x \in \mathcal{X}$, $\tau(\cdot|x) \in \mathcal{M}(\mathcal{Y};\mathcal{F})$; and *(ii)* for each $B \in \mathcal{F}$, $\tau(B|\cdot)$ is a bounded, $\mathcal{E}$-measurable function on $\mathcal{X}$.

If $\tau(\cdot|\cdot)$ is positive and has the additional property that $\tau(\mathcal{Y}|\cdot) = 1_{\mathcal{X}}$, then it is termed a *transition probability function*. The transition functions for specified ($\mathcal{E}$, $\mathcal{X}$, $\mathcal{F}$, $\mathcal{Y}$) clearly form a vector space. They form a Banach space using the norm

$$\|\tau(\cdot|\cdot)\| = \sup_{x\in\mathcal{X}} \|\tau(\cdot|x)\|_{\text{total variation}} \tag{A.6}$$

A transition function $\tau(\cdot|\cdot)$ with specified data ($\mathcal{E}$, $\mathcal{X}$, $\mathcal{F}$, $\mathcal{Y}$) induces a linear map $L \in \mathcal{B}(\mathcal{M}(\mathcal{X};\mathcal{E}), \mathcal{M}(\mathcal{Y};\mathcal{F}))$ via

$$(L\mu)(B) = \int_{x\in\mathcal{X}} \tau(B|x)\, d\mu(x) \tag{A.7}$$

Not every bounded linear map is induced by a transition function; in general, a *pseudo-transition function* is required (see **B2.7** and **B2.8**). However, we do have the following:

Proposition A3.2 If $L \in \mathcal{B}(\mathcal{M}(\mathcal{X};\mathcal{E}), \mathcal{M}(\mathcal{Y};\mathcal{F}))$ and $K \in \mathcal{B}(\mathcal{M}(\mathcal{Y};\mathcal{F}), \mathcal{M}(\mathcal{Z};\mathcal{G}))$ are both induced by transition functions, then the composition $K \circ L$ is also induced by a transition function.

Proof Let L and K be such maps, with associated transition functions $\tau(\cdot|\cdot)$ and $\nu(\cdot|\cdot)$. Define $\omega(\cdot|\cdot)$ by

$$\omega(C|x) = \int_{y\in\mathcal{Y}} \nu(C|y)\, d(\tau(\cdot|x))(y) \tag{A.8}$$

for $C \in \mathcal{G}$ and $x \in \mathcal{X}$. Then $\omega(\cdot|x) \in \mathcal{M}(\mathcal{Z};\mathcal{G})$. To see that $\omega(C|\cdot)$ is $\mathcal{E}$-measurable, note that, since $\nu(C|\cdot)$ is $\mathcal{F}$-measurable and bounded, there is a sequence of simple functions

$\langle \sum_k b_{jk} 1_{B_{jk}} \rangle$ converging uniformly to it. By the dominated convergence theorem [68], we then have

$$\omega(C|x) = \lim_{j\to\infty} \sum_k b_{jk} \tau(B_{jk}|x) \tag{A.9}$$

By [69], finite sums of measurable functions are measurable and the pointwise limit of a sequence of measurable functions is measurable, so $\omega(C|\cdot)$ is measurable. Hence, $\omega(\cdot|\cdot)$ is a transition function.

Now take any $\mu \in \mathcal{M}(\mathcal{X};\mathcal{E})$ and $C \in \mathcal{G}$. Then, using the preceding results,

$$\begin{aligned} ((K \circ L)(\mu))(C) &= \int_{y\in\mathcal{Y}} \nu(C|y)\, d(L\mu)(y) = \lim_{j\to\infty} \sum_k b_{jk} (L\mu)(B_{jk}) \\ &= \lim_{j\to\infty} \sum_k b_{jk} \int_{x\in\mathcal{X}} \tau(B_{jk}|x)\, d\mu(x) = \int_{x\in\mathcal{X}} \omega(C|x)\, d\mu(x) \end{aligned} \tag{A.10}$$

Hence, $K \circ L$ is induced by the transition function $\omega(\cdot|\cdot)$. □

Proposition A3.3 If $L \in \mathcal{B}(\mathcal{M}(\mathcal{X};\mathcal{E}), \mathcal{M}(\mathcal{Z};\mathcal{G}))$ and $K \in \mathcal{B}(\mathcal{M}(\mathcal{Y};\mathcal{F}), \mathcal{M}(\mathcal{W};\mathcal{H}))$ are both induced by transition functions, then the tensor product map $K \otimes L$ is well-defined and also induced by a transition function.

Proof Let L and K be such maps, with associated transition functions $\tau(\cdot|\cdot)$ and $\nu(\cdot|\cdot)$. Define $\omega(\cdot|\cdot)$ by $\omega(\cdot|x,y) = \tau(\cdot|x) \times \nu(\cdot|y)$ for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Then $\omega(\cdot|x,y)$ is clearly in $\mathcal{M}(\mathcal{X}\times\mathcal{Y};\mathcal{I})$, where $\mathcal{I}$ is the σ-algebra generated by the rectangular subsets $\mathcal{R} = \mathcal{E} \times \mathcal{F}$. The rectangular subsets form a semialgebra[2] Then the finite union of rectangular sets is an algebra.

Following Hausdorff [21], let R_0 be the rectangular subsets $\mathcal{R}$. For each ordinal α, let R_α be the collection of subsets of $\mathcal{X}\times\mathcal{Y}$ that are the countable intersection of subsets from the various collections R_β for ordinals $\beta < \alpha$ if α is even and that are the countable union of subsets from the various collections G_β for ordinals $\beta < \alpha$ if α is odd (where all limit ordinals–those without a predecessor–taken even). For clarification, using the standard notation [70], $R_1 = \mathcal{R}_\sigma$, $R_2 = \mathcal{R}_{\sigma\delta}$, and so on for the finite ordinals.

Fix any ordinal α and suppose that for all subsets B in all collections R_β for $\beta < \alpha$ we have the following property: there is an $\mathcal{I}$-measurable $\omega(B|\cdot)$ such that

$$((L \otimes K)\mu)(B) = \int_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \omega(B|x,y)\, d\mu(x,y) \tag{A.11}$$

for any $\mathcal{I}$-measure μ on $\mathcal{X}\times\mathcal{Y}$. Take any $C \in R_\alpha$. If α is odd, we have a sequence $\langle B_j \rangle$ of subsets from the various R_β with $\beta < \alpha$ such that $C = \bigcup_j B_j$. Let $\omega(C|\cdot) = \bigvee_j \omega(B_j|\cdot)$. Then $\omega(C|\cdot)$ is $\mathcal{I}$-measurable (see [69]) and we can extend $L \otimes K$ by

$$((L \otimes K)\mu)(C) = \int_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \omega(C|x,y)\, d\mu(x,y) \tag{A.12}$$

[2]A collection of sets is a *semialgebra* if it is closed under intersection and the complement of any set is a finite union of sets in the collection.

for any $\mathcal{I}$-measure μ on $\mathcal{X} \times \mathcal{Y}$; hence, since C was arbitrary, R_α has the property. Similarly, if α is even, we have a sequence $\langle B_j \rangle$ of subsets from the various R_β with $\beta < \alpha$ such that $C = \bigcap_j B_j$. Let $\omega(C|\cdot) = \bigwedge_j \omega(B_j|\cdot)$. Then $\omega(C|\cdot)$ is measurable and we can extend $L \otimes K$ by (A.12); hence, R_α also has the property.

However, R_0 has the property–for any $\mathcal{I}$-measure μ, $(L \otimes K)\mu$ is uniquely determined on the rectangular subsets by

$$((L \otimes K)\mu)(A \times B) = \int_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \tau(A|x)\nu(B|y)\, d\mu(x,y) \tag{A.13}$$

and $\omega(A \times B|x,y) = \tau(A|x)\nu(B|y)$ is $\mathcal{I}$-measurable (see [69])–so by transfinite induction [22] [86] all the R_α have the property. Following Kuratowski [34], the σ-algebra $\mathcal{I}$ is given by the union $\bigcup_\alpha R_\alpha$. Therefore, $\omega(\cdot|\cdot)$ is a transition function and induces a well-defined tensor product map $K \otimes L$. □

Comment The preceding proof does not depend on the axiom of choice since the union only needs to be taken up to the ordinal number for the minimal uncountable well-ordered set [21] [34], whose existence does not depend on the axiom of choice [44].

Appendix B

Propositions for option I

B.1 Measures

Notation For the following, sets will be denoted $\mathcal{X}$, $\mathcal{Y}$,... and σ-algebras by $\mathcal{E}$, $\mathcal{F}$,.... The spaces of finite, signed-measures on the given set with the given σ-algebra will be denoted $\mathcal{M}(\mathcal{X};\mathcal{E})$, $\mathcal{M}(\mathcal{Y};\mathcal{F})$,.... These are Banach spaces using the total variation norm.

Definition B1.1 A subset of signed-measures, $M \subset \mathcal{M}(\mathcal{X};\mathcal{E})$, is *absolutely-continuous-complete* if, for any μ in M, all signed-measures absolutely continuous with respect to $|\mu|$ are also in M.

Proposition B1.2 Any absolutely-continuous-complete subset that is a vector space is directed-complete[1].

Proof Let $M \subset \mathcal{M}(\mathcal{X};\mathcal{E})$ be such a subset. Given any upward-directed subset $B \subset M$, bounded by above by some $\rho \in M$, define $\bigvee_{\mu\in B}\mu$ by, for any $E \in \mathcal{E}$,

$$\left(\bigvee_{\mu\in B}\mu\right)(E) = \sup_{\mu\in B}\mu(E) \tag{B.1}$$

This exists since it is bounded from above by $\rho(E)$. It is readily seen that $\bigvee_{\mu\in B}\mu$ is greater than or equal to (using the partial ordering) any $\mu \in B$ and that, given any other $\nu \in M$ with that property, $\bigvee_{\mu\in B}\mu \leq \nu$. Given any countable collection of disjoint, measurable sets, $\{E\}_{j=1}^{\infty} \subset \mathcal{E}$, by the upward-directed property and the countable additivity of each $\mu \in B$,

$$\sum_{j=1}^{n}\left(\bigvee_{\mu\in B}\mu\right)(E_j) + \left(\bigvee_{\mu\in B}\mu\right)\left(\bigcup_{j=n+1}^{\infty}E_j\right) \leq \left(\bigvee_{\mu\in B}\mu\right)\left(\bigcup_{j=1}^{\infty}E_j\right) \leq \sum_{j=1}^{\infty}\left(\bigvee_{\mu\in B}\mu\right)(E_j) \tag{B.2}$$

[1] A partially-ordered subset is *upward-directed* if, given any two elements, there is a third that is greater than or equal to both. A set is *directed-complete* if any bounded, upward-directed subset has a supremum.

for any $n \in \{1,2,\ldots\}$. Since $\nu \le \bigvee_{\mu\in B}\mu \le \rho$ for any $\nu \in B$, and both ρ and ν are countably additive, both $\left(\bigvee_{\mu\in B}\mu\right)\left(\bigcup_{j=n+1}^{\infty} E_j\right) \to 0$ and the tail sum $\sum_{j=n+1}^{\infty}\left(\bigvee_{\mu\in B}\mu\right)(E_j) \to 0$ as $n\to\infty$, so

$$\sum_{j=1}^{\infty}\left(\bigvee_{\mu\in B}\mu\right)(E_j) \le \left(\bigvee_{\mu\in B}\mu\right)\left(\bigcup_{j=1}^{\infty}E_j\right) \le \sum_{j=1}^{\infty}\left(\bigvee_{\mu\in B}\mu\right)(E_j) \tag{B.3}$$

Therefore, $\bigvee_{\mu\in B}\mu$ is countably additive, so it is a signed $\mathcal{E}$-measure. Its total variation norm is bounded by the total variation norm of $\rho \vee (-\nu)$ for any particular $\nu \in B$. Since $\nu \le \bigvee_{\mu\in B}\mu \le \rho$, $\bigvee_{\mu\in B}\mu$ is absolutely continuous with respect to $|\rho| + |\nu|$, so $\bigvee_{\mu\in B}\mu \in M$. □

Comment The following proposition corrects the defect mentioned after **A1.2** for the particular case required.

For the following two propositions, let $\mathcal{G}$ be the σ-algebra generated by the rectangular subsets $\mathcal{E}\times\mathcal{F}$. Let the absolutely-continuous-complete subsets $M \subset \mathcal{M}(\mathcal{X};\mathcal{E})$, $N \subset \mathcal{M}(\mathcal{Y};\mathcal{F})$, and $Q \subset \mathcal{M}(\mathcal{X}\times\mathcal{Y};\mathcal{G})$ also be vector spaces.

Proposition B1.3 If $M\otimes N$ is weakly dense in Q, then it is norm-dense in Q.

Proof Suppose there were some measure $\mu \in C^+$ not in the norm-closure of $M\otimes N$, so there is some $\epsilon > 0$ such that $\|\mu - \nu\|_{\text{total variation}} > \varepsilon$ for all $\nu \in M\otimes N$. By Hahn-decomposition [71] and the absolutely-continuous-completeness property of M and N, if $\nu \in M\otimes N$, then $|\nu| \in M\otimes N$. Let $\mu \perp |\nu|$ be the singular part of μ with respect to $|\nu|$ using Lebesgue decomposition [55]. The set of $\mathcal{G}$-measures $\{\mu \perp |\nu| | \nu \in M\otimes N\}$ is lower-bound by the zero-measure. It is downward-directed since $\mu \perp (|\nu| + |\nu'|)$ is less than or equal to (in the partial ordering) both $\mu\perp|\nu|$ and $\mu\perp|\nu'|$ for any $\nu,\nu' \in M\otimes N$. Hence, by the preceding proposition, the $\mathcal{G}$-measure $\tau = \bigwedge_{\nu\in M\otimes N}\mu\perp|\nu|$ exists. Furthermore, from its definition in the preceding proof, this measure has total-variation norm greater than or equal to ε. Let $\rho\,\|\,\tau$ be the absolutely continuous part of ρ with respect to τ using Lebesgue decomposition. Now consider the bounded, linear functional $\Phi \in \mathcal{M}(\mathcal{X}\times\mathcal{Y};\mathcal{G})^*$ given by $\Phi\rho = (\rho\,\|\,\tau)(\mathcal{X}\times\mathcal{Y})$. This is zero on $M\otimes N$, yet $\Phi\mu = \|\tau\|_{\text{total variation}} \ge \varepsilon$; hence, μ is not in the weak-closure of $M\otimes N$. □

Proposition B1.4 If $M\otimes N$ is norm-dense in Q, then $M^+ \otimes_{\mathbb{R}^+} N^+$ is norm-dense in Q^+.

Proof By Hahn-decomposition and the absolutely-continuous-completeness property of M and N, the measures in $M\otimes N$ are total-variation norm-dense for the measures Q^+. Given any measure $\mu = \sum_j \nu_j\otimes\tau_j \in M\otimes N$, it is absolutely continuous with respect to the product measure $\rho = \left(\sum_j|\nu_j|\right)\otimes\left(\sum_k|\tau_k|\right)$, which is in $M\otimes N$ by Hahn-decomposition and the absolutely-continuous-completeness property of M and N. By the Radon-Nikodým theorem [53], there is a positive function $\frac{d\mu}{d\rho} \in L^1(\mathcal{X}\times\mathcal{Y};\rho)$ such that $\mu = \frac{d\mu}{d\rho}\rho$. By proposition **B3.1** below, $\frac{d\mu}{d\rho}$ can be arbitrarily well approximated in total-variation norm by elements of $L^1\left(\mathcal{X};\sum_j|\nu_j|\right)^+ \otimes_{\mathbb{R}^+} L^1\left(\mathcal{Y};\sum_k|\tau_k|\right)^+$. □

Proposition B1.5 The total-variation norm is a cross-norm for product measures.

Proof Let $\mu \in \mathcal{M}(\mathcal{X};\mathcal{E})$ and $\nu \in \mathcal{M}(\mathcal{Y};\mathcal{F})$. By Hahn decomposition, $|\mu \times \nu| = |\mu| \times |\nu|$, so

$$\|\mu \times \nu\| = |\mu \times \nu|(\mathcal{X} \times \mathcal{Y}) = |\mu|(\mathcal{X})|\nu|(\mathcal{Y}) = \|\mu\|\|\nu\| \qquad \square \tag{B.4}$$

Proposition B1.6 If μ is a σ-finite measure on $\mathcal{X}$, then there is no uncountable collection $\{A_\alpha\}$ of subsets of $\mathcal{X}$ with the properties that $\mu(A_\alpha) > 0$ for all α and $\mu(A_\alpha \cap A_\beta) = 0$ for all $\alpha \neq \beta$.

Proof Suppose otherwise. Since μ is σ-finite, there is a countable, disjoint collection $\{B_j\}$ of subsets of $\mathcal{X}$ such that each B_j has finite μ-measure and their union $\cup_j B_j$ is the entire space $\mathcal{X}$. Now suppose there is some $\varepsilon > 0$ and j such that for infinitely many of the $\{A_\alpha\}$, $\mu(B_j \cap A_\alpha) > \varepsilon$. That immediately contradicts $\mu(B_j)$ being finite. Therefore, it must be the case that for any $\varepsilon > 0$ and j, only finitely many of the $\{A_\alpha\}$ satisfy $\mu(B_j \cap A_\alpha) > \varepsilon$. Let $\{A_k\}$ be the countable collection formed by taking the union over all j and over all $\varepsilon \in \{1, 2^{-1}, 2^{-2}, \ldots\}$ of such $\{A_\alpha\}$. However, this also leads to a contradiction, since for any $A_\beta \notin \{A_k\}$, $\mu(A_\beta \cap B_j) = 0$ for all j, yet $0 < \mu(A_\beta) = \sum_j \mu(A_\beta \cap B_j) = 0$. $\square$

B.2 Maps on subspaces of the space of measures

Let $M \subset \mathcal{M}(\mathcal{X};\mathcal{E})$ be any absolutely-continuous-complete, norm-closed subspace.

Proposition B2.1 Any positive linear map $L \in \mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))$ has

$$\|L\|_{\text{op}} = \sup_{\mu \in M^+, \mu(\mathcal{X}) \leq 1} L\mu(\mathcal{Y}) \tag{B.5}$$

Proof Since L is positive, $|L\mu| = |L(\mu^+ - \mu^-)| = |L(\mu^+) - L(\mu^-)| \leq L(\mu^+) + L(\mu^-) = L|\mu|$ using Hahn decomposition, so

$$\sup_{\mu \in M^+, \mu(\mathcal{X}) \leq 1} L\mu(\mathcal{Y}) \leq \|L\|_{\text{op}} = \sup_{|\mu|(\mathcal{X}) \leq 1} |L\mu|(\mathcal{Y}) \leq \sup_{\mu \in M^+, \mu(\mathcal{X}) \leq 1} L\mu(\mathcal{Y}) \qquad \square \tag{B.6}$$

Proposition B2.2 The space of maps $\mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))$ is a vector lattice[2].

[2] A partially-ordered, vector space A is a *vector lattice* if for any $\mathsf{a}, \mathsf{b} \in \mathsf{A}$, there are elements $\mathsf{a} \vee \mathsf{b}$ and $\mathsf{a} \wedge \mathsf{b}$ such that $\mathsf{a} \vee \mathsf{b}$ is greater than or equal to both a and b, but is less than or equal to any other element with that property and $\mathsf{a} \wedge \mathsf{b}$ is less than or equal to both a and b, but is greater than or equal to any other element with that property.

