Fakultät für Mathematik
Technische Universität München

Technische Universität München
Lehrstuhl für Mathematische Physik

Compression and measurements in quantum information theory

Andreas Bluhm

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Felix Krahmer

Prüfende der Dissertation:
1. Prof. Dr. Michael M. Wolf
2. Prof. Dr. Matthias Christandl,
   Københavns Universitet

Die Dissertation wurde am 08.04.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 16.07.2019 angenommen.

Department of Mathematics
Technical Univeristy of Munich

Technical University of Munich
Chair of Mathematical Physics

Compression and measurements in quantum information theory

Andreas Bluhm

Full imprint of the dissertation approved by the Department of Mathematics of the Technical University of Munich to obtain the academic degree of

Doctor of Natural Sciences (Dr. rer. nat.)

Chairman:
Prof. Dr. Felix Krahmer

Examiners of the dissertation:
1. Prof. Dr. Michael M. Wolf
2. Prof. Dr. Matthias Christandl,
   Københavns Universitet

The dissertation was submitted to the Technical University of Munich on 08.04.2019 and was accepted by the Department of Mathematics on 16.07.2019.

Zusammenfassung

Diese Dissertation beschäftigt sich mit den Grenzen, die die Quantenmechanik der Quanteninformationsverarbeitung setzt. Zunächst betrachten wir die *Komprimierbarkeit* eines Quantensystems in Abhängigkeit von vorher festgelegten Messungen. Danach stellen wir eine Verbindung zwischen der *Kompatibilität von Messungen* und Inklusionsproblemen freier Spektraeder her. Abschließend erforschen wir, in welchem Maße sogenannte Sketching-Techniken benutzt werden können, um die Größe von semidefiniten Programmen zu reduzieren, welche wichtige Werkzeuge für die Quanteninformationstheorie darstellen.

Abstract

This dissertation treats the limitations quantum mechanics imposes on certain quantum information processing tasks. First, we explore the *compression* of a quantum system relative to a fixed set of measurements. Then, we establish a connection between the *compatiblity of measurements* and inclusion problems of free spectrahedra. Finally, we study how sketching techniques can be used to reduce the dimensionality of semidefinite programs, which are useful tools in quantum information theory.

"Was ihr machen wollt, weiß ich nicht", sagte Pippi. "Ich werde jedenfalls nicht auf der faulen Haut liegen. Ich bin nämlich ein Sachensucher, und da hat man niemals eine freie Stunde."

– aus „Pippi Langstrumpf" von Astrid Lindgren

Acknowledgements

There is a long list of people who, each in their own way, have made invaluable contributions to the completion of this thesis. First, I want to thank my adviser Michael M. Wolf for his patience, encouragement and good advice whenever I needed guidance.

Moreover, I want to thank the past and present members of the M5 group: Michael M. Wolf, Robert König and Michael Keyl for answering my many questions and creating a great work environment; Wojciech Dybalski for agreeing to be my mentor; Wilma Ghamam and Silvia Schulz for their kind help in dealing with the practical issues of university life; my office mate Javier Cuesta for his friendliness and aura of productivity; Alexander Kliesch and my fellow PhD students Anna-Lena Hashagen, Daniel Stilck França, Javier Cuesta, Luca Arceci, Margret Heinze, Martin Idel, Martina Gschwendtner, Matthias Caro, Michael Kech, Maximilian Duell, Stefan Huber and Yimin Ge for countless discussions about work and other more or less important aspects of life; the postdocs Anna Vershynina, Cambyse Rouzé, Daniela Cadamuro, Ion Nechita, Lisa Hänggli, Milán Mosonyi and Sabina Alazzawi for their own special contributions to the group life; Lukas Rauber, Nolan Peard, Shaun Datta, Ujan Chakraborty, Vu Phanh Thanh, Yifan Jia and all other students and visitors for sharing their ideas. Thank you all for creating such a pleasant and fruitful atmosphere at work.

During my PhD, I have had the possibility to work with a lot of amazing people from whom I learned important things about research. In particular, I thank Angela Capel, Eric Hanson, Ion Nechita, Lukas Rauber and Ujan Chakraborty for countless hours at the black board. Furthermore, I am obliged to the groups of David Pérez-García, Marcus Huber and Omar Fawzi as well as the Institut Henri Poincaré, the LPT at the Université Paul Sabatier in Toulouse and Maria Jivulescu and Nicolae Lupa at the University of Timişoara for hosting me on diverse occasions. Special thanks go to Matthias Christandl for agreeing to be an examiner for this thesis and to Felix Krahmer for accepting to be the chairman of the examination committee. Moreover, I want to thank Isabella Wiegand from the ISAM office especially for patiently answering all my questions concerning traveling and PhD related issues.

For the improvement of the present text, I am severely indebted to Christina Dirscherl, Lukas Nickel, Margret Heinze, Shaun Datta, Stefan Huber and my family.

I also want to thank all my friends who have supported me in various ways over the years. The list of names would be far to long, so I can only name a few. I am grateful to my flatmates Patrick Wilke and Philipp Geiger for the great time in Isarstan. Furthermore, I want to express my gratitude to my dear friend Christina Dirscherl for being her incomparable self, thus keeping me sane and happy over the last years.

Schlussendlich wäre all dies nicht ohne den Zuspruch und den Rückhalt meiner Eltern und meiner Schwester möglich gewesen. Außerdem danke ich meiner Großmutter, die mich stets fürsorglich unterstützt hat. Diese Arbeit ist meinem Großvater gewidmet, der ihre Vollendung, die er stets interessiert verfolgt hat, leider nicht mehr erleben durfte.