Proof For any maps $K, L \in \mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))$ define $K \vee L$ by $(K \vee L)(\mu)(F)$ for $\mu \in M^+$ and $F \in \mathcal{F}$ being

$$\sup\left\{\sum_{j=1}^{n}\sum_{k=1}^{m}\max\left\{(K\mu_j)(F_k),(L\mu_j)(F_k)\right\}\;\middle|\;\begin{array}{c} n,m \in \{1,2,\ldots\} \\ \mu_1,\ldots,\mu_n \in M^+ \\ \text{disjoint } F_1,\ldots,F_m \in \mathcal{F} \\ \sum_{j=1}^{n}\mu_j = \mu, \bigcup_{j=1}^{m} F_j = F \end{array}\right\} \tag{B.7}$$

Clearly, $K \vee L \geq K$ and $K \vee L \geq L$. Also, for any $J \in \mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))$ satisfying $J \geq K$ and $J \geq L$, then $J \geq K \vee L$. Also, clearly $(K \vee L)(c\,\cdot) = c(K \vee L)$ for any real scalar $c > 0$. Since M is a Banach lattice[3], it has the Riesz decomposition property [51] that for any $\rho, \nu \in M^+$, any $\mu_1,\ldots,\mu_n \in M^+$ such that $\sum_{j=1}^{n}\mu_j = \rho + \nu$ can be decomposed into $\mu_j = \rho_j + \nu_j$ with both $\rho_j, \nu_j \in M^+$ for each j and with $\sum_{j=1}^{n}\rho_j = \rho$ and $\sum_{j=1}^{n}\nu_j = \nu$. Also, given any partition of F into disjoint, $\mathcal{F}$-measurable subsets, $F = \bigcup_{j=1}^{m_1} G_j = \bigcup_{j=1}^{m_2} H_j$, there is a refinement of both partitions, $F = \bigcup_{j=1}^{m_1}\bigcup_{k=1}^{m_2} G_j \cap H_k$. Therefore, $(K \vee L)(\rho + \nu)(F)$ is equal to

$$\begin{aligned}
&\sup\left\{\sum_{j=1}^{n}\sum_{k=1}^{m}\max\left\{(K\mu_j)(F_k),(L\mu_j)(F_k)\right\}\;\middle|\;\begin{array}{c} n,m \in \{1,2,\ldots\} \\ \mu_1,\ldots,\mu_n \in M^+ \\ \text{disjoint } F_1,\ldots,F_m \in \mathcal{F} \\ \sum_{j=1}^{n}\mu_j = \rho+\nu, \bigcup_{j=1}^{m} F_j = F \end{array}\right\} \\
&= \sup\left\{\begin{array}{c} \sum_{j=1}^{n_1}\sum_{k=1}^{m_1}\max\left\{\begin{array}{c}(K\rho_j)(G_k),\\(L\rho_j)(G_k)\end{array}\right\} \\ +\sum_{j=1}^{n_2}\sum_{k=1}^{m_2}\max\left\{\begin{array}{c}(K\nu_j)(H_k),\\(L\nu_j)(H_k)\end{array}\right\}\end{array}\;\middle|\;\begin{array}{c} n_1,n_2,m_1,m_2 \in \{1,2,\ldots\} \\ \rho_1,\ldots,\rho_{n_1}, \\ \nu_1,\ldots,\nu_{n_2} \in M^+ \\ \text{disjoint } G_1,\ldots,G_{m_1} \in \mathcal{F}, \\ \text{disjoint } H_1,\ldots,H_{m_2} \in \mathcal{F} \\ \sum_{j=1}^{n_1}\rho_j = \rho, \sum_{j=1}^{n_2}\nu_j = \nu \\ \bigcup_{j=1}^{m_1} G_j = \bigcup_{j=1}^{m_2} H_j = F \end{array}\right\} \\
&= (K \vee L)(\rho)(F) + (K \vee L)(\nu)(F)
\end{aligned} \tag{B.8}$$

Since the cone of measures is generating for signed measures by Hahn decomposition and since M has the absolutely-continuous-complete property, $K \vee L$ extends to a linear map on all of M.

It remains to show the image of the map $K \vee L$ is indeed the signed measures. Since the cone of measures is generating, it suffices to show this with any $\mu \in M^+$. Take any countable collection of disjoint, measurable subsets $\{F_j\}_{j=1}^{\infty} \subset \mathcal{F}$. By the preceding argument using Riesz decomposition and refinement of partitions, for any $n \in \{1,2,\ldots\}$,

$$\sum_{j=1}^{n}(K \vee L)(\mu)(F_j) + (K \vee L)(\mu)\left(\bigcup_{j=n+1}^{\infty} F_j\right) \leq (K \vee L)(\mu)\left(\bigcup_{j=1}^{\infty} F_j\right) \leq \sum_{j=1}^{\infty}(K \vee L)(\mu)(F_j) \tag{B.9}$$

[3] A positive cone is *normal* if $0 \leq \mathsf{a} \leq \mathsf{b}$ implies $\|\mathsf{a}\| \leq \|\mathsf{b}\|$. A *Banach lattice* is a complete, normed vector lattice with a normal cone and with $\||\mathsf{a}|\| = \|\mathsf{a}\|$.

However, $K\mu \leq (K \vee L)\mu \leq |K\mu| + |L\mu|$ and both $K\mu$ and $|K\mu| + |L\mu|$ are countably additive, so both the tail of the series, $\sum_{j=n+1}^{\infty}(K \vee L)(\mu)(F_j)$, and $(K \vee L)(\mu)\left(\bigcup_{j=n+1}^{\infty} F_j\right)$ go to zero as $n \to \infty$. Therefore, $(K \vee L)\mu$ is countably additive.

Finally, $K \leq K \vee L \leq |K| + |L|$, so $K \vee L$ is absolutely continuous with respect to $|K| + |L|$. Hence, $K \vee L \in M$. Therefore, $\mathcal{B}(\mathcal{M}(\mathcal{X};\mathcal{E}), \mathcal{M}(\mathcal{Y};\mathcal{F}))$ is a vector lattice. □

Proposition B2.3 With the operator norm induced by the total variation norms, $\mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))$ has a normal cone.

Proof By **B2.1**, it only necessary to consider elements in the positive cone M^+ to calculate the operator norm of any $L \in \mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))^+$. Hence, for any $K \in \mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))^+$ with $K \leq L$, $\|K\|_{\text{op}} \leq \|L\|_{\text{op}}$. □

Proposition B2.4 For any map $L \in \mathcal{B}(M, \mathcal{M}(\mathcal{Y};\mathcal{F}))$, $\||L|\|_{\text{op}} = \|L\|_{\text{op}}$.

Proof For any such L,

$$\|L\|_{\text{op}} = \sup_{\mu \in M, \|\mu\| \leq 1} \|L\mu\| \tag{B.10}$$

which, using Hahn decomposition, is equal to (using $\mu \perp \nu$ to show they are mutually singular)

$$\|L\|_{\text{op}} = \sup\left\{ \|L(\mu - \nu)\| \;\middle|\; \begin{array}{l} \mu, \nu \in M^+ \\ \mu \perp \nu, \|\mu - \nu\| \leq 1 \end{array} \right\} \tag{B.11}$$

Since μ and ν are mutually singular,

$$\|\mu - \nu\| = \|\mu\| + \|\nu\| = \|\mu + \nu\| \tag{B.12}$$

Then, since $|L\mu| < |L|\mu$ and $|L\nu| < |L|\nu$,

$$\|L\|_{\text{op}} \leq \sup\left\{ |L|(\mu + \nu)(\mathcal{Y}) \;\middle|\; \begin{array}{l} \mu, \nu \in M^+ \\ \mu \perp \nu, (\mu + \nu)(\mathcal{Y}) \leq 1 \end{array} \right\} \tag{B.13}$$

which can only be increased by not requiring μ and ν to be mutually singular, so, rewriting $\mu + \nu \to \mu$, $\|L\|_{\text{op}}$ is less than or equal to $\sup_{\mu \in M^+, \|\mu\| \leq 1} |L|(\mu)(\mathcal{Y})$, which, by **B2.1**, is equal to $\||L|\|_{\text{op}}$.

On the other hand, by **B2.1**,

$$\||L|\|_{\text{op}} = \sup_{\mu \in M^+, \|\mu\| \leq 1} |L|(\mu)(\mathcal{Y}) \tag{B.14}$$

which, using $|L| = L \vee (-L)$ and the form of '$\vee$' in (B.7), is equal to

$$\sup\left\{ \sum_{j=1}^{n}\sum_{k=1}^{m} |(L\mu_j)(F_k)| \;\middle|\; \begin{array}{c} n, m \in \{1, 2, \ldots\} \\ \mu_1, \ldots, \mu_n \in M^+, \\ \left\|\sum_{j=1}^{n} \mu_j\right\| \leq 1, \\ \text{disjoint } F_1, \ldots, F_m \in \mathcal{F}, \\ \bigcup_{j=1}^{m} F_j = \mathcal{Y} \end{array} \right\} \tag{B.15}$$

which is less than or equal to

$$\sup\left\{\sum_{j=1}^{n}\|L\mu_j\| \;\middle|\; \begin{array}{c} n \in \{1,2,\dots\} \\ \mu_1,\dots,\mu_n \in M^+, \\ \left\|\sum_{j=1}^{n}\mu_j\right\| \le 1 \end{array}\right\} \tag{B.16}$$

Since, by the AL-space[4] property of M, $\left\|\sum_{j=1}^{n}\mu_j\right\| = \sum_{j=1}^{n}\|\mu_j\|$, this is less than or equal to

$$\sup_{\mu\in M^+,\|\mu\|\le 1}\|L\mu\| \le \sup_{\mu\in M,\|\mu\|\le 1}\|L\mu\| = \|L\|_{\text{op}} \qquad \square \tag{B.17}$$

Comment By **B2.2**, **B2.3**, and **B2.4**, $\mathcal{B}(M,\mathcal{M}(\mathcal{Y};\mathcal{F}))$ is a Banach lattice.

For the following proposition, let $\mathcal{G}$ be the σ-algebra generated by the rectangular subsets $\mathcal{E}\times\mathcal{F}$ and $\mathcal{J}$ be the σ-algebra generated by the rectangular subsets $\mathcal{H}\times\mathcal{I}$. Let the absolutely-continuous-complete subsets $M \subset \mathcal{M}(\mathcal{X};\mathcal{E})$, $N \subset \mathcal{M}(\mathcal{Y};\mathcal{F})$, and $Q \subset \mathcal{M}(\mathcal{X}\times\mathcal{Y};\mathcal{G})$ also be vector spaces.

Proposition B2.5 If $M \otimes N$ is norm-dense in Q, then for any linear maps $L \in \mathcal{B}(M,\mathcal{M}(\mathcal{Z};\mathcal{H}))$ and $K \in \mathcal{B}(N,\mathcal{M}(\mathcal{W};\mathcal{I}))$, the map $L\otimes K : Q \to \mathcal{M}(\mathcal{Z}\times\mathcal{W};\mathcal{J}))$ is well-defined and satisfies $|L\otimes K| = |L|\otimes|K|$ and $\|L\otimes K\|_{\text{op}} = \|L\|_{\text{op}}\|K\|_{\text{op}}$.

Proof Both

$$L\otimes K = (L^+ - L^-)\otimes(K^+ - K^-) = L^+\otimes K^+ - L^+\otimes K^- - L^-\otimes K^+ + L^-\otimes K^- \tag{B.18}$$

and

$$-L\otimes K = -(L^+ - L^-)\otimes(K^+ - K^-) = -L^+\otimes K^+ + L^+\otimes K^- + L^-\otimes K^+ - L^-\otimes K^- \tag{B.19}$$

are clearly less than or equal to (in the partial ordering)

$$|L|\otimes|K| = (L^+ + L^-)\otimes(K^+ + K^-) = L^+\otimes K^+ + L^+\otimes K^- + L^-\otimes K^+ + L^-\otimes K^- \tag{B.20}$$

so $|L\otimes K| \le |L|\otimes|K|$. However, by **B1.4**, **B1.5**, **B2.1**, **B2.4**, and the definition of operator norm, $\||L|\otimes|K|\| \le \|L\|\|K\|$. Hence, by **A1.3**, $L\otimes K$ is well-defined.

However, given any measures $\nu \in M^+$ and $\rho \in N^+$ and subsets $H \in \mathcal{H}$ and $I \in \mathcal{I}$,

$$|L\otimes K|(\nu\times\rho)(H\times I) = \sup\left\{\sum_{j=1}^{n}|(L\otimes K)\mu_j|(H\times I) \;\middle|\; \begin{array}{c} n \in \{1,2,\dots\} \\ \mu_1,\dots,\mu_n \in Q^+ \\ \sum_{j=1}^{n}\mu_j = \nu\times\rho \end{array}\right\} \tag{B.21}$$

[4]A Banach lattice is an *AL-space* if $\|\mathsf{a}+\mathsf{b}\| = \|\mathsf{a}\| + \|\mathsf{b}\|$ for a,b in the positive cone.

using a similar argument to that for the proof of **B2.4**. By Riesz decomposition, this is equal to

$$\sup\left\{\sum_{j=1}^{n}\sum_{k=1}^{m}|(L\otimes K)(\nu_j\times\rho_k)|(H\times I)\;\middle|\;\begin{array}{c}n,m\in\{1,2,\dots\}\\ \nu_1,\dots,\nu_n\in M^+,\rho_1,\dots,\rho_m\in N^+\\ \sum_{j=1}^{n}\nu_j=\nu,\sum_{k=1}^{m}\rho_k=\rho\end{array}\right\}\tag{B.22}$$

which, by the triangle inequality, is greater than or equal to $|(L\nu)\times(K\rho)|(H\times I)$. By the property of signed measures that $|\mu\times\mu'|=|\mu|\times|\mu'|$, this is equal to $|L\nu|(H)|K\rho|(I)$.

The rectangular subsets $\mathcal{H}\times\mathcal{I}$ generate the σ-algebra $\mathcal{J}$, so $|L\otimes K|(\nu\times\rho)(A)\geq((|L|\nu)\times(|K|\rho))(A)$ for any $A\in\mathcal{J}$. By the assumption in the proposition and **A1.3**, $|L\otimes K|(\mu)(A)\leq(|L|\otimes|K|)(\mu)(A)$ for any $\mu\in Q^+$. Hence, $|L\otimes K|\geq|L|\otimes|K|$, so $|L\otimes K|=|L|\otimes|K|$.

We already have $\|L\otimes K\|_{\text{op}}\leq\|L\|_{\text{op}}\|K\|_{\text{op}}$. On the other hand,

$$\begin{aligned}\||L|\|_{\text{op}}\||K|\|_{\text{op}}&=\sup\left\{|L|(\mu)(\mathcal{Z})|K|(\nu)(\mathcal{W})\;\middle|\;\begin{array}{l}\mu\in M^+,\mu(\mathcal{X})\leq1,\\ \nu\in N^+,\nu(\mathcal{Y})\leq1\end{array}\right\}\\ &\leq\sup\left\{(|L|\otimes|K|)(\mu)(\mathcal{Z}\times\mathcal{W})\;\middle|\;\begin{array}{l}\mu\in Q^+\\ \mu(\mathcal{X}\times\mathcal{Y})\leq1\end{array}\right\}=\||L|\otimes|K|\|_{\text{op}}\end{aligned}\tag{B.23}$$

so $\|L\otimes K\|_{\text{op}}=\||L|\otimes|K|\|_{\text{op}}=\||L|\|_{\text{op}}\||K|\|_{\text{op}}=\|L\|_{\text{op}}\|K\|_{\text{op}}$. □

Corollary B2.6 Positive maps $L\in\mathcal{B}(M,\mathcal{M}(\mathcal{Y};\mathcal{F}))^+$ are completely-positive[5].

Proof Use **B1.4**, **B2.5**, and the positivity of both maps L and $I_{\mathcal{B}(N)}$. □

Analogously to the pseudo-functions which compose the doubly-dual space $\mathcal{C}(\mathcal{Z})^{**}$ for $\mathcal{C}(\mathcal{Z})$ the continuous functions on some compact space $\mathcal{Z}$ (see [83]), we have the following:

Definition B2.7 A *pseudo-transition function* $\tau.(\cdot|\cdot)$ with data $(M,\ \mathcal{M}(\mathcal{Y};\mathcal{F}))$ has the properties: *(i)* for each measure $\mu\in M^+$, $\tau_\mu(\cdot|\cdot)$ is a $L^1(\mathcal{X};\mu)$-valued vector $\mathcal{F}$-measure on $\mathcal{Y}$; *(ii)* for each $B\in\mathcal{F}$ and $\mu\in M^+$, $\tau_\mu(B|\cdot)$ is essentially bounded with respect to μ (so it is in $L^\infty(\mathcal{X};\mu)$); and *(iii)* if $\mu\in M^+$ is absolutely continuous with respect to $\nu\in M^+$, then for any $B\in\mathcal{F}$, $\tau_\mu(B|\cdot)$ and $\tau_\nu(B|\cdot)$ differ only on a set of μ-measure zero.

The space of pseudo-functions is clearly a vector space. It is a Banach space under the norm

$$\|\tau.(\cdot|\cdot)\|=\sup_{\mu\in M^+,\|\mu\|\leq1}\|\tau_\mu(\cdot|\cdot)\|\tag{B.24}$$

Proposition B2.8 The space of maps $\mathcal{B}(M,\mathcal{M}(\mathcal{Y};\mathcal{F}))$ is isometrically isomorphic to the space of pseudo-transition functions with data $(M,\ \mathcal{M}(\mathcal{Y};\mathcal{F}))$. Furthermore, this isomorphism takes the positive cones in each space to one another.

[5] A map L is *completely-positive* if $L\otimes I_{\mathcal{B}(N)}$ is positive for every N of the form previously given.

Proof Given such a pseudo-transition function $\tau.(\cdot|\cdot)$, define L by

$$(L\mu)(B) = \int_{x\in\mathcal{X}} \tau_{|\mu|}(B|x)\, d\mu(x) \tag{B.25}$$

for any $B \in \mathcal{F}$ and $\mu \in M$. Then $L\mu$ is indeed a signed measure since it is countably additive because τ is a vector measure. Since $\tau_{|a\mu|}(\cdot|\cdot) = \tau_{|\mu|}(\cdot|\cdot)$ for any $a \in \mathbb{R} \smallsetminus \{0\}$, $L(a\mu) = a\, L\mu$ for any $a \in \mathbb{R}$. Also, for any $\nu \in M$,

$$(L(\mu+\nu))(B) = \int_{x\in\mathcal{X}} \tau_{|\mu|}(B|x)\, d\mu(x) + \int_{x\in\mathcal{X}} \tau_{|\nu|}(B|x)\, d\nu(x) \tag{B.26}$$

$$= \int_{x\in\mathcal{X}} \tau_{|\mu|+|\nu|}(B|x)\, d(\mu+\nu)(x) \tag{B.27}$$

$$= \int_{x\in\mathcal{X}} \tau_{|\mu+\nu|}(B|x)\, d(\mu+\nu)(x) \tag{B.28}$$

Hence, L is linear. To see that it is bounded, we have

$$\|L\|_{\text{op}} = \sup_{\mu\in M, \|\mu\|\le 1} \|L\mu\| = \sup_{\mu\in M, \|\mu\|\le 1} |L\mu|(\mathcal{Y}) \tag{B.29}$$

By Hahn decomposition [71], this is equal to

$$\sup\left\{ (L\mu)(A) - (L\mu)(B) \;\middle|\; \begin{array}{l} \mu \in M, \|\mu\| \le 1, \\ \text{disjoint } A, B \in \mathcal{F} \end{array} \right\} \tag{B.30}$$

which is equal to

$$\sup\left\{ \sum_j ((L\mu)(A_j) - (L\mu)(B_j)) \;\middle|\; \begin{array}{l} \mu \in M, \|\mu\| \le 1, \\ \text{finite, disjoint collections } \{A_j, B_j\} \subset \mathcal{F} \end{array} \right\} \tag{B.31}$$

This is bounded above by

$$\sup\left\{ \sum_j \|\tau_\mu(A_j|\cdot)\| \;\middle|\; \begin{array}{l} \mu \in M^+, \|\mu\| \le 1, \\ \text{finite, disjoint collections } \{A_j\} \subset \mathcal{F} \end{array} \right\} = \sup_{\mu\in M^+, \|\mu\|\le 1} \|\tau_\mu(\cdot|\cdot)\| = \|\tau.(\cdot|\cdot)\| \tag{B.32}$$

Finally, if $\tau.(\cdot|\cdot)$ is positive, L is clearly positive.