List of contributed articles

This thesis is based on the following articles:

Core articles as principal author

I) Andreas Bluhm, Lukas Rauber, and Michael M. Wolf (2018).
Quantum compression relative to a set of measurements.
Annales Henri Poincaré, 19(6), 1891–1937.
(see also article [1] in the bibliography)

II) Andreas Bluhm and Ion Nechita (2018).
Joint measurability of quantum effects and the matrix diamond.
Journal of Mathematical Physics 59, 112202.
(see also article [2] in the bibliography)

Further articles

III) Andreas Bluhm and Daniel Stilck França (2019).
Dimensionality reduction of SDPs through sketching.
Linear Algebra and its Applications, 563, 461–475.
(see also article [3] in the bibliography)

IV) Andreas Bluhm and Ion Nechita (2018).
Compatibility of quantum measurements and inclusion constants for the matrix jewel.
arXiv preprint arXiv:1809.04514.
Submitted to *SIAM Journal on Applied Algebra and Geometry.*
(see also article [4] in the bibliography)

I, Andreas Bluhm, am the principal author of Core Articles I and II.

Contents

1 Introduction **1**
- 1.1 Outline . . . 1
- 1.2 Summary of results . . . 2

2 Mathematical formulation of quantum mechanics **5**
- 2.1 Quantum states and measurements . . . 5
- 2.2 Composite systems . . . 8
- 2.3 Quantum operations . . . 9

3 Extensions and fixed-points of completely positive maps **11**

4 Semidefinite programming **13**
- 4.1 Basic notions and duality . . . 13
- 4.2 Verifying complete positivity . . . 14
- 4.3 Algorithms to solve semidefinite programs . . . 15

5 Sketching and convex optimization **17**
- 5.1 An introduction to sketching techniques . . . 17
- 5.2 Sketching for problems in convex optimization . . . 18

6 Compression in quantum information theory **21**
- 6.1 Quantum source coding . . . 21
- 6.2 Measurement compression . . . 22
- 6.3 Sparsification of POVMs . . . 23
- 6.4 Model compression . . . 23
- 6.5 Shadow tomography . . . 24
- 6.6 Quantum compression relative to a set of measurements . . . 25

7 Free spectrahedra **27**

8 Compatibility of quantum measurements **31**
- 8.1 Compatible quantum measurements . . . 31
- 8.2 Compatibility regions . . . 32
- 8.3 Approximate quantum cloning . . . 34
- 8.4 The connection to free spectrahedra . . . 35

Bibliography **37**

A Core Articles **43**
- A.1 Quantum compression relative to a set of measurements . . . 43
- A.2 Joint measurability of quantum effects and the matrix diamond . . . 94

B Further articles **125**
B.1 Dimensionality reduction of SDPs through sketching 125
B.2 Compatibility of quantum measurements and inclusion constants for the matrix jewel 144

1 Introduction

In the field of quantum information theory, one studies how the properties of quantum mechanical systems can be used for information theoretic tasks. In this endeavor, it is not only important to find new protocols for such tasks, but also to understand which limitations quantum mechanics imposes on information processing. This is the overarching topic of this thesis.

1.1 Outline

In this section, we explain which kinds of limitations we investigate in this thesis and conclude with an outline of the different chapters.

An important operation in information theory, quantum or classical, is compression. As an example, in order to perform complex cryptographic protocols or quantum computations, it is essential to store quantum information for a certain amount of time such that it can be used at a later step of the protocol or the computation at hand. However, the presence of decoherence is a severe threat to quantum information, which makes the storage of quantum states very challenging. It is thus desirable to store as little quantum information as possible. Our aim is therefore to understand to what extend quantum theory restricts our ability to reduce the dimensionality of the system under study, depending on the information we would like to preserve. An important tool which we use to obtain bounds on the compression dimension in this thesis is fixed-point theorems for completely positive maps, which describe the evolution of quantum mechanical systems.

High dimensionality is not only a challenge in quantum physics, but also in classical optimization. An important class of optimization problems is semidefinite programs, which often arise in engineering and quantum information theory. Although there are well-known efficient algorithms to solve such problems, these algorithms have memory requirements and runtimes which scale badly with the problem size, thus making them unsuitable for large problems. In this thesis, we give conditions under which we can find a smaller semidefinite program which still gives an approximate solution with high probability. This smaller problem can subsequently be treated using standard solvers.

Another important property of quantum mechanics is the incompatibility of quantum measurements. Incompatible observables, such as position and momentum, are required for the violation of Bell inequalities, which constitutes the basis of device-independent quantum cryptography. Incompatibility is thus a resource for quantum information tasks, similarly to entanglement. The minimal amount of white noise we need to add to a collection of measurements in order to make them compatible can be used to quantify the resource incompatibility. Hence, it is important to characterize how robust incompatibility is under the addition of noise, to obtain a detailed picture of how much incompatibility is available in an experimental situation. In this thesis, we develop a connection between the compatibility of quantum measurements and the inclusion of free spectrahedra. The latter are objects which are actively studied in convex optimization as a tractable relaxation for computationally hard problems. The interplay between convex optimization and quantum information allows us not only to obtain a greatly improved picture of the amount of incompatibility available under certain constraints, but also to use tools from quantum information theory to find new inclusion constants for a class of free spectrahedra.

In the rest of this chapter, we briefly discuss the contributed articles in this thesis. In Chapter 2, we give an introduction to the basic concepts of quantum information theory such as quantum states, measurements, and channels. More operator-algebraic aspects of quantum channels are reviewed in Chapter 3. We then move to an introduction to semidefinite programming in Chapter 4 before we discuss in Chapter 5 how sketching techniques can be used to compress such convex optimization problems. In Chapter 6, we present different notions of compression in quantum information theory and discuss their relation to our work. Chapter 7 then provides a brief introduction to the study of free spectrahedra. We conclude in Chapter 8 with an outline of previous results on the compatibility of quantum measurements and the improvements we obtain by using results from the study of free spectrahedra.

After this overview, we include the contributed articles. Every article is preceded by a summary of the contributions of the respective work and a description of the individual contribution of the author of this thesis. Furthermore, we include for each article the permission to use it in this thesis.

1.2 Summary of results

The contributed articles deal with different aspects of classical and quantum compression and the compatibility of quantum measurements. Core Article I investigates the compression of quantum states such that the outcomes of a set of measurements that we fix beforehand are not affected. The incompatibility of quantum measurements is the topic of Core Article II and Article IV. Core Article II focuses on binary measurements, whereas the case of measurements with an arbitrary number of outcomes is treated in Article IV. In Article III, we investigate how sketching techniques can be used to reduce the dimensionality of semidefinite programs. Note that the author of this thesis does not claim to be the principal author of the Articles III and IV.

Core articles as principal author

- *Article I [1]: Quantum compression relative to a set of measurements*
  In this work, we investigate the possibility of compressing a quantum system to one of smaller dimension such that we preserve the measurement statistics of a set of observables that we fix in advance. Here, we allow for an arbitrary amount of classical side information because classical storage is readily available whereas quantum memories are hard to construct in practice. We prove both upper and lower bounds on the minimal dimension of the compressed quantum systems and give an algorithm for the computation of this dimension, based on a semidefinite program. We find that the presence of symmetries in the measurements allows for compression, but that arbitrary measurements are incompressible. The bounds are proven following two independent approaches: The first one uses methods from operator algebras and is based on a result by Arveson on fixed-points of completely positive maps. The second one uses Bézout's theorem from classical algebraic geometry. Both approaches are complementary in the sense that the operator-algebraic approach gives sharper bounds whereas the algebro-geometric techniques allow us to show that our results still apply if the compression map acts on multiple copies of the quantum state. While we focus on the case in which the measurements have to be preserved exactly, we prove that the minimal dimension of the compressed system cannot

be decreased by allowing for an arbitrarily small error. The underlying reason for this is that we find an upper bound on the amount of classical side information that can be used for compression.

- *Article II [2]: Joint measurability of quantum effects and the matrix diamond*
In the second core article, we investigate the compatibility of binary quantum measurements and connect this problem to the inclusion of free spectrahedra. The latter arise as matricial relaxations of linear matrix inequalities and are studied in convex optimization to find approximate solutions to computationally hard problems. An important example of a free spectrahedron for this work is the matrix diamond, which is a matricial relaxation of the ℓ_1-ball. We show that the compatibility of binary quantum measurements is equivalent to the inclusion of the matrix diamond into the free spectrahedron defined by the measurements. Furthermore, we prove that the noise robustness of binary quantum measurements corresponds to finding inclusion constants for the matrix diamond. This allows us to use results from the study of free spectrahedra to bound the maximal amount of incompatibility present if the number of measurements and the dimension of the quantum system are fixed. In particular, we solve the case in which the system size is exponential in the number of measurements. Conversely, we can use the connection we establish here to study the free spectrahedral inclusion for the matrix diamond using techniques from quantum information theory such as asymmetric approximate cloning.

Further articles

- *Article III [3]: Dimensionality reduction of SDPs through sketching*
In this article, we show how to use sketching techniques to obtain semidefinite programs (SDPs) of reduced dimension which yield an approximate solution to the original problem with high probability. The main tool we use to achieve this is Johnson-Lindenstrauss transforms. If the matrices specifying the problem have Schatten 1-norms which are constant in the problem size and the same holds for an optimal solution, our techniques yield significant time and memory savings. Moreover, we show that the above approach cannot work for arbitrary semidefinite programs and that the condition on the Schatten 1-norms is necessary.

- *Article IV [4]: Compatibility of quantum measurements and inclusion constants for the matrix jewel*
This work is a follow-up to Article II. Here, we extend the connection between the compatibility of quantum measurements and the inclusion of free spectrahedra to measurements with an arbitrary number of outcomes. The universal object here is the matrix jewel which we introduce to generalize the matrix diamond. We prove that a set of measurements is compatible if and only if the matrix jewel is included in the free spectrahedron defined by these measurements. With the proper adjustments, the inclusion constants for the matrix jewel again correspond to the noise robustness of quantum measurements. This correspondence enables us to use results on approximate quantum cloning and mutually unbiased bases to bound the set of inclusion constants for the matrix jewel. Moreover, we develop symmetrization techniques to obtain bounds on the incompatibility of quantum measurements. Finally, we introduce the notion of incompatibility witnesses, which yield tractable relaxations to verify the compatibility of concrete measurements.

2 Mathematical formulation of quantum mechanics

In this chapter, we give a brief introduction to the mathematical formulation of quantum mechanics in finite dimensions, focussing on concepts used in quantum information theory. This material can be found in textbooks such as [5, 6, 7, 8] and articles such as [9]. In Section 2.1, we discuss the preparation and measurement of single quantum systems, before moving to composite systems in Section 2.2. Finally, we introduce the mathematical description of valid transformations on quantum systems in Section 2.3.

Before we start, let us introduce some basic notation. For any $n \in \mathbb{N}$, $[n]$ is a shorthand notation for the set $\{1, \ldots, n\}$. The set of real vectors with non-negative entries is written as $\mathbb{R}^n_+$. By the canonical basis of $\mathbb{C}^n$ we mean the set of vectors with one entry equal to 1 and the others equal to 0. Furthermore, we write $\mathcal{M}_{m,n}$ for the set of complex $m \times n$ matrices, where m, $n \in \mathbb{N}$. If $n = m$, we just write $\mathcal{M}_n$ for brevity. For $A \in \mathcal{M}_{m,n}$, the matrix A^* is the Hermitian conjugate whereas A^T is the transpose. For the set of Hermitian matrices, we write $\mathcal{M}_n^{\mathrm{herm}}$. We say that $A \geq B$ for A, $B \in \mathcal{M}_n^{\mathrm{herm}}$ if $A - B$ is positive semidefinite and write $A > 0$ to indicate that A is positive definite. The identity matrix in dimension n is denoted by $\mathbb{1}_n$, where we sometimes omit the subscript. The notation id_m and id_B is used for the identity map on $\mathcal{M}_m$ and system B, respectively. The trace on $A \in \mathcal{M}_n$ is written as $\mathrm{Tr}\,(A)$. If $A \in (\mathcal{M}_n)^g$, $g \in \mathbb{N}$, we often write $A_i \in \mathcal{M}_n$, $i \in [g]$, for the entries of A.