Now suppose we are given such a map L. For any measure $\mu \in M^+$, let M_μ be the subspace that is absolutely continuous with respect to μ. By the Radon-Nikodým theorem, M_μ is isometrically isomorphic to $L^1(\mathcal{X};\mu)$. Therefore, the adjoint map $(L|_{M_\mu})^*$ takes $\mathcal{M}(\mathcal{Y};\mathcal{F})^*$ to $L^1(\mathcal{X};\mu)^* \cong L^\infty(\mathcal{X};\mu)$, which is a subspace of $L^1(\mathcal{X};\mu)$ since μ is finite. For each $B \in \mathcal{F}$, define $\tau_\mu(B|\cdot)$ to be $(L|_{M_\mu})^*\Phi_B$, where $\Phi_B \in \mathcal{M}(\mathcal{Y};\mathcal{F})^*$ is the linear functional that evaluates a signed measure on the set B. Then $\tau_\mu(\cdot|\cdot)$ is a $L^1(\mathcal{X};\mu)$-valued vector measure since the countable additivity of μ implies $\tau_\mu(\cdot|\cdot)$ is countably additive. If μ is absolutely continuous with respect to $\nu \in M^+$, then $M_\mu \subset M_\nu$, so $\tau_\mu(B|\cdot)$ differs from $\tau_\nu(B|\cdot)$ only on a μ-measure zero set. Therefore, $\tau.(\cdot|\cdot)$ is a pseudo-transition function. To see that $\tau.(\cdot|\cdot)$ is bounded in norm, we have

$$\begin{aligned} \|\tau.(\cdot|\cdot)\| &= \sup_{\mu\in M^+, \|\mu\|\le 1} \|\tau_\mu(\cdot|\cdot)\| \\ &= \sup\left\{ \sum_j \|\tau_\mu(A_j|\cdot)\| \;\middle|\; \begin{array}{l} \mu\in M^+, \|\mu\|\le 1, \\ \text{finite, disjoint collections } \{A_j\}, \subset \mathcal{F} \end{array} \right\} \\ &\le \sup\left\{ \sum_j \mu(A_j)\, \|\tau_\mu(A_j|\cdot)\|_{L^\infty(\mathcal{X};\mu)} \;\middle|\; \begin{array}{l} \mu\in M^+, \|\mu\|\le 1, \\ \text{finite, disjoint collections } \{A_j\}, \subset \mathcal{F} \end{array} \right\} \end{aligned} \tag{B.33}$$

However,

$$\sum_j \mu(A_j)\, \|\tau_\mu(A_j|\cdot)\|_{L^\infty(\mathcal{X};\mu)} = \sum_j \mu(A_j)\, \left\|(L|_{M_\mu})^*\Phi_{A_j}\right\|_{L^\infty(\mathcal{X};\mu)} \le \mu\left(\bigcup_j A_j\right) \|(L|_{M_\mu})^*\|_{\text{op}} \le \|L\|_{\text{op}} \tag{B.34}$$

so $\|\tau.(\cdot|\cdot)\| \le \|L\|_{\text{op}}$. Lastly, if L is positive, so must be $\tau.(\cdot|\cdot)$; otherwise, if there were some $\mu\in M^+$ and $B\in\mathcal{F}$ such that $\tau_\mu(B|\cdot)$ were strictly less than zero on a set $A\in\mathcal{E}$ with $\mu(A)>0$, then

$$(L(1_A\mu))(B) = \int_A (L|_{M_\mu})^*\Phi_B\, d\mu = \int_A \tau_\mu(A|x)\, d\mu < 0 \tag{B.35}$$

which would be a contradiction. □

It is also possible to define $\tau.(\cdot|\cdot)$ in terms of L rather than adjoints of restrictions of L using the Radon-Nikodým derivative:

$$\tau_\mu(B|\cdot) = (L|_{M_\mu})^*\Phi_B = \frac{d\left(\left((L|_{M_\mu})^*\Phi_B\right)\mu\right)}{d\mu} = \frac{d\mu_B}{d\mu} \tag{B.36}$$

where μ_B is the $\mathcal{E}$-measure on $\mathcal{X}$ given by $\mu_B(A) = (L(1_A\mu))(B)$. Using this, we have the following:

Proposition B2.9 The positive map $L\in\mathcal{B}(M,\mathcal{M}(\mathcal{Y};\mathcal{F}))^+$ is norm-preserving on the positive cone if and only if the associated pseudo-transition function $\tau.(\cdot|\cdot)$ satisfies $\tau_\mu(\mathcal{Y}|\cdot) = 1_\mathcal{X}$ for every measure $\mu\in M^+$ (with equality in the $L^1(\mathcal{X};\mu)$-sense of almost everywhere with respect to μ).

Proof Suppose such a positive map L is norm-preserving on the positive cone; then, for any measure $\mu\in M^+$ and subset $A\in\mathcal{E}$,

$$(L(1_A\mu))(\mathcal{Y}) = (1_A\mu)(\mathcal{X}) = \mu(A) \tag{B.37}$$

Therefore, using the notation of the preceding comment, $\mu_\mathcal{Y} = \mu$, so $\tau_\mu(\mathcal{Y}|\cdot) = \frac{d\mu}{d\mu} = 1_\mathcal{X}$ (in the $L^1(\mathcal{X};\mu)$-sense). Conversely, suppose $\tau.(\cdot|\cdot)$ is such that $\tau_\mu(\mathcal{Y}|\cdot) = 1_\mathcal{X}$ for every measure $\mu\in M^+$ (with equality in the $L^1(\mathcal{X};\mu)$-sense). Then, for any measure $\mu\in M^+$,

$$(L\mu)(\mathcal{Y}) = \int_{x\in\mathcal{X}} \tau(\mathcal{Y}|x)\, d\mu(x) = \mu(X) \tag{B.38}$$

so L is norm-preserving on the positive cone. □

B.3 L^1-spaces

Notation In the following, let $\mathcal{X}$, $\mathcal{Y}$,... denote sets and μ, ν,... denote σ-finite measures. Hilbert spaces, denoted H, J,... are complete, sesquilinear inner-product spaces, with no restriction as to their dimension or separability. $\mathcal{D}(\mathsf{H})$, $\mathcal{D}(\mathsf{J})$,... denote the spaces of density matrices (trace-class, self-adjoint operators) on the specified Hilbert space. These spaces are Banach spaces employing the trace norm. $L^1(\mathcal{X};\mu)$, $L^1(\mathcal{Y};\nu)$,... denote the space of integrable, real-valued functions on the given sets with respect the given measures. These spaces are Banach spaces employing the L^1-norm. $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$, $L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))$,... denote the space of Bochner-integrable, density-matrix-valued functions on the given sets with respect the given measures. These spaces are Banach spaces employing first the trace norm pointwise, then the L^1-norm. For $n \in \{1,2,\ldots\}$, $\mathcal{M}_n$ is the space of $n \times n$-matrices.

Comment The following proposition strengthens the well-known result, which is a special case of a result by Grothendieck [13], that $L^1(\mathcal{X}\times\mathcal{Y};\mu\times\nu) = L^1(\mathcal{X};\mu)\hat{\otimes}L^1(\mathcal{Y};\nu)$, where $\hat{\otimes}$ indicates completion in the projective norm[6].

Proposition B3.1 The finite-nonnegative-tensor-rank[7] functions in $L^1(\mathcal{X}\times\mathcal{Y};\mu\times\nu)^+$, with respect to functions in $L^1(\mathcal{X};\mu)^+$ and $L^1(\mathcal{Y};\nu)^+$, are dense in the norm topology.

Proof Take any $f \in L^1(\mathcal{X}\times\mathcal{Y};\mu\times\nu)^+$. f can be arbitrarily well-approximated in $L^1(\mathcal{X}\times\mathcal{Y};\mu\times\nu)$-norm by simple functions: $\sum_j a_j 1_{A_j}$ for finite collections of positive reals $\{a_j\}$ and finite $\mu\times\nu$-measure subsets $\{A_j\}$. By the construction of product measures (see [72]), each A_j is covered by some finite collection of disjoint, measurable, rectangular subsets $\{B_k\times C_k\}$ with $\mu\times\nu\left(\left(\bigcup_k B_k\times C_k\right)\smallsetminus A_j\right)$ arbitrarily small. Hence, f can be arbitrarily well-approximated in $L^1(\mathcal{X}\times\mathcal{Y};\mu\times\nu)$-norm by simple functions: $\sum_j a_j 1_{B_j\times C_j} = \sum_j a_j 1_{B_j}\otimes 1_{C_j}$ for finite collections of positive reals $\{a_j\}$, finite μ-measure subsets $\{B_j\}$, and finite ν-measure subsets $\{C_j\}$. □

B.4 Density-matrix-valued L^1-spaces

Proposition B4.1 $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))\otimes L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))$ is trace-norm dense within $L^1(\mathcal{X}\times\mathcal{Y};\mu\times\nu;\mathcal{D}(\mathsf{H}\otimes\mathsf{J}))$.

[6] The *projective norm* on $\mathsf{A}\otimes\mathsf{B}$ is the norm induced by duality with Bilinear(A,B), $\|\mathsf{c}\|_\wedge = \inf\sum_j \|\mathsf{a}_j\|\|\mathsf{b}_j\|$, where the infinum is taken over all $\sum_j \mathsf{a}_j\otimes\mathsf{b}_j \in \mathsf{A}\otimes\mathsf{B}$ that equal c.

[7] Using only positive real scalars.

Proof Given operator $\rho \in \mathcal{D}(\mathsf{H} \otimes \mathsf{J})$, by the spectral theorem for compact operators, ρ can be arbitrarily well approximated in trace-norm by sums of the form $\sum_{j=1}^{n} a_j \mathbf{e}_j \otimes \mathbf{e}_j^*$ for some $\{a_j\}_{j=1}^n \subset \mathbb{R}$ and orthonormal collection of vectors $\{\mathbf{e}_j\}_{j=1}^n \subset \mathsf{H} \otimes \mathsf{J}$. By definition, each $\mathbf{e}_j$ can be arbitrarily well approximated in $\mathsf{H} \otimes \mathsf{J}$-norm (so $\mathbf{e}_j \otimes \mathbf{e}_j^*$ will be arbitrarily well-approximated in trace-norm) by sums of the form $\sum_{k=1}^{m} b_k \mathbf{f}_k \otimes \mathbf{g}_k$ for some $b_k \in \mathbb{C}$, $\mathbf{f}_k \in \mathsf{H}$, and $\mathbf{g}_k \in \mathsf{J}$. Using polarization,

$$\left(\sum_{k=1}^{m} b_k \mathbf{f}_k \otimes \mathbf{g}_k\right) \otimes \left(\sum_{\ell=1}^{m} b_\ell \mathbf{f}_\ell \otimes \mathbf{g}_\ell\right)^* \tag{B.39}$$

$$\begin{aligned}
&= \sum_{k,\ell=1}^{m} b_k \overline{b_\ell} \frac{1}{4} \left((\mathbf{f}_k + \mathbf{f}_\ell) \otimes (\mathbf{f}_k + \mathbf{f}_\ell)^* - (\mathbf{f}_k - \mathbf{f}_\ell) \otimes (\mathbf{f}_k - \mathbf{f}_\ell)^* \right.\\
&\quad \left. + \imath (\mathbf{f}_k + \imath \mathbf{f}_\ell) \otimes (\mathbf{f}_k + \imath \mathbf{f}_\ell)^* - \imath (\mathbf{f}_k - \imath \mathbf{f}_\ell) \otimes (\mathbf{f}_k - \imath \mathbf{f}_\ell)^* \right)\\
&\quad \otimes \frac{1}{4} \left((\mathbf{g}_k + \mathbf{g}_\ell) \otimes (\mathbf{g}_k + \mathbf{g}_\ell)^* - (\mathbf{g}_k - \mathbf{g}_\ell) \otimes (\mathbf{g}_k - \mathbf{g}_\ell)^* \right.\\
&\quad \left. + \imath (\mathbf{g}_k + \imath \mathbf{g}_\ell) \otimes (\mathbf{g}_k + \imath \mathbf{g}_\ell)^* - \imath (\mathbf{g}_k - \imath \mathbf{g}_\ell) \otimes (\mathbf{g}_k - \imath \mathbf{g}_\ell)^* \right)\\
&= \sum_{k=1}^{m} |b_k|^2 (\mathbf{f}_k \otimes \mathbf{f}_k^*) \otimes (\mathbf{g}_k \otimes \mathbf{g}_k^*)\\
&\quad + \frac{1}{8} \sum_{k<\ell} \Big(\Re(b_k \overline{b_\ell}) \big(\big((\mathbf{f}_k + \mathbf{f}_\ell) \otimes (\mathbf{f}_k + \mathbf{f}_\ell)^* - (\mathbf{f}_k - \mathbf{f}_\ell) \otimes (\mathbf{f}_k - \mathbf{f}_\ell)^* \big)\\
&\quad \otimes \big((\mathbf{g}_k + \mathbf{g}_\ell) \otimes (\mathbf{g}_k + \mathbf{g}_\ell)^* - (\mathbf{g}_k - \mathbf{g}_\ell) \otimes (\mathbf{g}_k - \mathbf{g}_\ell)^* \big)\\
&\quad + \big((\mathbf{f}_k + \imath \mathbf{f}_\ell) \otimes (\mathbf{f}_k + \imath \mathbf{f}_\ell)^* - (\mathbf{f}_k - \imath \mathbf{f}_\ell) \otimes (\mathbf{f}_k - \imath \mathbf{f}_\ell)^* \big)\\
&\quad \otimes \big((\mathbf{g}_k + \imath \mathbf{g}_\ell) \otimes (\mathbf{g}_k + \imath \mathbf{g}_\ell)^* - (\mathbf{g}_k - \imath \mathbf{g}_\ell) \otimes (\mathbf{g}_k - \imath \mathbf{g}_\ell)^* \big) \big)\\
&\quad - \Im(b_k \overline{b_\ell}) \big(\big((\mathbf{f}_k + \mathbf{f}_\ell) \otimes (\mathbf{f}_k + \mathbf{f}_\ell)^* - (\mathbf{f}_k - \mathbf{f}_\ell) \otimes (\mathbf{f}_k - \mathbf{f}_\ell)^* \big)\\
&\quad \otimes \big((\mathbf{g}_k + \imath \mathbf{g}_\ell) \otimes (\mathbf{g}_k + \imath \mathbf{g}_\ell)^* - (\mathbf{g}_k - \imath \mathbf{g}_\ell) \otimes (\mathbf{g}_k - \imath \mathbf{g}_\ell)^* \big)\\
&\quad + \big((\mathbf{f}_k + \imath \mathbf{f}_\ell) \otimes (\mathbf{f}_k + \imath \mathbf{f}_\ell)^* - (\mathbf{f}_k - \imath \mathbf{f}_\ell) \otimes (\mathbf{f}_k - \imath \mathbf{f}_\ell)^* \big)\\
&\quad \otimes \big((\mathbf{g}_k + \mathbf{g}_\ell) \otimes (\mathbf{g}_k + \mathbf{g}_\ell)^* - (\mathbf{g}_k - \mathbf{g}_\ell) \otimes (\mathbf{g}_k - \mathbf{g}_\ell)^* \big) \big) \Big)
\end{aligned}$$

Then, using the definition of Bochner integrable functions and following the argument in the proof of **B3.1** gives the desired result. □

B.5 Maps on density-matrix-valued L^1-spaces

Proposition B5.1 Any map $L \in \mathcal{B}(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J})))$ satisfies

$$\|L\|_{\mathrm{op}} = \sup \left\{ \|L\rho\| \;\middle|\; \begin{array}{l} \rho \in L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), \|\rho\| \le 1, \\ \rho \text{ is pointwise almost-every-} \\ \text{where rank one} \end{array} \right\}$$

Proof By the definition of Bochner integrable functions and by the spectral theorem for compact operators, simple functions with values in the finite-rank, self-adjoint operators are $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))$-norm dense. Therefore, $\|L\|_{\text{op}}$ is equal to

$$\sup\left\{ \left\| \sum_{j=1}^{n}\sum_{k=1}^{m} \lambda_{jk} L\left(1_{A_j}\mathbf{e}_j^k \otimes \mathbf{e}_j^{k*}\right)\right\| \;\middle|\; \begin{array}{l} n,m \in \{1,2,\ldots\},\text{ disjoint} \\ \mu\text{-measurable subsets } \{A_1,\ldots,A_n\}, \\ \text{collections of orthonormal} \\ \text{elements of } \mathsf{H} \\ \{\{\mathbf{e}_1^1,\ldots,\mathbf{e}_m^1\},\ldots,\{\mathbf{e}_1^n,\ldots,\mathbf{e}_m^n\}\}, \\ \{\lambda_{11},\ldots,\lambda_{nm}\}\subset\mathbb{R}, \\ \sum_{j=1}^n\sum_{k=1}^m |\lambda_{jk}|\mu(A_j)\le 1 \end{array} \right\} \tag{B.40}$$

where $\mathbf{e}_j^{k*}$ is the functional $\langle\cdot,\mathbf{e}_j^k\rangle$. Now fix a value of j, the $A_1,\ldots,A_n$, all the collections of orthonormal vectors, and all the λ's except $\lambda_{j1},\ldots,\lambda_{jm}$. The set of all elements of of $L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J})$ given by $\sum_{j=1}^n\sum_{k=1}^m \lambda_{jk}L\left(1_{A_j}\mathbf{e}_j^k\otimes\mathbf{e}_j^{k*}\right)$ for fixed $\sum_{k=1}^m|\lambda_{jk}|$ is a finite-dimensional, convex subset; hence, its maximum value of norm necessarily occurs at its extreme points where one λ_{jk} is nonzero whereas $\lambda_{j1},\ldots,\widehat{\lambda_{jk}},\ldots,\lambda_{jm}$ are all zero. Since the choice of j was arbitrary, this is true for all j, so the supremum is unchanged by restricting to $m=1$. □

Proposition B5.2 Any positive map $L\in\mathcal{B}(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})),L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))^+$ satisfies

$$\|L\|_{\text{op}} = \sup\left\{ \|L\rho\| \;\middle|\; \begin{array}{l} \rho\in L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))^+,\ \|\rho\|\le 1, \\ \rho \text{ is pointwise almost-every-} \\ \text{where rank one} \end{array}\right\}$$

Proof By the proof of the preceding proposition, $\|L\|_{\text{op}}$ is equal to

$$\sup\left\{ \left\| \sum_{j=1}^{n} \lambda_{j} L\left(1_{A_j}\mathbf{e}_j \otimes \mathbf{e}_j^{*}\right)\right\| \;\middle|\; \begin{array}{l} n \in \{1,2,\ldots\},\text{ disjoint} \\ \mu\text{-measurable subsets}\{A_1,\ldots,A_n\}, \\ \text{unit-norm } \mathbf{e}1,\ldots,\mathbf{e}_n\in\mathsf{H}, \\ \{\lambda_{1},\ldots,\lambda_{n}\}\subset\mathbb{R}, \\ \sum_{j=1}^n |\lambda_{j}|\mu(A_j)\le 1 \end{array} \right\} \tag{B.41}$$

The supremum can only be reduced or stay the same by restricting to positive λ_j's. However, since L is positive, by the triangle inequality and the quasi-AL-property[8] of $L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))$,

$$\left\|\sum_{j=1}^n \lambda_j L\left(1_{A_j}\mathbf{e}_j\otimes\mathbf{e}_j^*\right)\right\| \le \sum_{j=1}^n |\lambda_j|\,\left\|L\left(1_{A_j}\mathbf{e}_j\otimes\mathbf{e}_j^*\right)\right\| = \left\|\sum_{j=1}^n |\lambda_j| L\left(1_{A_j}\mathbf{e}_j\otimes\mathbf{e}_j^*\right)\right\| \tag{B.42}$$

so the supremum can also only be increased or stay the same by restricting to positive λ_j's. Therefore, it must have the same value. □

[8]A Banach space has the *quasi-AL-property* if $\|\mathsf{a}+\mathsf{b}\| = \|\mathsf{a}\| + \|\mathsf{b}\|$ for positive a, b.

Corollary B5.3 The cone of positive maps $\mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)^+$ is a normal cone for the induced operator norm.