Following the convention in mathematical physics, we define all inner products such that they are linear in the second argument and conjugate linear in the first. The usual inner product on $\mathbb{C}^n$ is written as $\langle \cdot, \cdot \rangle$ and $\|\cdot\|_2$ is the Euclidean norm on $\mathbb{C}^n$. The ℓ_p-norm on $\mathbb{C}^n$ for $p \in \mathbb{N}$ is defined as

$$\|x\|_p = (|x_1|^p + \ldots + |x_n|^p)^{1/p} \qquad \forall x \in \mathbb{C}^n.$$

Moreover, the ℓ_∞-norm is defined as $\|x\|_\infty = \max_{i \in [n]} |x_i|$ for all $x \in \mathbb{C}^n$.

We also use the Schatten p-norms on $\mathcal{M}_n$, where for $p \in \mathbb{N}$

$$\|A\|_p = \mathrm{Tr}\left((A^* A)^{p/2}\right)^{1/p} \qquad \forall A \in \mathcal{M}_{m,n}$$

and $\|\cdot\|_\infty$ is the operator norm. The Schatten 1-norm and Schatten 2-norm are also called *trace norm* and *Hilbert-Schmidt norm*, respectively.

We follow the custom in quantum mechanics and use bra-ket notation on several occasions. Let $d \in \mathbb{N}$ and $\psi \in \mathbb{C}^d$. Then, $|\psi\rangle$ is used to denote the vector ψ, whereas $\langle\psi|$ is the corresponding element in the dual space of $\mathbb{C}^d$. The latter means that $\langle\psi| : \mathbb{C}^d \to \mathbb{C}$ is the linear functional which maps $\varphi \mapsto \langle\psi|\varphi\rangle := \langle\psi, \varphi\rangle$ for all $\varphi \in \mathbb{C}^d$. Let $D \in \mathbb{N}$ and $|\chi\rangle \in \mathbb{C}^D$. The linear operator $|\chi\rangle\langle\psi| : \mathbb{C}^d \to \mathbb{C}^D$ is then defined as sending $|\varphi\rangle \mapsto \langle\psi|\varphi\rangle\,|\chi\rangle$ for all $|\varphi\rangle \in \mathbb{C}^d$.

2.1 Quantum states and measurements

An experiment in physics can abstractly be described in the following way: First, we prepare the system we want to study. Subsequently, we run the experiment on this system and register the measurement outcomes. As quantum mechanics is a statistical theory, the theory does not predict the individual measurement outcomes, but their probabilities. Therefore, what we obtain are the relative frequencies of the outcomes if we repeat this experiment very often. These can be compared to theoretical predictions.

Hence, we can divide such an idealized statistical experiment into a *preparation procedure* and a *measurement*. During the preparation, we prepare the quantum system in a *state*, which specifies the probability distributions of all possible measurements on this system. The state of a quantum system is independent of the particular preparation procedure and there can be many ways to create the same quantum state. A quantum state can thus be seen as the equivalence class of all procedures which result in the same outcomes for all possible measurements. During the measurement step, we measure an observable quantity, an *observable*. The observables are again independent of the particular measurement procedure and can be understood as equivalence classes of procedures which give the same probability distributions for all preparations. The set of observables gives rise to an *observable algebra*. In quantum mechanics, this algebra is the algebra of bounded operators on a separable Hilbert space $\mathcal{H}$, which we denote by $\mathcal{B}(\mathcal{H})$.

A state ρ in quantum mechanics is a linear functional which maps any element A in the observable algebra to its expectation value $\langle A\rangle_\rho$ with respect to ρ, which is a real number. Thus, quantum states are elements in the dual space of $\mathcal{B}(\mathcal{H})$. In order to obtain well-defined probabilities from the measurements, we require that states are *positive*, i.e.

$$\langle A^*A\rangle_\rho \geq 0 \qquad \forall A \in B(\mathcal{H}),$$

and *normalized*, i.e. for the identity operator $\mathbb{1} \in \mathcal{B}(\mathcal{H})$, we impose that

$$\langle \mathbb{1}\rangle_\rho = 1.$$

In this thesis, we focus on finite-dimensional quantum systems because this is the setting in which most research in quantum information theory is done. For such systems, we can choose $\mathcal{H} = \mathbb{C}^d$ for some $d \in \mathbb{N}$ and $\mathcal{B}(\mathcal{H}) \simeq \mathcal{M}_d$. Systems with $d = 2$ are often called *qubits*, which is short for quantum bits. We can equip $\mathcal{M}_d$ with the Hilbert-Schmidt inner product $\langle\cdot,\cdot\rangle_{\mathrm{HS}}$ to turn it into a Hilbert space, where

$$\langle A, B\rangle_{\mathrm{HS}} = \mathrm{Tr}\left(A^*B\right).$$

Using the Riesz representation theorem [10, Theorem I.3.4], we can identify quantum states with the set of d-dimensional density matrices

$$\mathcal{S}(\mathbb{C}^d) := \{\rho \in \mathcal{M}_d : \rho \geq 0, \mathrm{Tr}\left(\rho\right) = 1\},$$

where the constraints come from positivity and normalization. From the above definition, it can be seen that the set of density matrices is convex. The extreme points of this set are the rank 1 projectors which we call *pure* states [7, Proposition 2.11]. They are of the form $|\psi\rangle\langle\psi|$, where $|\psi\rangle \in \mathbb{C}^d$ and $\|\psi\|_2 = 1$. All states which are not pure are called *mixed* states. An important example of such a state is the *maximally mixed state* $\mathbb{1}_d/d \in \mathcal{S}(\mathbb{C}^d)$. Any mixed state can be written as a convex combination of pure states. A possible way to obtain such a representation is the spectral decomposition [11, Theorem 2.5.6], which for a state $\rho \in \mathcal{S}(\mathbb{C}^d)$ has the form

$$\rho = \sum_{i=1}^{d} \lambda_i \left|\psi_i\rangle\langle\psi_i\right|.$$

Here, the $\lambda_i \in [0,1]$ are the eigenvalues and the $|\psi_i\rangle$ are the corresponding normalized eigenvectors for all $i \in [d]$. A special class of quantum states are the *classical states*. These are the (possibly mixed) states which are diagonal in the canonical basis of $\mathbb{C}^d$. They correspond to classical probability distributions on the fixed basis states.