Proposition B5.4 If $L \in \mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)$ is completely bounded[9], then, for any space $\mathcal{Z}$, any measure τ, and any Hilbert space K, $\|L \otimes I\|_{\text{op}} \le \|L\|_{\text{matrix}}$ with I the identity map in $\mathcal{B}\left(L^1(\mathcal{Z};\tau;\mathcal{D}(\mathsf{K}))\right)$.

Proof By the definition of operator norm and the definition of the tensor product of maps, $\|L \otimes I\|_{\text{op}}$ is equal to

$$\sup\left\{ \|(L\otimes I)\rho\| \;\middle|\; \begin{array}{l} \text{finite-tensor-rank } \rho \in L^1(\mathcal{X}\times\mathcal{Z};\mu\times\tau;\mathcal{D}(\mathsf{H}\otimes\mathsf{K})) \\ \text{with } \|\rho\| \le 1 \end{array} \right\} \tag{B.43}$$

By the proof of **B5.1**, it is only necessary to take the supremum over simple functions taking values with rank one. By the argument in the proof of **B3.1**, the sets in the simple functions can be restricted to being rectangular. This eliminates consideration of $\mathcal{Z}$ and τ, replacing them with positive, real scalars that can be incorporated into the operators. Finally, rank-one, tensor-rank-n operators live on a n-dimensional subspace of K, which can be identified with $\mathbb{C}^n$. Making the identification $L^1(\text{one point};\text{trivial measure};\mathcal{D}(\mathbb{C}^n)) \leftrightarrow \mathcal{D}(\mathbb{C}^n)$ then gives that $\|L\otimes I\|_{\text{op}}$ is equal to $\sup_{n\le\dim\mathsf{K}} \|L\otimes I_{\mathcal{M}_n}\|_{\text{op}}$. □

Proposition B5.5 If the positive map $L \in \mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)^+$ is such that, for some space $\mathcal{Z}$, some measure τ, and some Hilbert space K, $L\otimes I$ is positive with I the identity map in $\mathcal{B}\left(L^1(\mathcal{Z};\tau;\mathcal{D}(\mathsf{K}))\right)$, then $\|L\otimes I\|_{\text{op}} = \|L\|_{\text{op}}$.

Proof Starting as with the preceding proof, we arrive at the point where $\|L\otimes I\|_{\text{op}}$ is given as

$$\sup\left\{ \left\| \sum_{j=1}^{n} (L\otimes I)(a_j \mathbf{v}_j \otimes \mathbf{v}_j^*\, 1_{A_j} \otimes 1_{B_j}) \right\| \;\middle|\; \begin{array}{l} n \in \{1,2,\ldots\}, \\ \text{finite-tensor-rank} \\ \{\mathbf{v}_1,\ldots,\mathbf{v}_n\} \in \mathsf{H}\otimes\mathsf{K}, \\ \{a_1,\ldots,a_n\} \subset \mathbb{R}, \\ \mu\text{-measurable } \{A_1,\ldots,A_n\}, \\ \tau\text{-measurable } \{B_1,\ldots,B_n\}, \\ \text{with} \sum_{j=1}^n |a_j| \|\mathbf{v}_j\|^2 \mu(A_j)\tau(B_j) \le 1 \end{array} \right\} \tag{B.44}$$

where $\mathbf{v}_j^*$ is the functional $\langle\cdot,\mathbf{v}_j\rangle$. Since $L\otimes I$ is positive, as in the proof for **B5.2**, it is possible to restrict to positive a_j's without changing the result. Then a_j can be combined with $\tau(B_j)$ and both incorporated into a change in the norm of $\mathbf{v}_j$. Furthermore, since $\mathbf{v}_j$ is of finite-tensor-rank, it is necessarily in $\mathsf{H}\otimes\mathsf{L}_j$ for some finite-dimensional subspace $\mathsf{L}_j \subset \mathsf{K}$, so it can be written as $\sum_{k=1}^{\dim \mathsf{L}_j} \mathbf{x}_k^j \otimes \mathbf{e}_k^j$ with $\{\mathbf{e}_k^j\}$ an orthonormal basis for L_j.

[9] L is *completely bounded* if it has finite matrix-norm, $\|L\|_{\text{matrix}} = \sup_n \|L\otimes I_{\mathcal{M}_n}\|_{\text{op}}$.

Then we have for $\|L \otimes I\|_{\text{op}}$,

$$\sup\left\{\sum_{j=1}^{n}\sum_{k=1}^{m}\left\|L(\mathbf{x}_k^j \otimes \mathbf{x}_k^{j*}1_{A_j})\right\| \;\middle|\; \begin{array}{l} n \in \{1,2,\ldots\}, m \in \{1,2,\ldots\}, m \leq \dim \mathsf{K}, \\ \mu\text{-measurable } \{A_1,\ldots,A_n\}, \\ \{\mathbf{x}_j^k\} \subset \mathsf{H}, \sum_{j=1}^{n}\sum_{k=1}^{m}\left\|\mathbf{x}_k^j\right\|^2 \mu(A_j) \leq 1 \end{array}\right\} \tag{B.45}$$

By taking the $\mathbf{x}$'s to be unit length and introducing new, real scalar variables for their squared norm, then by following the argument in the proof of **B5.1**, the supremum is the same if $m = 1$. Hence, $\|L \otimes I\|_{\text{op}}$ is equal to

$$\sup\left\{\sum_{j=1}^{n}\text{tr }\left\|L(\mathbf{x}_j \otimes \mathbf{x}_j^{*}1_{A_j})\right\| \;\middle|\; \begin{array}{l} n \in \{1,2,\ldots\}, \\ \mu\text{-measurable } \{A_1,\ldots,A_n\}, \\ \{\mathbf{x}_j\} \subset \mathsf{H}, \sum_{j=1}^{n}\left\|\mathbf{x}_j\right\|^2 \mu(A_j) \leq 1 \end{array}\right\} \tag{B.46}$$

which is $\|L\|_{\text{op}}$. □

Corollary B5.6 The completely-positive[10] maps

$$\mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)^{\text{cp}}$$

are completely bounded.

Corollary B5.7 The cone of completely-positive maps,

$$\mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)^{\text{cp}}$$

is a normal cone for either the induced operator norm or the matrix norm.

Proposition B5.8 If $L \in \mathcal{B}(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J})))$ is completely positive, then, for any space $\mathcal{Z}$, any measure τ, and any Hilbert space K, $L \otimes I$ is positive, with I the identity map in $\mathcal{B}(L^1(\mathcal{Z};\tau;\mathcal{D}(\mathsf{K})))$.

Proof Since L is completely positive, by **B5.6** it is completely bounded. Hence, $L \otimes I$ exists by **B5.4**. Furthermore, by **B4.1** and **A1.3**, $L \otimes I$ is unique, so it is meaningful to speak of it being positive.

Now suppose there were some space $\mathcal{Z}$, some measure τ, and some Hilbert space K such that $L \otimes I$ were not positive. Then there would be some positive $\rho \in L^1(\mathcal{X}\times\mathcal{Z};\mu\times\tau;\mathcal{D}(\mathsf{H}\otimes\mathsf{K}))^+$ such that $(L \otimes I)\rho$ is not positive. Since the cone $L^1(\mathcal{Y}\times\mathcal{Z};\nu\times\tau;\mathcal{D}(\mathsf{J}\otimes\mathsf{K}))^+$ is norm-closed and $L \otimes I$ is continuous, that implies there is a relatively open neighborhood of ρ in the cone $L^1(\mathcal{X}\times\mathcal{Z};\mu\times\tau;\mathcal{D}(\mathsf{H}\otimes\mathsf{K}))^+$ whose image under $L \otimes I$ does not intersect $L^1(\mathcal{Y}\times\mathcal{Z};\nu\times\tau;\mathcal{D}(\mathsf{J}\otimes\mathsf{K}))^+$.

Now approximating ρ as in the proof of **B5.4**, one finds that for this to occur there must be some $n \in \{1,2,\ldots\}$ for which $L \otimes I_{\mathcal{M}_n}$ is not positive; however, that is a contradiction. □

[10] L is *completely positive* if $L \otimes I_{\mathcal{M}_n}$ is positive for every $n \in \{1,2,\ldots,\}$.

Proposition B5.9 If either $\dim \mathsf{H}$ or $\dim \mathsf{J}$ is finite and if a positive map

$$L \in \mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)^+$$

is such that $L \otimes I_{\mathcal{M}_m}$ is positive for $m = \min\{\dim \mathsf{H}, \dim \mathsf{J}\}$, then L is completely positive.

Proof Clearly, since $L \otimes I_{\mathcal{M}_m}$ is positive, so is $L \otimes I_{\mathcal{M}_n}$ for all $n < m$. Now take $n > m$. $L \otimes I_{\mathcal{M}_n}$ will be positive if for every $\rho \in L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H} \otimes \mathbb{C}^n))^+$, $\mathbf{y} \in \mathsf{J} \otimes \mathbb{C}^n$, and ν-measurable $B \subset \mathcal{Y}$,

$$\int_B \langle((L \otimes I_{\mathcal{M}_n})\rho)\mathbf{y}, \mathbf{y}\rangle_{\mathsf{J}\otimes\mathbb{C}^n}\, d\nu \geq 0 \tag{B.47}$$

By the definition of Bochner integrable functions and the spectral theorem for compact operators, it is enough to show this for ρ that are simple functions with value in the rank-one operators. We then have to consider, for any finite collection of vectors $\{\mathbf{x}_j\} \subset \mathsf{H} \otimes \mathbb{C}^n$ and μ-measurable subsets $\{A_j\}$

$$\sum_j \int_B \langle(L \otimes I_{\mathcal{M}_n})(\mathbf{x}_j \otimes \mathbf{x}_j^*\, 1_{A_j})\mathbf{y}, \mathbf{y}\rangle_{\mathsf{J}\otimes\mathbb{C}^n}\, d\nu \tag{B.48}$$

where $\mathbf{x}_j^*$ is the functional $\langle\cdot, \mathbf{x}_j\rangle$. Writing $\mathbf{y} = \sum_{k=1}^n \mathbf{v}_k \otimes \mathbf{e}_k$ and $\mathbf{x}_j = \sum_{k=1}^n \mathbf{w}_{jk} \otimes \mathbf{e}_k$ for $\{\mathbf{e}_k\}$ an orthonormal basis for $\mathbb{C}^n$ gives (B.48) as

$$\sum_j \sum_{k,l=1}^n \int_B \langle L(\mathbf{w}_{jk} \otimes \mathbf{w}_{jl}^*\, 1_{A_j})\mathbf{v}_l, \mathbf{v}_k\rangle_{\mathsf{J}}\, d\nu \tag{B.49}$$

However, $\sum_{l=1}^n \mathbf{w}_{jl}^* \otimes \mathbf{v}_l$ is of rank at most m, so there are $\{\tilde{\mathbf{w}}_{jl}\}$ and $\{\tilde{\mathbf{v}}_l\}$ such that

$$\sum_{l=1}^n \mathbf{w}_{jl}^* \otimes \mathbf{v}_l = \sum_{l=1}^m \tilde{\mathbf{w}}_{jl}^* \otimes \tilde{\mathbf{v}}_l \tag{B.50}$$

The condition for $L \otimes I_{\mathcal{M}_n}$ to be positive is then that for every ν-measurable $B \subset \mathcal{Y}$, finite collection of μ-measurable subsets $\{A_j\}$, and finite collections of vectors $\{\tilde{\mathbf{w}}_{jk}\} \subset \mathsf{H}$ and $\{\tilde{\mathbf{v}}_k\} \subset \mathsf{J}$,

$$\sum_j \sum_{k,l=1}^m \int_B \langle L(\tilde{\mathbf{w}}_{jk} \otimes \tilde{\mathbf{w}}_{jl}^*\, 1_{A_j})\tilde{\mathbf{v}}_l, \tilde{\mathbf{v}}_k\rangle_{\mathsf{J}}\, d\nu \geq 0 \tag{B.51}$$

However, this condition is independent of n, as long as it is greater than or equal to m. □

Proposition B5.10 The space of completely-bounded maps,

$$\mathcal{CB}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)$$

is a Banach space with respect to the matrix norm.

Proof Let $\langle L_j \rangle$ be a Cauchy sequence in the matrix norm of such maps. Since the matrix-norm is greater than or equal to the operator norm, this is a Cauchy sequence in operator norm, so since the space is a Banach space with respect to the operator norm, it converges to some L_∞ in that norm. It remains to show that L_∞ is completely bounded. For each $n \in \{1, 2, \ldots\}$, by the triangle inequality, $\|K \otimes I_{\mathcal{M}_n}\|_{\text{op}} \leq n^2 \|K\|_{\text{op}}$ for any linear map $K \in \mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)$; hence, the functional $K \to \|K \otimes I_{\mathcal{M}_n}\|_{\text{op}}$ is continuous in the operator-norm topology. Since $\langle L_j \rangle$ converges to L_∞ in this topology, it must be that

$$\|L_\infty \otimes I_{\mathcal{M}_n}\|_{\text{op}} = \lim_{j\to\infty} \|L_j \otimes I_{\mathcal{M}_n}\|_{\text{op}} \leq \lim_{j\to\infty} \|L_j\|_{\text{matrix}} \tag{B.52}$$

The right-hand limit necessarily exists since $\langle L_j \rangle$ is a Cauchy sequence. Therefore, $\|L_\infty\|_{\text{matrix}} \leq \lim_{j\to\infty} \|L_j\|_{\text{matrix}}$, so L_∞ is completely bounded. □

Proposition B5.11 The subset of $\mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)$ for which the tensor product with $I_{\mathcal{M}_n}$ is positive for some fixed $n \in \{1, 2, \ldots\}$ is closed in the weak topology.

Proof We will show the complement is open. Take such a map L that is not in the subset. By the argument in the proof for **B5.9**, that implies there are some ν-measurable $B \subset \mathcal{Y}$, finite collection of μ-measurable subsets $\{A_j\}$, finite collections of vectors $\{\mathbf{w}_{jk}\} \subset \mathsf{H}$ and $\{\mathbf{v}_k\} \subset \mathsf{J}$, and $\varepsilon > 0$ such that

$$\sum_{j=1}^{m} \sum_{k,l=1}^{n} \int_B \langle L(\mathbf{w}_{jk} \otimes \mathbf{w}_{jl}^* \, 1_{A_j})\mathbf{v}_l, \mathbf{v}_k \rangle_{\mathsf{J}} \, d\nu < -\varepsilon \tag{B.53}$$

Then, by the triangle inequality, all the maps in the weak neighborhood

$$\bigcap_{j=1}^{m} \bigcap_{k,l=1}^{n} \mathcal{N}\left(L; \mathbf{w}_{jk}\mathbf{w}_{jk}^* \, 1_{A_j}; \mathbf{v}_l \otimes \mathbf{v}_l^* \, 1_B; \frac{\varepsilon}{2n^4 m}\right) \tag{B.54}$$

$$+ \bigcap_{j=1}^{m} \bigcap_{k=1}^{n} \bigcap_{r<l} \mathcal{N}\left(L; (\mathbf{w}_{jr} \otimes \mathbf{w}_{jl}^* + \mathbf{w}_{jl} \otimes \mathbf{w}_{jr}^*) \, 1_{A_j}; \mathbf{v}_k \otimes \mathbf{v}_k^* \, 1_B; \frac{\varepsilon}{n^4 m}\right)$$

$$+ \bigcap_{j=1}^{m} \bigcap_{k=1}^{n} \bigcap_{r<l} \mathcal{N}\left(L; \mathbf{w}_{jk} \otimes \mathbf{w}_{jk}^* \, 1_{A_j}; (\mathbf{v}_k \otimes \mathbf{v}_m^* + \mathbf{v}_m \otimes \mathbf{v}_k^*) \, 1_B; \frac{\varepsilon}{n^4 m}\right)$$

$$+ \bigcap_{j=1}^{m} \bigcap_{k<l} \bigcap_{q<r} \mathcal{N}\left(L; (\mathbf{w}_{jk} \otimes \mathbf{w}_{jl}^* + \mathbf{w}_{jl} \otimes \mathbf{w}_{jk}^*) \, 1_{A_j}; (\mathbf{v}_q \otimes \mathbf{v}_r^* + \mathbf{v}_r \otimes \mathbf{v}_q^*) \, 1_B; \frac{2\varepsilon}{n^4 m}\right)$$

will also fail to yield a positive tensor product with $I_{\mathcal{M}_n}$. □

Corollary B5.12 The cone of completely positive maps is weakly closed in $\mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)$.

Comment The preceding result also follows from showing that the spaces are norm-closed, then using the separating theorem to argue that the weak and norm topologies have the same closed, convex subsets. The approach followed here is preferable since, as has already been noted, the separating theorem depends on the axiom of choice through the Hahn-Banach theorem. For the case where the measures are trivial, the space is a dual space, $\mathcal{D}(\mathsf{H}) = \mathcal{K}(\mathsf{H})^*$, so it is possible to do better.

Proposition B5.13 The subset of $\mathcal{B}(\mathcal{D}(\mathsf{H}), \mathcal{D}(\mathsf{J}))$ for which the tensor product with $I_{\mathcal{M}_n}$ is positive for some fixed $n \in \{1, 2, \ldots\}$ is closed in the weak* topology.

Proof Following the proof of **B5.9**, if $L \otimes I_{\mathcal{M}_n}$ is not positive, then there are some finite collections of vectors $\{\mathbf{w}_k\} \subset \mathsf{H}$ and $\{\mathbf{v}_k\} \subset \mathsf{J}$, and $\varepsilon > 0$ such that all the maps in the weak* neighborhood

$$\begin{aligned} &\bigcap_{k,l=1}^{n} \mathcal{N}\left(L; \mathbf{w}_k \mathbf{w}_k^* 1_{A_j}; \mathbf{v}_l \otimes \mathbf{v}_l^* 1_B; \frac{\varepsilon}{2n^4}\right) \\ &+ \bigcap_{k=1}^{n} \bigcap_{r<l} \mathcal{N}\left(L; (\mathbf{w}_r \otimes \mathbf{w}_l^* + \mathbf{w}_l \otimes \mathbf{w}_r^*) 1_{A_j}; \mathbf{v}_k \otimes \mathbf{v}_k^* 1_B; \frac{\varepsilon}{n^4}\right) \\ &+ \bigcap_{k=1}^{n} \bigcap_{r<l} \mathcal{N}\left(L; \mathbf{w}_k \otimes \mathbf{w}_k^* 1_{A_j}; (\mathbf{v}_k \otimes \mathbf{v}_m^* + \mathbf{v}_m \otimes \mathbf{v}_k^*) 1_B; \frac{\varepsilon}{n^4}\right) \\ &+ \bigcap_{k<l} \bigcap_{q<r} \mathcal{N}\left(L; (\mathbf{w}_k \otimes \mathbf{w}_l^* + \mathbf{w}_l \otimes \mathbf{w}_k^*) 1_{A_j}; (\mathbf{v}_q \otimes \mathbf{v}_r^* + \mathbf{v}_r \otimes \mathbf{v}_q^*) 1_B; \frac{2\varepsilon}{n^4}\right) \end{aligned} \tag{B.55}$$

will also fail to yield a positive tensor product with $I_{\mathcal{M}_n}$. □

Corollary B5.14 The cone of completely positive maps in $\mathcal{B}(\mathcal{D}(\mathsf{H}), \mathcal{D}(\mathsf{J}))$ is closed in the weak* topology.

Proposition B5.15 Any bounded, positively-linear map L to $L^1(\mathcal{Y}; \nu; \mathcal{D}(\mathsf{J}))$ that is given on the positive cone of $L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H}))$ extends uniquely by linearity to a map $L \in \mathcal{B}(L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H})), L^1(\mathcal{Y}; \nu; \mathcal{D}(\mathsf{J})))^+$.

Proof Let L be any such map. Extend L to $\mathcal{B}(L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H})), L^1(\mathcal{Y}; \nu; \mathcal{D}(\mathsf{J})))^+$ by

$$L(\rho) = L\left(\frac{|\rho| + \rho}{2}\right) - L\left(\frac{|\rho| - \rho}{2}\right) \text{ for any } \rho \in L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H})) \tag{B.56}$$

where $|\cdot|$ is applied pointwise with $|\rho|(x) = |\rho(x)| = \sqrt{\rho(x)^2}$. It is readily seen that positive linearity implies the extended L is now linear over $\mathbb{R}$. Furthermore, by **B5.2**, this extension does not increase the operator norm.