Now that we have defined the set of quantum states, let us come to the description of measurements. Abstractly, the outcome of an experiment can be described as a set of measurement outcomes $\{\alpha_i\}_{i\in\Sigma}$, where Σ labels the outcomes, and a probability distribution p which encodes the probabilities with which these outcomes occur. For simplicity, we assume that Σ is a finite set and identify it with $[n]$ for some $n \in \mathbb{N}$. Hence, $p = (p_1, \ldots, p_n)$, where $p_i \in [0,1]$ for all $i \in [n]$ and the p_i sum to 1. For a treatment of real-valued outcomes, we refer to [7, Section 3.1.4]. In a statistical experiment, measurements are linear maps from quantum states to probability distributions. We assume that the system has been prepared in the state $\rho \in \mathcal{S}(\mathbb{C}^d)$. Using the Riesz representation theorem again, there is therefore an $E_i \in \mathcal{M}_d^{\mathrm{herm}}$ for all $i \in [n]$ such that

$$p_i = \mathrm{Tr}\left(E_i \rho\right).$$

As p_i is a probability for any state ρ and therefore lies in the interval $[0,1]$, we require that E_i is positive semidefinite and that $E_i \leq \mathbb{1}_d$. This justifies the definition of the set of *effect operators* or *effects*

$$\mathcal{E}(\mathbb{C}^d) := \{E \in \mathcal{M}_d : E \geq 0, E \leq \mathbb{1}_d\}.$$

Furthermore, the probabilities have to sum to 1 for any ρ. On the level of effect operators, we thus require that the E_i sum to the identity operator.

Definition 2.1. *Let d, $n \in \mathbb{N}$, and $E_i \in \mathcal{E}(\mathbb{C}^d)$ for all $i \in [n]$. This family of effect operators is called a* positive operator-valued measure *(POVM) if*

$$\sum_{i=1}^{n} E_i = \mathbb{1}_d.$$

If we are not interested in the measurement outcomes α_i, we can identify every measurement with a POVM and the probability for a specific outcome with an effect operator. A measurement with two outcomes is determined by one effect operator E alone, as the probability for the second outcome is determined by $\mathbb{1}_d - E$ due to normalization. Therefore, we can identify binary measurements with effect operators.

If, however, we are interested in the measurement outcomes, an observable A is given as a set of measurement outcomes $\{a_i\}_{i\in[m]}$ and a POVM $(E_1, \ldots, E_m)$. We can compute the expectation values with respect to this observable as

$$\langle A\rangle_\rho = \sum_{i=1}^{n} p_i a_i = \sum_{i=1}^{n} a_i \mathrm{Tr}\left(\rho E_i\right).$$

The above can be written more concisely using an operator $\hat{A} \in \mathcal{M}_d^{\mathrm{herm}}$ defined as

$$\hat{A} := \sum_{i=1}^{n} a_i E_i.$$

By using the linearity of the trace, it follows that $\langle A\rangle_\rho = \mathrm{Tr}(\rho\hat{A})$. This connects POVMs to the usual notion of observables in quantum mechanics which can be found in textbooks such as [12]. There, observables are identified with Hermitian operators and the associated measurements follow from the spectral decomposition. Let $B \in \mathcal{M}_d^{\text{herm}}$ be an operator with spectral decomposition

$$B = \sum_{i=1}^{m} \lambda_i P_i,$$

where the $\lambda_i \in \mathbb{R}$ are the eigenvalues, the P_i the orthogonal projections onto the respective eigenspaces, and $m \in \mathbb{N}$. The corresponding measurement is then assumed to have outcomes $\{\lambda_i\}_{i=1}^m$ with probabilities

$$q_i := \operatorname{Tr}(\rho P_i) \qquad \forall i \in [m].$$

The projectors P_i clearly form a POVM, but POVM elements need not be projectors in general. We refer to a POVM in which all effect operators are projections as a *projective measurement*. Therefore, POVMs generalize projective measurements. This generalization allows us to perform tasks such us unambiguous discrimination of non-orthogonal pure states (see e.g. [5, Section 2.2.6]).

2.2 Composite systems

Now that we have defined states and measurements on single systems, we can consider composite systems. Let A be a system with Hilbert space $\mathbb{C}^{d_A}$, $d_A \in \mathbb{N}$, and B one with Hilbert space $\mathbb{C}^{d_B}$, $d_B \in \mathbb{N}$. Then, the joint system AB has Hilbert space $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ and dimension $d_A d_B$. Observables and states on the composite system AB are defined with respect to this larger Hilbert space. In order to consider subsystems of this system, let us define the partial trace over a subsystem.

Definition 2.2. *Let d_A, $d_B \in \mathbb{N}$. Then, the* partial trace over the system B *is the map* $\operatorname{Tr}_B(\rho_{AB}) : \mathcal{M}_{d_A} \otimes \mathcal{M}_{d_B} \to \mathcal{M}_{d_A}$ *which satisfies*

$$\operatorname{Tr}(\operatorname{Tr}_B(X)\, Y) = \operatorname{Tr}(X(Y \otimes \mathbb{1}_B)) \qquad \forall X \in \mathcal{M}_{d_A} \otimes \mathcal{M}_{d_B},\ \forall Y \in \mathcal{M}_{d_A}.$$

Then, the *reduced state* ρ_A on the system A alone is obtained from a state $\rho_{AB} \in \mathcal{S}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$ as $\rho_A := \operatorname{Tr}_B(\rho_{AB})$. An important example of a bipartite state with $d_A = d_B = d$ is the *maximally entangled state* Ω. For an orthonormal basis $\{|i\rangle\}_{i\in[d]}$ of $\mathbb{C}^d$, it is defined as

$$\Omega := \frac{1}{d} \sum_{i,j=1}^{d} |i\rangle\langle j| \otimes |i\rangle\langle j| \,. \tag{2.1}$$

The reduced states of Ω on the first and second subsystem are both $\mathbb{1}/d$. This example shows that the maximally mixed state arises as the reduced state of a pure state, the maximally entangled state. The following proposition demonstrates that any mixed state arises from a pure state on a larger system (see e.g. [5, Section 2.5] for a proof).

Proposition 2.3 (Purification)**.** *Let $\rho \in \mathcal{S}(\mathbb{C}^{d_A})$, $d_A \in \mathbb{N}$. Then, there exists a system B of dimension $d_B \in \mathbb{N}$ and a pure state $|\psi\rangle\langle\psi| \in \mathcal{S}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$ such that $\rho = \operatorname{Tr}_B(|\psi\rangle\langle\psi|)$. The state $|\psi\rangle\langle\psi|$ is called a* purification *of ρ. We can choose $d_B = \operatorname{rank}\rho$. If $|\varphi\rangle\langle\varphi| \in \mathcal{S}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d'_B})$ is another purification of ρ with $d'_B \geq d_B$, $d'_B \in \mathbb{N}$, there exists an isometry $V : \mathbb{C}^{d_B} \hookrightarrow \mathbb{C}^{d'_B}$ such that $|\varphi\rangle = (\mathbb{1}_{d_A} \otimes V)\,|\psi\rangle$.*

The last part of the proposition shows that all purifications are equivalent up to isometries and it usually does not matter which purification we choose.

In the same way as we can see mixed states as pure states reduced to a subsystem, we can understand the relation between the projective measurements usually used in quantum physics and the more general concept of POVMs.

Theorem 2.4 ([8, Theorem 2.42]). *Let d, $n \in \mathbb{N}$, and $(E_1, \ldots, E_n)$ be a d-dimensional POVM. Then, there exists an isometry $V : \mathbb{C}^d \hookrightarrow \mathbb{C}^d \otimes \mathbb{C}^n$ such that*

$$E_i = V^* \left(\mathbb{1}_d \otimes |i\rangle\langle i|\right) V \qquad \forall i \in [n],$$

where $\{|i\rangle\}_{i=1}^n$ is an orthonormal basis of $\mathbb{C}^n$.

The above theorem states that any POVM can be seen as a projective measurement restricted to a subsystem. It is a simpler version of Naimark's theorem [13, Theorem 4.6].

In bipartite systems, we can witness that there is an observable difference between quantum and classical correlations. This difference shows in the violation of Bell inequalities (see [14] for a review). The arguably best-known Bell inequality is the CHSH-inequality, which has been named after its inventors Clauser, Horne, Shimony and Holt [15]. Let A_i, B_j be observables on the systems $\mathbb{C}^{d_A}$ and $\mathbb{C}^{d_B}$, respectively, for i, $j \in \{1, 2\}$. Each observable is defined by a binary POVM with outcomes $\{+1, -1\}$. To obtain observables on the bipartite system $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$, we consider the tensor products $A_i \otimes B_j$. Let $\rho \in \mathcal{S}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$. Then, the CHSH inequality is given as

$$\langle A_1 \otimes B_1\rangle_\rho + \langle A_1 \otimes B_2\rangle_\rho + \langle A_2 \otimes B_1\rangle_\rho - \langle A_2 \otimes B_2\rangle_\rho \leq 2. \tag{2.2}$$

This inequality is satisfied by all quantum states admitting a local hidden variable model. These are all states which we can describe using classical probability theory under the assumption that outcomes on one system do not influence the outcomes on the other system (see [9, Section 2.2.4] for details). In quantum mechanics, however, the left hand side of Equation (2.2) can attain values up to $2\sqrt{2}$. The fact that such assignments of states and measurements exist has been demonstrated experimentally [16]. The violation of Bell inequalities proves the existence of *entanglement* in quantum mechanics. However, a discussion of quantum entanglement is beyond the scope of this thesis and we refer the reader to [17] instead.

2.3 Quantum operations

We now turn to operations which map quantum states to quantum states. These are the valid operations which we can theoretically implement on a physical system.

Definition 2.5. *Let d, D, $n \in \mathbb{N}$ and let $\Phi : \mathcal{M}_d \to \mathcal{M}_D$ be a linear map. Then, Φ is*

- positive *if $\Phi(A) \geq 0$ for all $A \geq 0$.*
- n-positive *if the map $\Phi \otimes \mathrm{id}_n : \mathcal{M}_d \otimes \mathcal{M}_n \to \mathcal{M}_D \otimes \mathcal{M}_n$ is positive.*
- completely positive *(CP) if Φ is n-positive for all $n \in \mathbb{N}$.*
- trace-preserving *if $\mathrm{Tr}\,(\Phi(A)) = \mathrm{Tr}\,(A)$ for all $A \in \mathcal{M}_d$.*
- unital *if $\Phi(\mathbb{1}_d) = \mathbb{1}_D$.*

As states are positive operators, quantum operations have to conserve this property. We thus require a quantum operation to be a positive map. Moreover, it should not matter whether we act with an operation Φ on a quantum system A alone or whether we see A as part of a larger system AB on which we act with $\Phi \otimes \mathrm{id}_{d_B}$, since we only act non-trivially on the subsystem A. Here, B is an arbitrary system of finite dimension d_B. Therefore, we require a quantum operation to be not only positive but completely positive. Furthermore, it has to be trace preserving since quantum states have trace 1. A map which is completely positive and trace preserving (CPTP) is called a *quantum channel*.

Example 2.6 (Measurement channel)**.** *Let $E := (E_1, \ldots, E_k)$ be a d-dimensional POVM, where $k,\ d \in \mathbb{N}$. Then, the* measurement channel *$\Psi_E : \mathcal{M}_d \to \mathcal{M}_k$ corresponding to E is defined as*

$$\Psi_E(A) := \sum_{i=1}^{k} \mathrm{Tr}\left(E_i A\right) |i\rangle\langle i| \,.$$

Here, $\{|i\rangle\}_{i\in[k]}$ is the canonical basis of $\mathbb{C}^k$. Note that the above channel maps quantum states to classical probability distributions.