Proposition B5.16 Any map $L \in \mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))\right)^+$ extends uniquely by linearity to a map $L \in \mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{S}_1(\mathsf{J}))\right)^+$. This extension has operator norm less than twice that of the restricted map.

Proof Let L be any such map. Extend L to $L \in \mathcal{B}(L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{S}_1(\mathsf{J}))^+)$ by using Cartesian decomposition,

$$L(\rho) = L\left(\frac{\rho+\rho^*}{2}\right) + \imath L\left(\frac{\rho-\rho^*}{2\imath}\right) \text{ for any } \rho \in L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})) \tag{B.57}$$

where $*$ is applied pointwise with $\rho^*(x) = \rho(x)^*$. It is readily seen that real linearity implies the extended L is now linear over $\mathbb{C}$. By definition, this extended map is a positive one. The triangle inequality implies the extension has operator norm less than twice that of the restricted map. □

Comment Morally, one would hope for $L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H}))^*$ to be $L^\infty(\mathcal{X};\mu;\mathcal{B}(\mathsf{H}))$, which has the structure of an algebra. That is almost correct, as the following proposition indicates.

Definition B5.17 Let $\pi : L^\infty(\mathcal{X};\mu) \to \mathcal{B}(L^2(\mathcal{X};\mu))$ be defined by $\pi(f)g$ being the pointwise multiplication fg.

Definition B5.18 Let $\mathcal{W}(\mathcal{X};\mu;\mathsf{H})$ be the subset of $\mathcal{B}(L^2(\mathcal{X};\mu)\otimes\mathsf{H})$ that is the bicommutant[11] of the C^*-algebra $\pi(L^\infty(\mathcal{O};\mu))\otimes\mathcal{B}(\mathsf{H})$ considered as a subset of $\mathcal{B}\left(L^2(\mathcal{X};\mu)\otimes\mathsf{H}\right)$.

By Von Neumann's bicommutant theorem [80], $\mathcal{W}(\mathcal{X};\mu;\mathsf{H})$ is a von Neumann algebra.

Proposition B5.19 There is a weak*-continuous, isometric bijection
$\eta : L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H}))^* \to \mathcal{W}(\mathcal{X};\mu;\mathsf{H})$.

Proof Using **A1.3**, since finite-tensor-rank elements are dense in $L^2(\mathcal{X};\mu)\otimes\mathsf{H}$, define η by

$$\left\langle \eta(\Phi)\left(\sum_{j=1}^{l} f_j\otimes\psi_j\right), \sum_{k=1}^{m} f'_k\otimes\psi'_k \right\rangle = \sum_{j=1}^{l}\sum_{k=1}^{m} \Phi\left(\psi_j{\psi'_k}^* f_j\overline{f'_k}\right) \tag{B.58}$$

for every $\{\psi_j\},\{\psi'_k\}\subset\mathsf{H}$, $\{f_j\},\{f'_k\}\subset L^2(\mathcal{X};\mu)$, and $\Phi\in L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H}))^*$. Using the definition of Bochner integrability and the density of finite rank operators in $\mathcal{S}_1(\mathsf{H})$, by **A1.3** it is only necessary to specify the inverse, ζ^{-1}, on simple functions of finite rank:

$$\eta^{-1}(w)\left(\sum_{j=1}^{n}\sum_{k=1}^{m}\psi_{jk}{\psi'}^*_{jk}1_{B_j}\right) = \sum_{j=1}^{n}\sum_{k=1}^{m}\left\langle w(1_{B_j}\otimes\psi_{jk}), 1_{B_j}\otimes\psi'_{jk}\right\rangle \tag{B.59}$$

[11] The *commutant* of a subset of an algebra is composed of all elements of the algebra that commute with the subset. The *bicommutant* is the commutant of the commutant.

for every $\{\psi_{jk}\}, \{\psi'_{jk}\} \subset \mathsf{H}$, collection $\{B_j\}$ of disjoint, μ-measurable subsets of $\mathcal{X}$, and $w \in \mathcal{W}(\mathcal{X};\mu;\mathsf{H})$. To see the inverse is well-defined, it is only necessary to check two cases: firstly, that

$$\eta^{-1}(w)\left(1_B \sum_{k=1}^{m} \psi_k {\psi'}_k^{*}\right) = \eta^{-1}(w)\left(1_B \sum_{k=1}^{l} \xi_k {\xi'}_k^{*}\right) \tag{B.60}$$

if $\sum_{k=1}^{m} \psi_k {\psi'}_k^{*} = \sum_{k=1}^{l} \xi_k {\xi'}_k^{*}$; and, secondly, that

$$\eta^{-1}(w)\left(1_{B_1} \sum_{k=1}^{m} \psi_k {\psi'}_k^{*}\right) + \eta^{-1}(w)\left(1_{B_2} \sum_{k=1}^{m} \psi_k {\psi'}_k^{*}\right) = \eta^{-1}(w)\left(1_{B_1 \cup B_2} \sum_{k=1}^{m} \psi_k {\psi'}_k^{*}\right) \tag{B.61}$$

if B_1, B_2 are disjoint. The first is obvious since $\psi\psi'^* \to \psi \otimes \psi'^* \to \langle w(1_B \otimes \psi), 1_B \otimes \psi'\rangle$ is a linear map. The second follows from

$$\begin{aligned}\langle w(1_{B_1} \otimes \psi), 1_{B_2} \otimes \psi'\rangle &= \langle(\pi(1_{B_1}) \otimes I_{\mathsf{H}})w(1_{B_1} \otimes \psi), 1_{B_2} \otimes \psi'\rangle \\ &= \langle w(1_{B_1} \otimes \psi), (\pi(1_{B_1}) \otimes I_{\mathsf{H}})(1_{B_2} \otimes \psi')\rangle = 0\end{aligned} \tag{B.62}$$

To show η is weak*-continuous, let $\Theta : \mathcal{S}_1\left(L^2(\mathcal{X};\mu) \otimes \mathsf{H}\right) \to L^1\left(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})\right)$ be defined, using **A1.3**, by its value for finite-rank, finite-tensor-rank elements, which are dense in $\mathcal{S}_1\left(L^2(\mathcal{X};\mu) \otimes \mathsf{H}\right)$,

$$\Theta\left(\sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{r}(f_{jk} \otimes \psi_{jk})(f'_{jl} \otimes \psi'_{jl})^*\right) = \sum_{j=1}^{n}\sum_{k=1}^{m}\sum_{l=1}^{r} \psi_{jk}{\psi'}_{jl}^{*} f_{jk}\overline{f'_{jl}} \tag{B.63}$$

The restriction of Θ to this dense subset is positive–in the sense that it takes $\mathcal{D}\left(L^2(\mathcal{X};\mu) \otimes \mathsf{H}\right)^+$ into $L^1\left(\mathcal{X};\mu;\mathcal{D}(\mathsf{H})\right)^+$–and has operator norm at most two by an argument similar to that for the proofs of **B5.15** and **B5.16** since it is an isometry on the positive cone. Then, for any $\rho \in \mathcal{S}_1\left(L^2(\mathcal{X};\mu) \otimes \mathsf{H}\right)$ and $\Phi \in L^1\left(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})\right)^*$, tr $\eta(\Phi)\rho = \Phi(\Theta(\rho))$. Therefore, defining $\eta_* : \mathcal{S}_1\left(L^2(\mathcal{X};\mu) \otimes \mathsf{H}\right)/\ker\Theta \to L^1\left(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})\right)$ by $\eta_*(\rho + \ker\Theta) = \Theta(\rho)$, we have $(\eta_*)^* = \eta$ and $\mathcal{W}(\mathcal{X};\mu;\mathsf{H}) \cong \left(\mathcal{S}_1\left(L^2(\mathcal{X};\mu) \otimes \mathsf{H}\right)/\ker\Theta\right)^*$. By the properties of bounded, linear maps on Banach spaces, η_* is invertible since η is and $\eta^{-1} = \left((\eta_*)^*\right)^{-1} = \left((\eta_*)^{-1}\right)^*$, so η^{-1} is also weak*-continuous.

To show η is isometric, take any $\Phi \in L^1\left(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})\right)^*$ and start with

$$\|\Phi\|_{\mathrm{op}} = \sup\left\{\Phi\left(\sum_{j=1}^{m}\rho_j 1_{B_j}\right) \,\middle|\, \begin{array}{l}\{\rho_j\} \subset \mathcal{S}^1(\mathsf{H}),\ \text{disjoint } \{B_j\} \subset \mathcal{E} \\ \left\|\sum_{j=1}^{m}\rho_j 1_{B_j}\right\| \le 1\end{array}\right\} \tag{B.64}$$

Fixing μ-measurable $B \subset \mathcal{X}$, $\rho \to \Phi(\rho 1_B)$ is a linear functional on $\mathcal{S}^1(\mathsf{H})$; hence, there is some $L \in \mathcal{B}(\mathsf{H})$ such that $\Phi(\rho 1_B) = \mathrm{tr}\ L\rho$ for every $\rho \in \mathcal{S}^1(\mathsf{H})$. Using the well-known property of operators on Hilbert spaces (which follows from the Cauchy-Schwarz inequality) that

$$\|K\|_{\mathrm{op}} = \sup_{\|\psi\|,\|\psi'\| \le 1} \langle K\psi, \psi'\rangle \tag{B.65}$$

then

$$\sup_{\|\rho\|_{\text{trace}}\le 1} \text{tr } L\rho = \|L\|_{\text{op}} = \sup_{\|\psi\|,\|\psi'\|\le 1} \langle L\psi, \psi'\rangle = \sup_{\|\psi\|,\|\psi'\|\le 1} \text{tr } \left((L\psi)\psi'^*\right) \tag{B.66}$$

Hence, the supremum in (B.64) need only be taken over collections of rank-one operators $\{\psi_j \psi'^*_j\}$. Since $\|\psi\psi'^*\|_{\text{trace}} = \|\psi\|\|\psi'\|$, the condition that

$$\left\|\sum_{j=1}^m \psi_j \psi'^*_j 1_{B_j}\right\| = \sum_{j=1}^m \|\psi_j\|\|\psi'_j\|\mu(B_j) \le 1 \tag{B.67}$$

is then implied by the pair of conditions

$$\left\|\sum_{j=1}^m 1_{B_j} \otimes \psi_j\right\| = \sum_{j=1}^m \|\psi_j\|^2 \mu(B_j) \le 1 \text{ and } \left\|\sum_{j=1}^m 1_{B_j} \otimes \psi'_j\right\| = \sum_{j=1}^m \|\psi'_j\|^2 \mu(B_j) \le 1 \tag{B.68}$$

using the Cauchy-Schwartz inequality. Therefore, a lower bound to $\|\Phi\|_{\text{op}}$ is provided by

$$\sup\left\{\left\langle \eta(\Phi)\left(\sum_{j=1}^m 1_{B_j}\otimes\psi_j\right), \sum_{k=1}^m 1_{B_k}\otimes\psi'_k\right\rangle \,\middle|\, \begin{array}{l} \{\psi_j\}, \{\psi'_j\} \subset \mathsf{H} \\ \text{disjoint } \mu\text{-measurable } \{B_j\} \\ \left\|\sum_{j=1}^m 1_{B_j}\otimes\psi_j\right\| \le 1 \\ \left\|\sum_{j=1}^m 1_{B_j}\otimes\psi'_j\right\| \le 1 \end{array}\right\} \tag{B.69}$$

Since the simple functions are dense in $L^2(\mathcal{O};\mu)$ and finite-tensor-rank elements are dense in $L^2(\mathcal{O};\mu)\otimes\mathsf{H}$, this is simply $\|\zeta(\Phi)\|_{\text{op}}$.

On the other hand, for the collections $\{\psi_j\psi'^*_j\}$ it is certainly possible to add the additional constraint that $\|\psi_j\| = \|\psi'_j\|$ for each j. Then (B.67) and (B.68) are equivalent, so $\|\Phi\|_{\text{op}}$ is equal to (B.69) with the additional constraint. However, then $\|\eta(\Phi)\|_{\text{op}}$ is certainly an upper bound to $\|\Phi\|_{\text{op}}$; hence, $\|\Phi\|_{\text{op}} = \|\eta(\Phi)\|_{\text{op}}$. □

Remark The preceding proposition allows $L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H}))^*$ to be given an algebraic structure, with product $\Phi\Psi = \eta^{-1}(\eta(\Phi)\eta(\Psi))$, conjugate $\Phi^\star = \eta^{-1}(\eta(\Phi)^*)$, and unit $\eta^{-1}\left(I_{L^2(\mathcal{X};\mu)\otimes\mathsf{H}}\right)$ being the functional $\rho \to \int_{\mathcal{X}} \text{tr}_{\mathsf{H}}\rho\, d\mu$ (in other words, the functional that sends $\rho \to \|\rho\|$ for ρ in the positive cone).

If a map is positive, its adjoint map is of course positive as well. Hence if a map is completely-positive, then its adjoint map is completely-positive as well (with the definition we use). The following shows that this holds as well with the definition employed in Stinespring's theorem [88].

Proposition B5.20 If $L \in \mathcal{B}\left(L^1(\mathcal{X};\mu;\mathcal{S}_1(\mathsf{H})), L^1(\mathcal{Y};\nu;\mathcal{S}_1(\mathsf{J}))\right)$ is completely-positive, then the adjoint map L^* satisfies

$$\sum_{j,k} \langle \eta\circ L^*(\Phi_j\Phi_k^\star)\psi_k, \psi_j\rangle \ge 0$$

for every finite collection $\{\psi_j\} \subset L^2(\mathcal{X};\mu)\otimes\mathsf{H}$ and $\{\Phi_j\} \subset L^1(\mathcal{Y};\nu;\mathcal{S}_1(\mathsf{J}))^*$.

Proof Employing the notation of the preceding proof, we have, for any $n \in \{1, 2, \ldots\}$,

$$\sum_{j,k=1}^{n} \langle \eta \circ L^*(\Phi_j \Phi_k^\star)\psi_k, \psi_j \rangle = \sum_{j,k=1}^{n} (\Phi_j \Phi_k^\star) L \circ \Theta(\psi_k \otimes \psi_j^*) \tag{B.70}$$

Let $\{\mathbf{e}_j\}$ be an orthonormal basis for $\mathbb{C}^n$. Then let $\tilde{\psi} \in L^2(\mathcal{X};\mu) \otimes \mathsf{H} \otimes \mathbb{C}^n$ be $\sum_{j=1}^n \psi_j \otimes \mathbf{e}_j$ and $\tilde{\Phi} \in L^1(\mathcal{Y};\nu;\mathcal{S}_1(\mathsf{J} \otimes \mathbb{C}^n))^*$ be $\sum_{j=1}^n \Phi_j \otimes (\mathbf{e}_1 \otimes \mathbf{e}_j^*)^*$. Then

$$\tilde{\Phi}\tilde{\Phi}^\star = \sum_{j,k=1}^{n} (\Phi_j \Phi_k^\star) \otimes (\mathbf{e}_k \otimes \mathbf{e}_j^*)^* \tag{B.71}$$

so (B.70) is equal to

$$(\tilde{\Phi}\tilde{\Phi}^\star)(L \otimes I_{\mathcal{M}_n}) \circ \Theta(\tilde{\psi} \otimes \tilde{\psi}^*) \tag{B.72}$$

with the positive map Θ appropriately adjusted to the larger spaces. Since L is completely positive, $L \otimes I_{\mathcal{M}_n}$ is positive, so this quantity is clearly greater than or equal to zero. □

Proposition B5.21 The Hilbert space[12] $L^2(\mathcal{O};\nu) \otimes \mathcal{S}_2(\mathsf{J}) \cong L^2(\mathcal{O};\nu;\mathcal{S}_2(\mathsf{J}))$ is relatively dense within $\mathcal{W}(\mathcal{O};\nu;\mathsf{J})$ in the ultrastrong-operator topology.

Proof For any finite-dimensional subspace $\mathsf{G} \subset \mathsf{J}$, let p_{G} be orthonormal projection from J to G and let π be defined as in **B5.17**. The net of operators, $\langle \pi(1_B) \otimes p_{\mathsf{G}} \rangle$, over all finite-dimensional $\mathsf{G} \subset \mathsf{H}$ and finite ν-measure $B \subset \mathcal{O}$, ordered by inclusion, converges to the identity operator in the ultrastrong-operator topology. These operators are all in $\mathcal{W}(\mathcal{O};\nu;\mathsf{J})$, so for any $w \in \mathcal{W}$, the net $\langle \pi(1_B) \otimes p_{\mathsf{G}} \circ w \circ \pi(1_B) \otimes p_{\mathsf{G}} \rangle$ converges to w in the ultrastrong-operator topology, with each operator in $\mathcal{W} \cap L^2(\mathcal{O};\nu) \otimes \mathcal{S}_2(\mathsf{J})$. □

B.6 Vector measures

Notation The spaces of finite-norm, vector measures on the given set with the given σ-algebra will be denoted $\mathcal{M}(\mathcal{X};\mathcal{E};\mathsf{A})$, $\mathcal{M}(\mathcal{Y};\mathcal{F};\mathsf{B})$,.... These are Banach spaces using the total variation norm (give reference). The subset of these that is atomic will be denoted $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{A})$,....

Proposition B6.1 The subset $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{B})$ is a closed, linear subspace of the Banach space $\mathcal{M}(\mathcal{X};\mathcal{E};\mathsf{B})$ (hence, it is itself a Banach space).

[12] $\mathcal{S}_2$ denotes the Hilbert-Schmidt operators.

Proof The subset $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{B})$ is clearly a linear subspace. To show it is closed, take any μ in the complement of $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{B})$. Suppose that for any $\varepsilon > 0$ there were some countable collection of atoms $\{A_j^\varepsilon\} \subset \mathcal{E}$ such that $\|\mu\left(\mathcal{X} \smallsetminus \bigcup_j A_j^\varepsilon\right)\| < \varepsilon$. Then the countable union

$$B = \bigcup_{\varepsilon \in \{1, 2^{-1}, 2^{-3}, \ldots\}} \bigcup_j A_j^\varepsilon \tag{B.73}$$

would satisfy $\|\mu\left(\mathcal{X} \smallsetminus B\right)\| = 0$, which contradicts μ being in the complement. Hence, there is some $\varepsilon > 0$ such that, for any countable collection of atoms $\{A_j\} \subset \mathcal{E}$, $\|\mu\left(\mathcal{X} \smallsetminus \bigcup_j A_j\right)\| > \varepsilon$. Consequently, the distance from μ to $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{B})$ is greater than ε. Since μ was arbitrary, the complement of $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{B})$ is open; hence, $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{B})$ is closed. □

Proposition B6.2 If the Banach spaces A, B, and C are such that $\mathsf{A} \otimes \mathsf{B}$ is norm-dense in C, then $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{A}) \otimes \mathcal{A}(\mathcal{Y};\mathcal{F};\mathsf{B})$ is norm-dense in $\mathcal{A}(\mathcal{X} \times \mathcal{Y};\mathcal{G};\mathsf{C})$, where $\mathcal{G}$ is the σ-algebra generated by the rectangular subsets $\mathcal{E} \times \mathcal{F}$.