We have already noted that the $n \times n$-matrices with the Hilbert-Schmidt inner product form a Hilbert space. To any linear map Φ as in Definition 2.5, there is a dual map Φ^* with respect to the Hilbert-Schmidt inner product. The dual map is defined by

$$\mathrm{Tr}\left(B\Phi(A)\right) = \mathrm{Tr}\left(\Phi^*(B)A\right) \qquad \forall A \in \mathcal{M}_d,\ \forall B \in \mathcal{M}_D.$$

By the above, Φ is trace preserving if and only if Φ^* is unital and a map is completely positive if and only if its dual map is completely positive [7, Section 4.1.2]. This duality allows us to either talk about a quantum operation as a CPTP map acting on states or as a unital CP map acting on observables. This means that we can either see the quantum operation as part of the preparation or the measurement procedure. The former point of view is called the Schrödinger picture and the latter is called the Heisenberg picture.

The following theorem gives a very useful criterion to determine whether a given map is completely positive [18].

Theorem 2.7 ([8, Theorem 2.22])**.** *Let $\Phi : \mathcal{M}_d \to \mathcal{M}_D$ be a linear map. The* Choi matrix *$J(\Phi)$ of Φ is defined as*

$$J(\Phi) := (\Phi \otimes \mathrm{id}_d)(\Omega).$$

Here, Ω is the maximally entangled state from Equation (2.1)*. Then, $J(\Phi)$ is a positive semidefinite matrix if and only if Φ is completely positive.*

Furthermore, it is very useful for us that every completely positive map can be written in a special form, its *Kraus representation* [19].

Theorem 2.8 ([5, Theorem 8.1])**.** *Let $\Phi : \mathcal{M}_d \to \mathcal{M}_D$ be a linear map. Then, Φ is completely positive if and only if there exist operators $\{K_i\}_{i\in[m]} \subset \mathcal{M}_{D,d}$, $0 \leq m \leq Dd$, such that*

$$\Phi(A) = \sum_{i=1}^{m} K_i A K_i^* \qquad \forall A \in \mathcal{M}_d.$$

The K_i are called Kraus operators*. If Φ is trace preserving, then $\sum_{i=1}^{m} K_i^* K_i = \mathbb{1}_d$.*

The above theorem immediately shows that the measurement channel defined in Example 2.6 is completely positive. The Choi matrix and the Kraus representation make it a routine task to verify that a given linear map is completely positive.

In this chapter, we have reviewed some basic concepts from quantum information theory which are used throughout this thesis. In the next chapter, we continue the study of completely positive maps from a more abstract point of view.

3 Extensions and fixed-points of completely positive maps

Completely positive maps are not only important in quantum information theory, but also in the study of operator algebras and operator spaces. In this chapter, we review some results of a more operator-algebraic nature on map extensions and fixed-points of completely positive maps. These results form the basis of our work on the compression of quantum states in Chapter 6 and on the incompatibility of quantum measurements in Chapter 8.

Definition 3.1 (Matrix algebra)**.** *Let $n \in \mathbb{N}$. A* matrix algebra *$\mathcal{A}$ is a subalgebra of $\mathcal{M}_n$ such that $\mathbb{1} \in \mathcal{A}$ and $A \in \mathcal{A}$ if and only if $A^* \in \mathcal{A}$.*

Equivalently, we can think of a matrix algebra equipped with the operator norm as a finite-dimensional C* or von Neumann algebra, since the topologies which distinguish the two coincide in finite dimensions. All matrix algebras can be brought into the following standard form:

Theorem 3.2 ([20, Theorem 5.6])**.** *Let $n \in \mathbb{N}$ and let $\mathcal{A} \subseteq \mathcal{M}_n$ be a matrix algebra. Then, there is a unitary $U \in \mathcal{M}_n$, an $m \in \mathbb{N}$, and positive integers k_i, d_i, $i \in [m]$, such that $k_1 d_1 + \ldots + k_m d_m = n$ and*

$$\mathcal{A} \simeq U^* \left(\bigoplus_{i=1}^{m} \mathcal{M}_{d_i} \otimes \mathbb{1}_{k_i} \right) U.$$

Hitherto, we have always defined completely positive maps between full matrix algebras $\mathcal{M}_d$ and $\mathcal{M}_D$, d, $D \in \mathbb{N}$. Sometimes, however, we are interested in a set of observables which come only from a matrix subalgebra or from a linear subspace of $\mathcal{M}_d$ containing the identity, a so called operator system.

Definition 3.3 (Operator system)**.** *Let $\mathcal{A}$ be a matrix algebra and let $\mathcal{L} \subseteq \mathcal{A}$ be a linear subspace containing $\mathbb{1}$ and such that $A \in \mathcal{L}$ if and only if $A^* \in \mathcal{L}$. Then, $\mathcal{M}$ is called an* operator system*.*

As in Definition 2.5, we can now define n-positive maps on a linear subspace $\mathcal{L}$ of $\mathcal{M}_d$, where $n \in \mathbb{N}$. Note that an element from the subspace is positive if it is a positive semidefinite matrix. Let $\Phi : \mathcal{L} \to \mathcal{M}_D$ be a linear map. Then, Φ is positive, n-positive, completely positive or unital if the corresponding condition from Definition 2.5 holds with $\mathcal{M}_d$ replaced by $\mathcal{L}$. In particular, these definitions apply to operator systems. The following proposition shows that we do not need to check n-positivity for all n to conclude that Φ is completely positive.

Proposition 3.4 ([13, Theorem 6.1])**.** *Let $\mathcal{L}$ be an operator system and $\Phi : \mathcal{L} \to \mathcal{M}_D$ be a linear map for $D \in \mathbb{N}$. Then, Φ is completely positive if and only if it is D-positive.*

The next theorem connects complete positivity on operator systems to complete positivity on matrix algebras.

Theorem 3.5 ([13, Theorem 6.2])**.** *Let $D \in \mathbb{N}$, $\mathcal{A}$ be a matrix algebra, $\mathcal{L} \subseteq \mathcal{A}$ be an operator system and $\Phi : \mathcal{L} \to \mathcal{M}_D$ be a completely positive map. Then, there exists a completely positive map $\Psi : \mathcal{A} \to \mathcal{M}_D$ which extends Φ.*

This is a finite-dimensional version of Arveson's extension theorem [13, Theorem 7.5]. Note that the extension is usually not unique.