Proof The atoms of $\mathcal{G}$ are in the rectangular subsets $\mathcal{E} \times \mathcal{F}$. Therefore, any vector measure in $\mathcal{A}(\mathcal{X} \times \mathcal{Y};\mathcal{G};\mathsf{C})$ can be arbitrarily well-approximated in norm by vector measures in $\mathcal{A}(\mathcal{X} \times \mathcal{Y};\mathcal{G};\mathsf{C})$ of the form $\sum_j \mathsf{c}_j \delta_{E_j} \times \delta_{F_j}$ for finite collections of atoms $\{E_j\} \subset \mathcal{E}$ and $\{F_j\} \subset \mathcal{F}$ and a finite collection $\{\mathsf{c}_j\} \subset \mathsf{C}$. By assumption, each $\mathsf{c} \in \mathsf{C}$ can be arbitrarily well-approximated in norm by elements in $\mathsf{A} \otimes \mathsf{B}$. Hence, $\mathcal{A}(\mathcal{X};\mathcal{E};\mathsf{A}) \otimes \mathcal{A}(\mathcal{Y};\mathcal{F};\mathsf{B})$ is norm-dense in $\mathcal{A}(\mathcal{X} \times \mathcal{Y};\mathcal{G};\mathsf{C})$.
□

Appendix C

Propositions for option II

Notation All topological spaces $\mathcal{X}, \mathcal{Y}, \ldots$ are compact and Hausdorff. $\mathcal{C}(\mathcal{X})$, $\mathcal{C}(\mathcal{Y}), \ldots$ are the spaces of real-valued, continuous functions, which are Banach space employing the maximum norm. All partitions of unity will be assumed composed of continuous functions. Hilbert spaces, denoted H, $\mathsf{J}, \ldots$ are complete, sesquilinear inner-product spaces, with no restriction as to their dimension or separability. $\mathcal{K}(\mathsf{H})$, $\mathcal{K}(\mathsf{J}), \ldots$ denote the spaces of compact[1], self-adjoint operators on the specified Hilbert space; these are Banach spaces using the operator norm. $\mathcal{C}(\mathcal{X}; \mathcal{K}(\mathsf{H}))$, $\mathcal{C}(\mathcal{Y}; \mathcal{K}(\mathsf{J}))), \ldots$ are the spaces of compact-operator-valued, continuous functions, which are Banach spaces employing the norm given by first applying the operator norm on the operators pointwise, then the maximum norm over the space.

C.1 Real-valued, continuous functions

Comment The following proposition strengthens the well-known result, which is a special case of a result by Grothendieck [13], that $\mathcal{C}(\mathcal{X} \times \mathcal{Y}) = \mathcal{C}(\mathcal{X}) \check{\otimes} \mathcal{C}(\mathcal{Y})$, where $\check{\otimes}$ indicates completion in the injective norm[2].

Proposition C1.1 The finite-nonnegative-tensor-rank[3] continuous functions are dense in the maximum norm topology for positive, continuous functions on $\mathcal{X} \times \mathcal{Y}$.

Proof Let f be any positive, continuous function on $\mathcal{X} \times \mathcal{Y}$ and take any $\varepsilon > 0$. Since f is continuous, for each $(x, y) \in \mathcal{X}$ there are open subsets $U_{(x,y)} \subset \mathcal{X}$ and $V_{(x,y)} \subset \mathcal{Y}$ such that $f(U_{(x,y)} \times V_{(x,y)}) \subset \left(f(x,y) - \frac{\varepsilon}{2}, f(x,y) + \frac{\varepsilon}{2}\right)$. Since $\mathcal{X} \times \mathcal{Y}$ is necessarily compact, there is a finite subcover, $\left\{U_{(x,y)_j} \times V_{(x,y)_j}\right\}_{j=1}^{n}$. For each $x \in \mathcal{X}$, define U_x by the intersection over all

[1] An operator is *compact* if the image of a bounded sequence necessarily contains a convergent subsequence.

[2] The *injective norm* on $\mathsf{A} \otimes \mathsf{B}$ is the norm induced by its being a subspace of Bilinear$(\mathsf{A}^*, \mathsf{B}^*)$, $\|\mathsf{c}\|_\vee = \sup \mathsf{c}(\varphi, \psi)$, where the supremum is taken over φ in the unit ball of A^* and ψ in the unit ball of B^*.

[3] Using only positive real scalars.

$U_{(x,y)_j}$ containing x. Similarly, for each $y \in \mathcal{Y}$, define V_y by the intersection over all $V_{(x,y)_j}$ containing y. Since $\mathcal{X}$ and $\mathcal{Y}$ are compact, there are finite subcovers, $\left\{U_{x_j}\right\}_{j=1}^{l}$ and $\left\{V_{y_j}\right\}_{j=1}^{m}$, where each $U_{x_i} \times V_{y_k}$ is a subset of one of the $U_{(x,y)_j} \times V_{(x,y)_j}$. Since compact, Hausdorff sets are normal, there are finite partitions of unity, $\{\phi_j\}_{j=1}^{l}$ and $\{\psi_j\}_{j=1}^{m}$, dominated by $\left\{U_{x_j}\right\}_{j=1}^{l}$ and $\left\{V_{y_j}\right\}_{j=1}^{m}$ respectively [45] [73]. Then

$$\sum_{i=1}^{l}\sum_{k=1}^{m} f(x_i, y_l)\phi_i\psi_k \tag{C.1}$$

is a finite-nonnegative-tensor-rank continuous function that is everywhere on $\mathcal{X} \times \mathcal{Y}$ within ε of f. □

C.2 Maps on real-valued, continuous functions

Corollary C2.1 Positive maps $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y}), \mathcal{C}(\mathcal{X}))^+$ are completely-positive[4].

Proof Use **C1.1** and the positivity of both maps L and $I_{\mathcal{B}(\mathcal{C}(\mathcal{Z}))}$. □

Corollary C2.2 Maps $L \in \mathcal{B}(\mathcal{C}(\mathcal{Z}), \mathcal{C}(\mathcal{X}))$ and $K \in \mathcal{B}(\mathcal{C}(\mathcal{W}), \mathcal{C}(\mathcal{Y}))$ satisfy $(K \otimes L)^*(\mu \times \nu) = K^*\mu \times L^*\nu$ for any Radon measures μ on $\mathcal{X}$ and ν on $\mathcal{Y}$.

Proof By the preceding proposition, it is sufficient to demonstrate that for any finite collections $\{f_j\} \subset \mathcal{C}(\mathcal{Z})$ and $\{g_j\} \subset \mathcal{C}(\mathcal{W})$,

$$\sum_j \int_{\mathcal{Z}\times\mathcal{W}} f_j \otimes g_j \, d(K \otimes L)^*(\mu \times \nu) = \sum_j \int_{\mathcal{Z}\times\mathcal{W}} f_j \otimes g_j \, d(K^*\mu \times L^*\nu) \tag{C.2}$$

but this follows from Tonelli's theorem [74], which is applicable since μ, ν are necessarily finite measures,

$$\begin{aligned}
&\sum_j \int_{\mathcal{Z}\times\mathcal{W}} f_j \otimes g_j \, d(K \otimes L)^*(\mu \times \nu) = \sum_j \int_{\mathcal{X}\times\mathcal{Y}} K f_j \otimes L g_j \, d(\mu \times \nu) = \sum_j \int_{\mathcal{X}} K f_j \, d\mu \int_{\mathcal{Y}} L g_j \, d\nu \\
&= \sum_j \int_{\mathcal{Z}} f_j \, dK^*\mu \int_{\mathcal{W}} g_j \, dL^*\nu = \sum_j \int_{\mathcal{Z}\times\mathcal{W}} f_j \otimes g_j \, d(K^*\mu \times L^*\nu) \qquad \square
\end{aligned} \tag{C.3}$$

Comment For any positive map $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y}), \mathcal{C}(\mathcal{X}))^+$, clearly $\|L\|_{\text{op}} = \|L(1_{\mathcal{Y}})\|_{\max}$. Hence, the cone of positive maps is clearly normal.

[4] A map L is *completely-positive* if $L \otimes I_{\mathcal{B}(\mathcal{C}(\mathcal{Z}))}$ is positive for every $\mathcal{Z}$.

Proposition C2.3 For any $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y}),\mathcal{C}(\mathcal{X}))$, $\|L\|_{\text{op}} = \sup_{x\in\mathcal{X}} \|L^*(\delta_x)\|$.

Proof By definition of the operator norm,

$$\begin{aligned}
\|L\|_{\text{op}} &= \sup_{f\in\mathcal{C}(\mathcal{Y}),\|f\|_{\max}\le 1} \|Lf\|_{\max} = \sup\left\{ |(Lf)(x)| \;\middle|\; \begin{array}{l} f\in\mathcal{C}(\mathcal{Y}) \\ \|f\|_{\max}\le 1 \\ x\in\mathcal{X} \end{array} \right\} \\
&= \sup\left\{ \left|\int_{\mathcal{X}} Lf\, d\delta_x\right| \;\middle|\; \begin{array}{l} f\in\mathcal{C}(\mathcal{Y}) \\ \|f\|_{\max}\le 1 \\ x\in\mathcal{X} \end{array} \right\} = \sup\left\{ \left|\int_{\mathcal{X}} f\, dL^*(\delta_x)\right| \;\middle|\; \begin{array}{l} f\in\mathcal{C}(\mathcal{Y}) \\ \|f\|_{\max}\le 1 \\ x\in\mathcal{X} \end{array} \right\} \\
&= \sup_{x\in\mathcal{X}} \|L^*(\delta_x)\|_{\text{total variation}} \qquad \square
\end{aligned} \tag{C.4}$$

Corollary C2.4 Maps $L \in \mathcal{B}(\mathcal{C}(\mathcal{Z}),\mathcal{C}(\mathcal{X}))$ and $K \in \mathcal{B}(\mathcal{C}(\mathcal{W}),\mathcal{C}(\mathcal{Y}))$ satisfy $\|K\otimes L\|_{\text{op}} = \|K\|_{\text{op}}\|L\|_{\text{op}}$.

Proof Using the preceding proposition,

$$\|K\otimes L\|_{\text{op}} \ge \sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \|(K\otimes L)^*(\delta_{(x,y)})\|_{\text{total variation}} \tag{C.5}$$

Since $\delta_{(x,y)} = \delta_x \times \delta_y$, from **C2.3**, this is equal to

$$\sup_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \|K^*(\delta_x)\times L^*(\delta_y)\|_{\text{total variation}} \tag{C.6}$$

By **B1.5**, this is equal to

$$\sup_{x\in\mathcal{X}} \|K^*(\delta_x)\|_{\substack{\text{total}\\ \text{variation}}} \quad \sup_{y\in\mathcal{Y}} \|L^*(\delta_y)\|_{\substack{\text{total}\\ \text{variation}}} \tag{C.7}$$

which, using the preceding proposition again, is equal to $\|K\|_{\text{op}}\|L\|_{\text{op}}$. □

Proposition C2.5 $\mathcal{B}(\mathcal{C}(\mathcal{Y}),\mathcal{C}(\mathcal{X}))$ is not in general a vector lattice.

Counterexample Take $\mathcal{X} = \mathcal{Y} = [-1,1]$ with the usual topology. Take $L \in \mathcal{B}(\mathcal{C}([-1,1]))$ to be

$$Lf(x) = \begin{cases} 0 & \text{if } x\le 0 \\ f(x) - f(-x) & \text{if } x > 0 \end{cases} \tag{C.8}$$

Then $L \vee 0$ should be

$$(L\vee 0)f(x) = \begin{cases} 0 & \text{if } x\le 0 \\ f(x) & \text{if } x > 0 \end{cases} \tag{C.9}$$

but this sends some continuous functions to discontinuous ones. □

Proposition C2.6 $\mathcal{B}(\mathcal{C}(\mathcal{Y}),\mathcal{C}(\mathcal{X}))$ is <u>not</u> in general directed-complete.

Counterexample Take $\mathcal{X} = \mathcal{Y} = [-1,1]$ with the usual topology. Take $L_j \in \mathcal{B}(\mathcal{C}([0,1]))$ to be

$$L_j f(x) = \begin{cases} 0 & \text{if } x \le 0 \\ \sqrt[j]{x}\, f(x) & \text{if } x > 0 \end{cases}$$

Then $\sup_j L_j$ should be

$$(L \vee 0) f(x) = \begin{cases} 0 & \text{if } x \le 0 \\ f(x) & \text{if } x > 0 \end{cases}$$

but this sends some continuous functions to discontinuous ones. □

Proposition C2.7 For any positive $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y}),\mathcal{C}(\mathcal{X}))^+$, there is a transition function $\tau(\cdot|\cdot) : \mathrm{Open}_{\mathcal{Y}} \times \mathcal{X}$, which is: *(i)* an additive set function on the open subsets of $\mathcal{Y}$; *(ii)* pointwise countably additive on the open subsets of $\mathcal{Y}$; and *(iii)* lower-semi-continuous for any fixed open subset $O \subset \mathcal{Y}$–such that

$$(L^*\mu)(O) = \int_{x\in\mathcal{X}} \tau(O|x)\, d\mu(x)$$

for any open $O \subset \mathcal{Y}$ and Radon measure μ on $\mathcal{X}$.

Proof Let $O \subset \mathcal{Y}$ be open. Define $\tau(O|\cdot)$ by

$$\tau(O|x) = \left(\bigvee_{\substack{f \in \mathcal{C}(\mathcal{Y}) \\ 0 \le f \le 1_O}} Lf \right)(x) \tag{C.10}$$

Then $\tau(O|\cdot)$ is lower-semi-continuous, since, for any $a \in \mathbb{R}$,

$$\tau(O|\cdot)^{-1}((a,\infty)) = \bigcup_{\substack{f \in \mathcal{C}(\mathcal{Y}) \\ 0 \le f \le 1_O}} (Lf)^{-1}((a,\infty)) \tag{C.11}$$

which is the union of open sets; hence, it is open. Also, for any Radon measure μ, by inner regularity,

$$(L^*\mu)(O) = \sup_{\text{compact } K\subset O} (L^*\mu)(K) \tag{C.12}$$

By Urysohn's lemma [75], this is equal to

$$\sup_{\substack{f \in \mathcal{C}(\mathcal{Y}) \\ 0 \le f \le 1_O}} \int_{y\in\mathcal{Y}} f(y)\, d(L^*\mu)(y) = \sup_{\substack{f \in \mathcal{C}(\mathcal{Y}) \\ 0 \le f \le 1_O}} \int_{x\in\mathcal{X}} (Lf)(x)\, d\mu(x) \tag{C.13}$$

which is less than or equal to $\int_{x\in\mathcal{X}} \tau(O|x)\, d\mu(x)$, which is less than or equal to

$$\left(\bigvee_{\substack{f \in \mathcal{C}(\mathcal{Y}) \\ 0 \le f \le 1_O}} L^{**} f \right)(\mu) \tag{C.14}$$

which exists since dual spaces are necessarily directed-complete. This is in turn less than or equal to $(L^*\mu)(O)$; hence, $(L^*\mu)(O) = \int_{x\in\mathcal{X}} \tau(O|x)\, d\mu(x)$. □

The bounded, Borel functions on $\mathcal{Y}$, $\mathcal{B}(\mathcal{Y})$, form a Banach space using the supremum norm. The space $\mathcal{B}(\mathcal{Y})$, acting on Radon measures via integration, can be considered a subspace of the linear functionals $\mathcal{C}(\mathcal{Y})^{**}$. Note the preceding shows $\tau(O|x) = (L^*\delta_x)(B)$ and that $L^{**}1_O = \tau(O|\cdot)$ is a function in $\mathcal{B}(\mathcal{X})$ rather than just a general pseudo-function in $\mathcal{C}(\mathcal{X})^{**}$ [83].

Proposition C2.8 For any positive $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y}), \mathcal{C}(\mathcal{X}))^+$, if $\mathcal{Y}$ is metric there is a transition function $\tau(\cdot|\cdot) : \mathrm{Borel}_{\mathcal{Y}} \times \mathcal{X}$, which is: *(i)* an additive set function on the Borel subsets of $\mathcal{Y}$; *(ii)* pointwise countably additive on the Borel subsets of $\mathcal{Y}$; and *(iii)* Borel measurable for any fixed Borel subset $B \subset \mathcal{Y}$–such that

$$(L^*\mu)(B) = \int_{x\in\mathcal{X}} \tau(B|x)\, d\mu(x)$$

for any Borel $B \subset \mathcal{Y}$ and Radon measure μ on $\mathcal{X}$.

Proof Following Hausdorff [21], let G_0 be the collection of the open subsets of $\mathcal{Y}$. For each ordinal α, let G_α be the collection of subsets of $\mathcal{Y}$ that are the countable intersection of subsets from the various collections G_β for ordinals $\beta < \alpha$ if α is odd and that are the countable union of subsets from the various collections G_β for ordinals $\beta < \alpha$ if α is even (where all limit ordinals–those without a predecessor–taken even). For clarification, using the standard notation [76], $G_1 = G_\delta$, $G_2 = G_{\delta\sigma}$, and so on for the finite ordinals.

Fix any ordinal α and suppose that for all subsets B in all collections G_β for $\beta < \alpha$ we have the following property: there is a Borel measurable $\tau(B|\cdot)$ such that $(L^*\mu)(B) = \int_{x\in\mathcal{X}} \tau(B|x)\, d\mu(x)$ for any Radon measure μ on $\mathcal{X}$. Take any $C \in G_\alpha$. If α is even, we have a sequence $\langle B_j \rangle$ of subsets from the various G_β with $\beta < \alpha$ such that $C = \bigcup_j B_j$. Let $\tau(C|\cdot) = \bigvee_j \tau(B_j|\cdot)$. Then $\tau(C|\cdot)$ is measurable (see [69]) and, by the dominated convergence theorem [68], $(L^*\mu)(C) = \int_{x\in\mathcal{X}} \tau(C|x)\, d\mu(x)$ for any Radon measure μ on $\mathcal{X}$; hence, since C was arbitrary, G_α has the property. Similarly, if α is odd, we have a sequence $\langle B_j \rangle$ of subsets from the various G_β with $\beta < \alpha$ such that $C = \bigcap_j B_j$. Let $\tau(C|\cdot) = \bigwedge_j \tau(B_j|\cdot)$. Then $\tau(C|\cdot)$ is measurable and, by the dominated convergence theorem, $(L^*\mu)(C) = \int_{x\in\mathcal{X}} \tau(C|x)\, d\mu(x)$ for any Radon measure μ on $\mathcal{X}$; hence, G_α also has the property. However, by the above proposition **C2.7**, G_0 has the property, so by transfinite induction [22] [86] all the G_α have the property.

Since $\mathcal{Y}$ is metric, its closed subsets are also in $G_1 = G_\delta$ [35]; hence, following Kuratowski [34], the Borel subsets of $\mathcal{Y}$ are in the union $\bigcup_\alpha G_\alpha$. Therefore, the Borel subsets also have the property. □

Comment The preceding proof does not depend on the axiom of choice since the union only needs to be taken up to the ordinal number for the minimal uncountable well-ordered set [21] [34], whose existence does not depend on the axiom of choice [44].

Corollary C2.9 For any positive $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y}), \mathcal{C}(\mathcal{X}))^+$, if $\mathcal{Y}$ is metric then L^{**} acting on $\mathcal{B}(\mathcal{Y}) \subset \mathcal{C}(\mathcal{Y})^{**}$ has image in $\mathcal{B}(\mathcal{X}) \subset \mathcal{C}(\mathcal{X})^{**}$.

Proof Use the density of simple functions in $\mathcal{B}(\mathcal{Y})$ together with the above propositions **C2.8** and **A1.3**.

C.3 Compact-operator-valued, continuous functions

Proposition C3.1 $\mathcal{C}(\mathcal{X}; \mathcal{K}(\mathsf{H})) \otimes \mathcal{C}(\mathcal{Y}; \mathcal{K}(\mathsf{J}))$ is norm dense within $\mathcal{C}(\mathcal{X} \times \mathcal{Y}; \mathcal{K}(\mathsf{H} \otimes \mathsf{J}))$.

Proof Finite-rank operators are norm-dense among compact operators, so proceed as in the first part of the proof for **B4.1**. Then following the argument in the proof of **C1.1** involving partitions-of-unity gives the desired result. □

C.4 Operator inequalities

Proposition C4.1 For any operators $\varphi, \xi \in \mathcal{K}(\mathsf{H})$,

$$\| |\varphi| - |\xi| \|_{\text{op}} \le \sqrt{\|\varphi - \xi\|_{\text{op}} \|\varphi + \xi\|_{\text{op}}}$$

where $|\chi| = \sqrt{\chi^2}$.

Proof By the properties of compact operators, $|\varphi| - |\xi|$ is compact. By the spectral theorem for compact operators, there is a unit-length $\psi \in \mathsf{H}$ which is an eigenvector ψ of $|\varphi| - |\xi|$ and with corresponding eigenvalue λ equal to $\| |\varphi| - |\xi| \|_{\text{op}}$ in magnitude. Then, using the triangle inequality,

$$\begin{aligned}\||\varphi|-|\xi|\|_{\mathrm{op}}^2 &= |\lambda\langle(|\varphi|-|\xi|)\psi,\psi\rangle| \le |\lambda\langle(|\varphi|+|\xi|)\psi,\psi\rangle| \\ &= \left|\frac{1}{2}\langle(|\varphi|+|\xi|)(|\varphi|-|\xi|)\psi,\psi\rangle + \frac{1}{2}\langle(|\varphi|-|\xi|)(|\varphi|+|\xi|)\psi,\psi\rangle\right| \\ &= \left|\left\langle\left(\varphi^2-\xi^2\right)\psi,\psi\right\rangle\right| \le \|\varphi^2-\xi^2\|_{\mathrm{op}} = \left\|\frac{1}{2}(\varphi+\xi)(\varphi-\xi)+\frac{1}{2}(\varphi-\xi)(\varphi+\xi)\right\|_{\mathrm{op}} \\ &\le \frac{1}{2}\|(\varphi+\xi)(\varphi-\xi)\|_{\mathrm{op}} + \frac{1}{2}\|(\varphi-\xi)(\varphi+\xi)\|_{\mathrm{op}} \le \|\varphi-\xi\|_{\mathrm{op}}\|\varphi+\xi\|_{\mathrm{op}} \qquad \square\end{aligned} \tag{C.15}$$

The second needed inequality, which involves subsets of direct products of spaces of operators, is lengthy to state, although it has a very short proof. For real numbers $a = (a_1,\ldots,a_n)$ in the simplex $\{x \in [0,1]^{\times n} | x_1+\cdots+x_n = 1\}$, let $A_n(a,\mathsf{H}) \subset \mathcal{K}(\mathsf{H})^{\times n}$ be the set

$$\left\{(\phi_1,\ldots,\phi_n) \in \mathcal{K}(\mathsf{H})^{\times n} \,\middle|\, \left\|\sum_{j=1}^{n} a_j\phi_j\right\|_{\mathrm{op}} \le 1\right\} \tag{C.16}$$

Let $\mathbb{B}(\mathcal{B}(\mathsf{H}))$ denote the closed, unit ball (using the operator norm) in $\mathcal{B}(\mathsf{H})$. By the triangle inequality, $\mathbb{B}(\mathcal{B}(\mathsf{H}))^{\times n} \subset A_n(a,\mathsf{H})$ for any allowed choice of a. Now consider the set $C_n(\varepsilon,\mathsf{H}) \subset \mathcal{K}(\mathsf{H})^{\times n}$ given by

$$\left\{(\phi_1,\ldots,\phi_n) \in \{B \in \mathcal{K}(\mathsf{H})^{\times n} \,\middle|\, \max_{j<k}\|\phi_j-\phi_k\|_{\mathrm{op}} \le \varepsilon\right\} \tag{C.17}$$

for $\varepsilon > 0$.

Define a distance from $A \in \mathcal{K}(\mathsf{H})^{\times n}$ to a subset $F \subset \mathcal{K}(\mathsf{H})^{\times n}$ by

$$\mathrm{dist}\,(A,F) = \inf_{E\in F}\max_j \|A_j - E_j\|_{\mathrm{op}} \tag{C.18}$$

Using this to define a Hausdorff distance between subsets then gives:

Proposition C4.2 The Hausdorff distance between $A_n(a,\mathsf{H}) \cap C_n(\varepsilon,\mathsf{H})$ and $\mathbb{B}(\mathcal{B}(\mathsf{H}))^{\times n} \cap C_n(\varepsilon,\mathsf{H})$ is bounded by ε.

Proof Take any allowed a and any $\phi \in A_n(a,\mathsf{H}) \cap C_n(\varepsilon,\mathsf{H})$. Then, by the triangle inequality, for any $j \in \{1,2,\ldots,n\}$,

$$\left\|\phi_j - \sum_{k=1}^{n} a_k\phi_k\right\|_{\mathrm{op}} \le \sum_{k=1}^{n} a_k\|\phi_j-\phi_k\|_{\mathrm{op}} \le \varepsilon \tag{C.19}$$

However, by the definition of $A_n(a,\mathsf{H})$, $\sum_{k=1}^{n} a_k\phi_k$ is in $\mathbb{B}(\mathcal{B}(\mathsf{H}))$. $\square$

Note the compactness of the operators was not used in the preceding proof so the proposition holds for the corresponding subsets of self-adjoint operators in $\mathcal{B}(\mathsf{H})$, although we will not require that generalization.

For the following proposition, let $\{A_j\}$ be any finite collection of positive, Hermitian $n \times n$-matrices and let P be any orthogonal projector.

Proposition C4.3 There is a constant c_0, independent of P, n, and r, such that

$$\sum_{j=1}^{r} \|A_j - PA_jP\|_{\text{trace}} \le c_0 \sqrt{\left(\sum_{j=1}^{r} \text{tr } (A_j - PA_jP)\right)\left(\sum_{j=1}^{r} \text{tr } A_j\right)}$$

Proof Observe that for any $a_1, \ldots, a_m \in \mathbb{R} \setminus \{0\}$, the quantity $\sum_{j=1}^{m} \frac{a_j^2}{x_j}$ is minimized over $x_1, \ldots, x_m \in \mathbb{R}^+$ for fixed $\sum_{j=1}^{m} x_j$ when for each j, $x_j = |a_j|$. Now consider the problem of minimizing the product tr D tr $CD^{-1}C^*$ over strictly-positive, Hermitian, $m \times m$-matrices D, given any $n \times m$-matrix C. First take the case of $n \ge m$ and C^*C strictly positive. The minimization over D is the same as minimizing over its eigenvalues, $x_1, \ldots, x_m \in \mathbb{R}^+$, and eigenvectors, W, with W unitary. For fixed value of W and the trace of D, $x_1 + \cdots + x_m$, using the preceding observation, this will occur for

$$x_j = \frac{(x_1 + \cdots + x_m)e_j}{e_1 + \cdots + e_m} \tag{C.20}$$

where e_j is the square root of the jth diagonal entry of W^*C^*CW. Then, tr D tr $CD^{-1}C^* = (e_1 + \cdots + e_m)^2$. Since squaring is monotonic for positive reals and the square root function is concave, this is minimized over all W when W^*C^*CW is diagonal, so $e_1 + \cdots + e_m = \text{tr } \sqrt{C^*C}$. Therefore, tr D tr $CD^{-1}C^* \ge \left(\text{tr } \sqrt{C^*C}\right)^2$. By the ability to embed matrices into larger matrices and the density of invertible matrices, this inequality actually holds for all $n \times m$-matrices C.

By the triangle inequality,

$$\left\|\begin{bmatrix} 0 & C \\ C^* & D \end{bmatrix}\right\|_{\text{trace}}^2 \le \left(\left\|\begin{bmatrix} 0 & C \\ C^* & 0 \end{bmatrix}\right\|_{\text{trace}} + \|D\|_{\text{trace}}\right)^2 = \left(2\text{tr } \sqrt{C^*C} + \text{tr } D\right)^2 \tag{C.21}$$

By the arithmetic-geometric mean inequality, this is less than or equal to

$$8\left(\text{tr } \sqrt{C^*C}\right)^2 + 2\left(\text{tr } D\right)^2 \le 8\left(\left(\text{tr } \sqrt{C^*C}\right)^2 + \left(\text{tr } D\right)^2\right) \tag{C.22}$$

which, by the above result, is less than or equal to

$$8\left(\left(\text{tr } CD^{-1}C^* + \text{tr } D\right)\text{tr } D\right) \tag{C.23}$$

For any Hermitian matrix of the form $\begin{bmatrix} B & C \\ C^* & D \end{bmatrix}$ to be positive, it is necessary (and sufficient) that $B - CD^{-1}C^*$ is positive; otherwise, if there were some vector $\mathbf{v}$ such that

$\langle (B - CD^{-1}C^*)\mathbf{v}, \mathbf{v}\rangle < 0$, then the vector $\mathbf{w} = \begin{bmatrix} \mathbf{v} \\ -D^{-1}C^*\mathbf{v} \end{bmatrix}$ would give

$$\left\langle \begin{bmatrix} B & C \\ C^* & D \end{bmatrix} \mathbf{w}, \mathbf{w} \right\rangle = \langle (B - CD^{-1}C^*)\mathbf{v}, \mathbf{v}\rangle + \left\langle D \begin{bmatrix} D^{-1}C^* & I_m \end{bmatrix} \mathbf{w}, \begin{bmatrix} D^{-1}C^* & I_m \end{bmatrix} \mathbf{w} \right\rangle \quad \text{(C.24)}$$
$$= \langle (B - CD^{-1}C^*)\mathbf{v}, \mathbf{v}\rangle < 0$$

which is a contradiction. Therefore, for any positive, Hermitian matrix of the form $\begin{bmatrix} B & C \\ C^* & D \end{bmatrix}$ with D strictly positive,

$$\left\| \begin{bmatrix} 0 & C \\ C^* & D \end{bmatrix} \right\|_{\text{trace}}^2 \leq 8\left((\text{tr } B + \text{tr } D)\, \text{tr } D \right) \quad \text{(C.25)}$$

By the density of strictly-positive matrices among positive ones, this holds even if D is not strictly positive. Since any orthogonal projector can be brought into the form where there are ones along the upper-left diagonal and zeros everywhere else, this proves the case of $r = 1$,

$$\|A - PAP\|_{\text{trace}} \leq c_0 \sqrt{\text{tr } (A - PAP) \text{tr } A} \quad \text{(C.26)}$$

with $c_0 = 2\sqrt{2}$.

The case of $r > 1$ then follows. Take any row vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \subset \mathbb{R}^+$ that are independent and whose projections to any of the subspace of dimension greater than two using diagonal projectors are still independent. Then, using Lagrange multipliers, the maximum value of $\sum_{j=1}^r c_j x_j$ over $x_1, \ldots, x_r \geq 0$ for fixed value of the product $\left(\sum_{j=1}^r a_j x_j\right)\left(\sum_{k=1}^r b_k x_k\right)$ can only occur if all but two of the x_j's are zero. Even then, the maximum value can only be attained for both x_j's nonzero if it is also attained for one of them zero. Since the given conditions on $\mathbf{a}, \mathbf{b}, \mathbf{c}$ describe a dense subset, then it always the case that the maximizing value is attained when all the x_j's are zero except for one. Then, since

$$\sup \frac{\sum_{j=1}^r \|A_j - PA_jP\|_{\text{trace}}}{\sqrt{\left(\sum_{k=1}^r (\text{tr } (A_k - PA_kP))\right)\left(\sum_{l=1}^r \text{tr } A_l\right)}} \quad \text{(C.27)}$$

over all P, n, r, and $\{A_j\}$, is the same as

$$\sup \frac{\sum_{j=1}^r x_j \|A_j - PA_jP\|_{\text{trace}}}{\sqrt{\left(\sum_{k=1}^r x_k (\text{tr } (A_k - PA_kP))\right)\left(\sum_{l=1}^r x_l \text{tr } A_l\right)}} \quad \text{(C.28)}$$

over all P, n, r, $\{x_j\} \subset \mathbb{R}^+ \cup \{0\}$, and $\{A_j\}$, it is the same as

$$\sup \frac{\|A - PAP\|_{\text{trace}}}{\sqrt{\text{tr } (A - PAP) \text{tr } A}} \quad \text{(C.29)}$$

over all P, n, and A, which is bounded by $2\sqrt{2}$ by the above. □

Remark Apparently $c_0 = 2$ is sufficient, since, by numerical calculation, (C.23) can be replaced by

$$\left\| \begin{bmatrix} 0 & C \\ C^* & D \end{bmatrix} \right\|_{\text{trace}}^2 \le 4\Big(\big(\text{tr } CD^{-1}C^* + \text{tr } D\big) \text{tr } D\Big) \tag{C.30}$$

where the bound is tight since equality is approached as $\varepsilon \to 0^+$ for $D = \varepsilon\sqrt{C^*C}$. However, a proof of this tight bound is lacking.

C.5 Maps on compact-operator-valued, continuous functions

Proposition C5.1 Any positive map $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))^+$ satisfies

$$\|L\|_{\text{op}} = \sup_{\substack{f \in \mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J}))^+ \\ \|f\| \le 1}} \|Lf\| = \sup_{\text{finite-dimensional } \mathsf{K} \subset \mathsf{J}} \|L(I_{\mathsf{K}} 1_{\mathcal{Y}})\| \tag{C.31}$$

Proof From the proof of **C3.1**, functions of the form $\sum_j g_j \phi_j$ for finite collections $\{\phi_j\} \subset \mathcal{K}(\mathsf{J})$ and partition-of-unity $\{g_j\}$ are norm-dense in $\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J}))$. Furthermore, for any $\varepsilon > 0$, it is possible to enforce the constraint that $\|\phi_j - \phi_k\|_{\text{op}} < \varepsilon$ if the support of g_j intersects the support of g_k while maintaining the density property.

Then, for any value of $\varepsilon > 0$, $\|L\|_{\text{op}}$ is equal to

$$\sup\left\{ \left\| \sum_{j=1}^n L(\phi_j g_j) \right\| \;\middle|\; \begin{array}{l} \text{partition-of-unity } \{g_1, \dots, g_n\}, \\ \{\phi_j\} \subset \mathcal{K}(\mathsf{J}), \|\phi_j - \phi_k\|_{\text{op}} < \varepsilon \\ \text{if support } g_j \cap \text{support } g_k \ne \varnothing, \\ \left\|\sum_{j=1}^n \phi_j g_j\right\| \le 1 \end{array} \right\} \tag{C.32}$$

Now consider $b(\varepsilon)$ given by

$$\sup\left\{ \left\| \sum_{j=1}^n L(\phi_j g_j) \right\| \;\middle|\; \begin{array}{l} \text{partition-of-unity } \{g_1, \dots, g_n\}, \\ \{\phi_j\} \subset \mathcal{K}(\mathsf{J}), \|\phi_j - \phi_k\|_{\text{op}} < \varepsilon \\ \text{if support } g_j \cap \text{support } g_k \ne \varnothing, \\ \|\phi_j\|_{\text{op}} \le 1 \end{array} \right\} \tag{C.33}$$

Note $b(\varepsilon)$ is a decreasing function of ε, so its limit as $\varepsilon \to 0^+$ certainly exists. By **C4.2**,

$$b(\varepsilon) \le \|L\|_{\text{op}} \le b(\varepsilon) + \varepsilon\|L\|_{\text{op}} \tag{C.34}$$

so

$$b(\varepsilon) \le \|L\|_{\text{op}} \le \frac{b(\varepsilon)}{1-\varepsilon} \tag{C.35}$$

Hence, $\|L\|_{\text{op}}$ is equal to $\lim_{\varepsilon\to 0^+} b(\varepsilon)$.

The additional constraint that all the $\{\phi_j\}$ are positive can only reduce the supremum or leave it unchanged. On the other hand, since L is positive

$$\left\|\sum_j L(g_j\phi_j)\right\| = \sup_{x\in\mathcal{X}} \sup_{\mathbf{v}\in\mathsf{H},\|\mathbf{v}\|\le 1} \sum_j \langle L(g_j\phi_j)(x)\mathbf{v},\mathbf{v}\rangle \tag{C.36}$$

is certainly less than

$$\left\|\sum_j L(g_j|\phi_j|)\right\| = \sup_{x\in\mathcal{X}} \sup_{\mathbf{v}\in\mathsf{H},\|\mathbf{v}\|\le 1} \sum_j \langle L(g_j|\phi_j|)(x)\mathbf{v},\mathbf{v}\rangle \tag{C.37}$$

In addition, examining the constraints on $\{\phi_j\}$ in (C.33), $\|\,|\phi_j|\,\|_{\text{op}} = \|\phi_j\|_{\text{op}}$ and, by **C4.1**,

$$\||\phi_j| - |\phi_k|\|_{\text{op}} \le \sqrt{\|\phi_j - \phi_k\|_{\text{op}}\,\|\phi_j + \phi_k\|_{\text{op}}} \le \sqrt{\varepsilon}\sqrt{\|\phi_j + \phi_k\|_{\text{op}}} \tag{C.38}$$

By the triangle inequality and the other constraints, this is less than or equal to $\sqrt{2\varepsilon}$ (so this new condition gives rise to a subset that necessarily includes the previous one), which can simply be replaced by ε since $\varepsilon \to 0$. Hence, the additional constraint that all the $\{\phi_j\}$ are positive can only reduce the supremum or leave it unchanged. Therefore, it must leave it unchanged.

However, now starting with $\sup_{\substack{f\,\in\,\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J}))^+\\ \|f\|\,\le\,1}} \|Lf\|$ and repeating the process would lead to the same result, so it is only necessary to take the supremum over the positive cone. Finite-rank operators are norm-dense among compact operators, so it is only necessary to take the supremum over them. Any collection of finite-rank operators live collectively on some finite-dimensional subspace $\mathsf{K} \subset \mathsf{J}$. Since L is positive, the supremum for that particular K then occurs for the constant function with value I_{K}. □

Proposition C5.2 If $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))$ is completely bounded, then, for any compact space $\mathcal{Z}$ and any Hilbert space K, $\|L\otimes I\|_{\text{op}} \le \|L\|_{\text{matrix}}$ with I the identity map in $\mathcal{B}(\mathcal{C}(\mathcal{Z};\mathcal{K}(\mathsf{K})))$.

Proof By the definition of operator norm and the definition of the tensor product of maps, $\|L\otimes I\|_{\text{op}}$ is equal to

$$\sup\left\{\|(L\otimes I)f\| \;\middle|\; \begin{array}{l}\text{finite-tensor-rank } f\in\mathcal{C}(\mathcal{Y}\times\mathcal{Z};\mathcal{K}(\mathsf{J}\otimes\mathsf{K})\\ \text{with } \|f\|\le 1\end{array}\right\} \tag{C.39}$$

By the preceding proof, this is the same as

$$\lim_{\varepsilon\to 0}\sup\left\{\left\|\sum_{j=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{m} L(\phi_{jkl}g_j)\otimes\tau_{jkl}h_k\right\|\right. \tag{C.40}$$

$$\left|\begin{array}{l} m \in \{1,2,\ldots\}, \\ \text{partition-of-unity } \{g_1,\ldots,g_n\}, \text{partition-of-unity } \{h_1,\ldots,h_p\}, \\ \{\phi_{jkl}\} \subset \mathcal{K}(\mathsf{J}), \{\tau_{jkl}\} \subset \mathcal{K}(\mathsf{K}), \\ \|\sum_{l=1}^{m}(\phi_{jkl}\otimes\tau_{jkl} - \phi_{j'k'l}\otimes\tau_{j'k'l})\|_{\text{op}} < \varepsilon \\ \text{if support } g_j \cap \text{support } g_{j'} \neq \varnothing \text{ and support } h_k \cap \text{support } h_{k'} \neq \varnothing \\ \|\sum_{l=1}^{m}\phi_{jkl}\otimes\tau_{jkl}\|_{\text{op}} \leq 1 \end{array}\right\}$$

Writing

$$\left\|\sum_{j=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{m} L(\phi_{jkl}g_j)\otimes\tau_{jkl}h_k\right\| \tag{C.41}$$

out as

$$\sup_{x\in\mathcal{X}}\sup_{z\in\mathcal{Z}}\left\|\sum_{j=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{m} L(\phi_{jkl}g_j)(x)\otimes\tau_{jkl}h_k(z)\right\|_{\text{op}} \tag{C.42}$$

then, since $\mathcal{Z}$ is compact, the maximum value is achieved for a certain z_* (which depends on all the other quantities the supremum is taken over). Incorporating the value of $h_k(z_*)$ into τ_{jkl}, then the constraint:

$$\left\|\sum_{l=1}^{m}\phi_{jkl}\otimes\tau_{jkl}\right\|_{\text{op}} \leq 1 \tag{C.43}$$

becomes

$$\left\|\sum_{l=1}^{m}\phi_{jkl}\otimes\tau_{jkl}\right\|_{\text{op}} \leq h_k(z_*) \tag{C.44}$$

which implies

$$\sum_{k=1}^{p}\left\|\sum_{l=1}^{m}\phi_{jkl}\otimes\tau_{jkl}\right\|_{\text{op}} \leq 1 \tag{C.45}$$

which by the triangle inequality implies

$$\left\|\sum_{k=1}^{p}\sum_{l=1}^{m}\phi_{jkl}\otimes\tau_{jkl}\right\|_{\text{op}} \leq 1 \tag{C.46}$$

Similarly, the constraint:

$$\left\|\sum_{l=1}^{m}(\phi_{jkl}\otimes\tau_{jkl} - \phi_{j'k'l}\otimes\tau_{j'k'l})\right\|_{\text{op}} < \varepsilon \tag{C.47}$$

$$\text{if support } g_j \cap \text{support } g_{j'} \neq \varnothing \text{ and support } h_k \cap \text{support } h_{k'} \neq \varnothing$$

becomes

$$\left\|\sum_{l=1}^{m}(\phi_{jkl}\otimes\tau_{jkl} - \phi_{j'kl}\otimes\tau_{j'kl})\right\|_{\text{op}} < \varepsilon h_k(z_*) \tag{C.48}$$

$$\text{if support } g_j \cap \text{support } g_{j'} \neq \varnothing$$

which implies

$$\sum_{k=1}^{p}\left\|\sum_{l=1}^{m}(\phi_{jkl}\otimes\tau_{jkl}-\phi_{j'kl}\otimes\tau_{j'kl})\right\|_{\text{op}}<\varepsilon \qquad \text{if support } g_j\cap\text{support } g_{j'}\neq\varnothing \tag{C.49}$$

which, by the triangle inequality, implies

$$\left\|\sum_{k=1}^{p}\sum_{l=1}^{m}(\phi_{jkl}\otimes\tau_{jkl}-\phi_{j'kl}\otimes\tau_{j'kl})\right\|_{\text{op}}<\varepsilon \qquad \text{if support } g_j\cap\text{support } g_{j'}\neq\varnothing \tag{C.50}$$

Using the new τ_{jkl} together with the new, weaker constraints can only increase the value of the supremum. Furthermore, the sums over k and l can now be combined, yielding a bound to $\|L\otimes I\|_{\text{op}}$ given by

$$\lim_{\varepsilon\to 0}\sup\left\{\left\|\sum_{j=1}^{n}\sum_{k=1}^{m}L(\phi_{jk}g_j)\otimes\tau_{jk}\right\| \;\middle|\; \begin{array}{l} m\in\{1,2,\ldots\},\\ \text{partition-of-unity } \{g_1,\ldots,g_n\},\\ \{\phi_{jk}\}\subset\mathcal{K}(\mathsf{J}),\{\tau_{jk}\}\subset\mathcal{K}(\mathsf{K}),\\ \|\sum_{k=1}^{m}(\phi_{jk}\otimes\tau_{jk}-\phi_{j'k}\otimes\tau_{j'kl})\|_{\text{op}}<\varepsilon\\ \text{if support } g_j\cap\text{support } g_{j'}\neq\varnothing\\ \|\sum_{k=1}^{m}\phi_{jk}\otimes\tau_{jk}\|_{\text{op}}\leq 1 \end{array}\right\} \tag{C.51}$$

Since the finite-rank operators are norm-dense among the compact operators, this bound is unchanged by requiring the τ_{jk} to be finite-rank. However, the finite-rank operators collectively live on a finite-dimensional subspace of K that is isomorphic to $\mathbb{C}^N$ for some integer $N\leq\dim\mathsf{K}$. Incorporating this into the preceding expression for the bound, we have

$$\|L\otimes I\|_{\text{op}}\leq\sup_{N\leq\dim\mathsf{K}}\|L\otimes I_{\mathcal{M}_N}\|_{\text{op}} \qquad \square \tag{C.52}$$

Proposition C5.3 If the positive map $L\in\mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))^+$, compact space $\mathcal{Z}$, and Hilbert space K are such that $L\otimes I$ is positive, for I the identity map in $\mathcal{B}(\mathcal{C}(\mathcal{Z};\mathcal{K}(\mathsf{K})))$, then $\|L\otimes I\|_{\text{op}}=\|L\|_{\text{op}}$.

Proof If $L\otimes I$ is positive, then clearly $L\otimes I_{\mathcal{M}_n}$ is positive for every positive integer n less than or equal to $\dim\mathsf{K}$. By the preceding proposition, it is therefore only necessary to show that if $L\otimes I_{\mathcal{M}_n}$ is positive, then $\|L\otimes I_{\mathcal{M}_n}\|_{\text{op}}=\|L\|_{\text{op}}$. However, if $L\otimes I_{\mathcal{M}_n}$ is positive, then by **C5.1**,

$$\begin{aligned}\|L\otimes I_{\mathcal{M}_n}\|_{\text{op}} &= \sup_{\text{finite-dimensional } \mathsf{L}\subset\mathsf{J}}\|(L\otimes I_{\mathcal{M}_n})(I_{\mathsf{L}\otimes\mathbb{C}^n}1_{\mathcal{Y}})\| \\ &= \sup_{\text{finite-dimensional } \mathsf{L}\subset\mathsf{J}}\sup_{x\in\mathcal{X}}\|L(I_{\mathsf{L}}1_{\mathcal{Y}})(x)\otimes I_n\|_{\text{op}} \\ &= \sup_{\text{finite-dimensional } \mathsf{L}\subset\mathsf{J}}\sup_{x\in\mathcal{X}}\|L(I_{\mathsf{L}}1_{\mathcal{Y}})(x)\|_{\text{op}}\end{aligned} \tag{C.53}$$

which is $\|L\|_{\text{op}}$ by **C5.1**. $\square$

Corollary C5.4 The completely-positive maps

$$\mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))^{\mathrm{cp}}$$

are completely bounded.

Corollary C5.5 The cone of completely-positive maps

$$\mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))^{\mathrm{cp}}$$

is normal in either the induced operator norm or the matrix norm.

Proposition C5.6 If $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))^{\mathrm{cp}}$, then, for any compact space $\mathcal{Z}$ and any Hilbert space K, $L \otimes I$ is positive, with I the identity map in $\mathcal{B}(\mathcal{C}(\mathcal{Z};\mathcal{K}(\mathsf{K})))$.

Proof Since L is completely positive, by **C5.4** it is completely bounded. Hence, $L \otimes I$ exists by **C5.2**. Furthermore, by **C3.1** and **A1.3**, $L \otimes I$ is unique, so it is meaningful to speak of it being positive.

Now suppose there were some compact space $\mathcal{Z}$ and some Hilbert space K such that $L \otimes I$ were not positive. Then there would be some positive $f \in \mathcal{C}(\mathcal{Y} \otimes \mathcal{Z};\mathcal{K}(\mathsf{J} \otimes \mathsf{K}))^+$ such that $(L \otimes I)f$ is not positive. Since the cone $\mathcal{C}(\mathcal{X} \otimes \mathcal{Z};\mathcal{K}(\mathsf{H} \otimes \mathsf{K}))^+$ is norm-closed and $L \otimes I$ is continuous, that implies there is a relatively open neighborhood of f in the cone $\mathcal{C}(\mathcal{Y} \otimes \mathcal{Z};\mathcal{K}(\mathsf{J} \otimes \mathsf{K}))^+$ whose image under $L \otimes I$ does not intersect $\mathcal{C}(\mathcal{X} \otimes \mathcal{Z};\mathcal{K}(\mathsf{H} \otimes \mathsf{K}))^+$.

Now approximating f as in the proof of **C5.2** as $\sum_{jkl} \varphi_{jkl} \otimes \tau_{jkl}\, g_j \otimes h_k$ (but without the need for any ε-constraints), one finds that for this to occur there must be some $n \in \{1, 2, \dots\}$ for which $L \otimes I_{\mathcal{M}_n}$ is not positive; however, that is a contradiction. □

Proposition C5.7 If either $\dim \mathsf{H}$ or $\dim \mathsf{J}$ is finite and a positive map $L \in \mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))^+$ is such that $L \otimes I_{\mathcal{M}_m}$ is positive for $m = \min\{\dim \mathsf{H}, \dim \mathsf{J}\}$, then L is completely positive.

Proof Clearly, since $L \otimes I_{\mathcal{M}_m}$ is positive, so is $L \otimes I_{\mathcal{M}_n}$ for all $n < m$. Now take $n > m$. $L \otimes I_{\mathcal{M}_n}$ will be positive if for every $f \in \mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J} \otimes \mathbb{C}^n))^+$, $\mathbf{y} \in \mathsf{H} \otimes \mathbb{C}^n$, and $x \in \mathcal{X}$,

$$\langle (((L \otimes I_{\mathcal{M}_n}) f)(x) \mathbf{y}, \mathbf{y} \rangle \geq 0 \tag{C.54}$$

It is enough to show this for f in a dense subset, so f can be restricted to the form $\sum_j \varphi_j g_j$ for $\{g_j\}$ a partition-of-unity and $\{\varphi_j\}$ a collection of compact operators. By the spectral theorem for compact operators, it is enough to show this for the φ_j's all rank one. Following the argument of **B5.9** then gives the desired result. □

Proposition C5.8 The space of completely bounded maps,

$$\mathcal{CB}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))$$

is a Banach space with respect to the matrix norm.

Proof Same argument as for **B5.10** □

Proposition C5.9 The subset of $\mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))$ for which the tensor product with $I_{\mathcal{M}_n}$ is positive for some fixed $n \in \{1,2,\ldots\}$ is closed in the weak topology.

Proof We will show the complement is open. Take such a map L that is not in the subset. By the argument in the proof for **C5.7** and **B5.9**, that implies there are some $x \in \mathcal{X}$, partition-of-unity $\{g_j\}$, finite collections of vectors $\{\mathbf{w}_{jk}\} \subset \mathsf{J}$ and $\{\mathbf{v}_k\} \subset \mathsf{H}$, and $\varepsilon > 0$ such that

$$\sum_{j=1}^{m}\sum_{k,l=1}^{n}\langle L(\mathbf{w}_{jk}\otimes\mathbf{w}_{jl}^{*}\,g_j)(x)\mathbf{v}_l,\mathbf{v}_k\rangle < -\varepsilon \tag{C.55}$$

Then, by the triangle inequality, all the maps in the weak neighborhood

$$\bigcap_{j=1}^{m}\bigcap_{k,l=1}^{n}\mathcal{N}\left(L;\mathbf{w}_{jk}\mathbf{w}_{jk}^{*}\,g_j;\mathbf{v}_l\otimes\mathbf{v}_l^{*}\,\delta_x;\frac{\varepsilon}{2n^4m}\right) \tag{C.56}$$

$$+\bigcap_{j=1}^{m}\bigcap_{k=1}^{n}\bigcap_{r<l}\mathcal{N}\left(L;(\mathbf{w}_{jr}\otimes\mathbf{w}_{jl}^{*}+\mathbf{w}_{jl}\otimes\mathbf{w}_{jr}^{*})\,g_j;\mathbf{v}_k\otimes\mathbf{v}_k^{*}\,\delta_x;\frac{\varepsilon}{n^4m}\right)$$

$$+\bigcap_{j=1}^{m}\bigcap_{k=1}^{n}\bigcap_{r<l}\mathcal{N}\left(L;\mathbf{w}_{jk}\otimes\mathbf{w}_{jk}^{*}\,g_j;(\mathbf{v}_k\otimes\mathbf{v}_m^{*}+\mathbf{v}_m\otimes\mathbf{v}_k^{*})\,\delta_x;\frac{\varepsilon}{n^4m}\right)$$

$$+\bigcap_{j=1}^{m}\bigcap_{k<l}\bigcap_{q<r}\mathcal{N}\left(L;(\mathbf{w}_{jk}\otimes\mathbf{w}_{jl}^{*}+\mathbf{w}_{jl}\otimes\mathbf{w}_{jk}^{*})\,g_j;(\mathbf{v}_q\otimes\mathbf{v}_r^{*}+\mathbf{v}_r\otimes\mathbf{v}_q^{*})\,\delta_x;\frac{2\varepsilon}{n^4m}\right)$$

will also fail to yield a positive tensor product with $I_{\mathcal{M}_n}$. □

Corollary C5.10 The cone of completely positive maps is weakly closed in $\mathcal{B}(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H})))$.

Comment See the comment following **B5.12** concerning the use of the axiom of choice.

Proposition C5.11 For any positive functional $\Phi \in \mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H}))^{*}$, there is some $\mathcal{D}(\mathsf{H})^{+}$-valued, Radon vector measure μ such that

$$\Phi f = \int_{\mathcal{X}} f\,d\mu$$

Proof Since $\mathcal{K}(\mathsf{H})^* \cong \mathcal{D}(\mathsf{H})$, there is some $\rho \in \mathcal{D}(\mathsf{H})^+$ such that $\Phi(\phi 1_{\mathcal{X}}) = \text{tr}\ \rho\phi$ for any $\phi \in \mathcal{K}(\mathsf{H})$. Since finite rank operators are dense in compact ones, ρ lives on a separable subspace of H; let $\{\mathbf{e}_j\}$ be an orthonormal basis for this subspace and $\langle P_j\rangle$ an increasing sequence of projectors onto the subspaces spanned by the first j basis vectors. Applying the Riesz-Markov theorem entry-wise, there is a sequence of Radon vector measures, $\langle \mu_j\rangle$, with each μ_j taking values in $\mathcal{D}(P_j\mathsf{H})^+$. Take $j > k$; then $\|\mu_j - \mu_k\|$ is given by

$$\sup\left\{\sum_{l=1}^{r}\|\mu_j(E_l) - \mu_k(E_l)\|_{\text{trace}} \;\middle|\; \begin{array}{l} r \in \{1,2,\ldots\}, \text{disjoint, Borel} \\ \text{subsets } \{E_1,\ldots,E_n\} \\ \text{with } \bigcup_{l=1}^{r} E_l = \mathcal{X} \end{array}\right\} \tag{C.57}$$

Applying **C4.3**, $\sum_{l=1}^{r}\|\mu_j(E_l) - \mu_k(E_l)\|_{\text{trace}}$ is less than or equal to

$$c_0\sqrt{\left(\sum_{l=1}^{r}\text{tr}\ \left(\mu_j(E_l) - \mu_k(E_l)\right)\right)\left(\sum_{l=1}^{r}\text{tr}\ \mu_j(E_l)\right)} \tag{C.58}$$

$$= c_0\sqrt{\text{tr}\ \left(\mu_j(\mathcal{X}) - \mu_k(\mathcal{X})\right)\text{tr}\ \mu_j(\mathcal{X})}$$

which is less than or equal to $c_0\sqrt{\text{tr}\ (\rho - \mu_k(\mathcal{X}))\,\text{tr}\ \rho}$. This goes to zero as $k \to \infty$ since the $\mu_k(\mathcal{X})$'s are truncations of ρ, which converge to ρ in trace norm, which can be seen either by using **C4.3** again or by first demonstrating that the truncations converge in norm for rank-one ρ (which is readily shown); by the spectral theorem for compact operators, the truncations converge in trace norm for every ρ. Therefore, $\langle\mu_j\rangle$ is a Cauchy sequence. Since $\mathcal{D}(\mathsf{H})^+$-valued vector measures are complete with the given norm [78], the limit is the desired μ. □

For the following proposition, restrict the Hilbert space H to be separable. Then there is some $\rho \in \mathcal{D}(\mathsf{H})^+$ that is strictly positive. Let $\{\mathbf{e}_j\}$ be an orthonormal basis for H composed of eigenvectors of ρ, with eigenvalues λ_j in decreasing order. For any Radon measure μ on $\mathcal{X}$, by the preceding proposition there is the induced variation measure $\nu = |L^*(\rho 1_{\mathcal{X}}\mu)|$ on $\mathcal{Y}$, which is also Radon [78]. Then we have the following result:

Proposition C5.12 For any positive $L \in \mathcal{B}\left(\mathcal{C}(\mathcal{Y};\mathcal{K}(\mathsf{J})),\mathcal{C}(\mathcal{X};\mathcal{K}(\mathsf{H}))\right)^+$, the adjoint map L^* induces a map K that sends $L^1(\mathcal{X};\mu;\mathcal{D}(\mathsf{H}))^+$ into $L^1(\mathcal{Y};\nu;\mathcal{D}(\mathsf{J}))^+$.

Proof Let $C \subset \mathcal{Y}$ be any closed, ν-null subset. By outer-regularity, there are open sets containing C with arbitrarily small ν-measure. Therefore, by Urysohn's lemma [75],

$$0 = \nu(A) = \inf_{f\in\mathcal{C}(\mathcal{Y}),1_{\mathcal{X}}\geq f\geq 1_C}\int_{\mathcal{Y}} f\,d\nu \tag{C.59}$$

Hence, by the definition of ν, for any $\phi \in \mathcal{K}(\mathsf{J})^+$,

$$0 = \inf_{f\in\mathcal{C}(\mathcal{Y}),1_{\mathcal{Y}}\geq f\geq 1_C}\int_{x\in\mathcal{X}}\text{tr}\ \rho L(\phi f)(x)\,d\mu(x) \tag{C.60}$$

For A any Borel subset of $\mathcal{X}$ and τ any positive, self-adjoint operator on the span of finitely many of the $\{\mathbf{e}_j\}$, any single-term simple function, $\tau 1_A$, can be scaled to be less than $\rho 1_{\mathcal{X}}$. Then

$$0 = \inf_{f \in \mathcal{C}(\mathcal{Y}), 1_{\mathcal{Y}} \geq f \geq 1_C} \int_{x \in \mathcal{X}} 1_A(x) \text{tr } \tau L(\phi f)(x)\, d\mu(x) \tag{C.61}$$

Now take $\{\tau_j\}$ and $\{A_j\}$ to be finite collections of such operators and subsets. For allowed $f_1, \ldots, f_n$, the pointwise product $f_1 \cdots f_n$ is also allowed and is less than or equal to each of the f_j; hence,

$$\begin{aligned} &\inf_{f \in \mathcal{C}(\mathcal{Y}), 1_{\mathcal{Y}} \geq f \geq 1_C} \sum_{j=1}^{n} \int_{x \in \mathcal{X}} 1_{A_j}(x) \text{tr } \tau_j L(\phi f)(x)\, d\mu(x) \\ &= \sum_{j=1}^{n} \inf_{f \in \mathcal{C}(\mathcal{Y}), 1_{\mathcal{X}} \geq f \geq 1_C} \int_{x \in \mathcal{X}} 1_{A_j}(x) \text{tr } \tau_j L(\phi f)(x)\, d\mu(x) = 0 \end{aligned} \tag{C.62}$$

Since simple functions of the form $\sum_{j=1}^{n} \tau_j 1_{A_j}$ are dense in $L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H}))^+$, then for any element $\xi \in L^1(\mathcal{X}; \mu; \mathcal{D}(\mathsf{H}))^+$,

$$0 = \inf_{f \in \mathcal{C}(\mathcal{Y}), 1_{\mathcal{Y}} \geq f \geq 1_C} \int_{x \in \mathcal{X}} \text{tr } \xi(x) L(\phi f)(x)\, d\mu(x) = \inf_{f \in \mathcal{C}(\mathcal{Y}), 1_{\mathcal{Y}} \geq f \geq 1_C} \int_{\mathcal{Y}} f\, d(\phi L^*(\xi\mu)) \tag{C.63}$$

where $\phi L^*(\xi\mu)$ is a Radon measure. Therefore, it must be that $(\phi L^*(\xi\mu))(C) = 0$. Since ϕ was arbitrary, $L^*(\xi\mu)(C)$ must be zero.

By inner regularity, any Borel set can have its measure approximated arbitrarily well by closed sets it contains. Therefore, for any ν-null, Borel subset A, $L^*(\xi\mu)(A)$ is also zero. Therefore, $L^*(\xi\mu)$ is absolutely continuous with respect to ν. Since $\mathcal{D}(\mathsf{H})$ has the Radon-Nikodým property [14], there is a $\psi \in L^1(\mathcal{Y}; \nu; \mathcal{D}(\mathsf{J}))^+$ such that $L^*\mu = \psi\nu$, which gives the desired map K by $K\xi = \psi$. □