As classical states are diagonal matrices, classical systems correspond to commutative matrix algebras [9, Section 2.1.3]. Commutative algebras have $d_i = 1$ for all $i \in [m]$ in Theorem 3.2 and we write $\mathbb{C}^d$ for the algebra of diagonal $d \times d$ matrices. For evolutions between two systems where one is classical, there is no difference between positivity and complete positivity.

Theorem 3.6 ([13, Theorem 3.9 and 3.11])**.** *Let $\mathcal{A}$, $\mathcal{B}$ be two matrix algebras and $\Phi : \mathcal{A} \to \mathcal{B}$ be a positive linear map. If $\mathcal{A}$ or $\mathcal{B}$ is commutative, then Φ is completely positive.*

In the rest of this chapter, we are concerned with fixed-points of completely positive maps.

Definition 3.7. *Let $\Phi : \mathcal{M}_d \to \mathcal{M}_d$ be a linear map, $d \in \mathbb{N}$. Then, the fixed-point set of Φ is*

$$\mathcal{F}_\Phi := \{A \in \mathcal{M}_d : A = \Phi(A)\}.$$

By Brouwer's fixed-point theorem (e.g. [21, p. 73]), it is clear that $\mathcal{F}_\Phi$ is non-empty if Φ is a quantum channel (e.g. [22, Theorem 6.11]). If Φ is unital, then $\mathbb{1}_d \in \mathcal{F}_\Phi$. It follows from results in [23] that fixed-points of quantum channels almost have the form of a matrix algebra (see also [22, Theorem 6.14]):

Theorem 3.8. *Let $\Phi : \mathcal{M}_d \to \mathcal{M}_d$, $d \in \mathbb{N}$, be a quantum channel. Then, there is a unitary $U \in \mathcal{M}_d$, an $m \in \mathbb{N}$, and positive integers d_0, d_i, k_i, $i \in [m]$, such that $d_0 + k_1 d_1 + \ldots + k_m d_m = d$ and*

$$\mathcal{F}_\Phi = U \left(0 \oplus \bigoplus_{i=1}^{m} \mathcal{M}_{d_i} \otimes \rho_i \right) U^*.$$

Here, the $\rho_i \in \mathcal{S}(\mathbb{C}^{k_i})$, $i \in [m]$, are positive definite states and the zero block has dimension d_0.

If we had $\rho_i = \mathbb{1}_{k_i}/k_i$ for all $i \in [m]$, then $\mathcal{F}_\Phi$ would be a matrix algebra. If Φ has a full-rank fixed-point, then [23] shows that the fixed-point set of the dual map Φ is a matrix algebra (see also [22, Theorem 6.12]).

Theorem 3.9. *Let $\Phi : \mathcal{M}_d \to \mathcal{M}_d$, $d \in \mathbb{N}$, be a quantum channel with full-rank fixed-point. Then, $\mathcal{F}_{\Phi^*}$ is a matrix algebra and therefore has the form in Theorem 3.2.*

If Φ does not have a full-rank fixed-point, $\mathcal{F}_{\Phi^*}$ is not necessarily a matrix algebra, as the example $\Phi^* : \mathcal{M}_3 \to \mathcal{M}_3$,

$$\Phi(A) = \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & \frac{1}{2}(a_{11} + a_{22}) \end{pmatrix} \qquad \forall A = (a_{ij})_{i,j=1}^3 \in \mathcal{M}_3$$

from [24, p. 288, Remark 2] shows. The above map is completely positive and unital, but its fixed-points do not form an algebra. Therefore, the fixed-points of a unital CP map only form an operator system in general. However, Arveson showed in [24] that we can infer more if this operator system generates the full matrix algebra.

Theorem 3.10 ([24, p. 288, Remark 2])**.** *Let $\Psi : \mathcal{M}_d \to \mathcal{M}_d$, $d \in \mathbb{N}$, be a unital CP map whose fixed-points algebraically generate $\mathcal{M}_d$. Then, Ψ is the identity map.*

In Core Article I [1], we generalize the proof technique of Theorem 3.10 to the setting in which the fixed-points of Ψ only generate a matrix subalgebra. This allows us to prove the lower bound on the compression dimension obtained in Theorem 6.1 of Core Article I [1]. We elaborate on this in Chapter 6.

4 Semidefinite programming

We have seen in the previous chapter how completely positive maps can be defined on linear subspaces, in particular on operator systems. For completely positive maps on full matrix algebras, Theorem 2.7 enables us to verify efficiently whether a given map is completely positive. For linear maps on proper linear subspaces, the Choi matrix no longer exists. However, it is possible to verify complete positivity algorithmically by means of a semidefinite program. This is essential for our work on the compression of quantum states in Chapter 6 and for the theory of free spectrahedra, which is the topic of Chapter 7.

In Section 4.1, we start with a brief introduction to semidefinite programming. Most of the material is standard and can be found in textbooks such as [8, 25, 26]. Semidefinite programming is then applied to the verification of complete positivity in Section 4.2. Finally, we review some of the methods used to solve semidefinite programs in Section 4.3 and compare their complexity.

4.1 Basic notions and duality

Semidefinite programs form a class of convex optimization problems which can be solved efficiently under suitable conditions (see Chapter 4.3). As such, they also play an important role in quantum information theory, both for numerical and analytical results. For example, they can be used to compute completely bounded norms of quantum channels [27] or to detect entanglement [28]. For us, they are very useful tools in Chapters 6, 7 and 8.

Let us start by giving a semidefinite program in standard form:

Definition 4.1. *Let m, $d \in \mathbb{N}$ and A, $B_i \in \mathcal{M}_d$ for all $i \in [m]$. Moreover, let $b_i \in \mathbb{R}$ for all $i \in [m]$. Then, the* semidefinite program (SDP) in standard form *defined by A, B_i and b_i is the optimization problem*

$$\begin{aligned} \text{maximize} \quad & \operatorname{Tr}(AX) \\ \text{subject to} \quad & \operatorname{Tr}(B_i X) \leq b_i \qquad \forall i \in [m] \\ & X \geq 0, \; X \in \mathcal{M}_d. \end{aligned} \tag{4.1}$$

This is also called the primal form *of an SDP.*

In textbooks on optimization theory such as [25, 26], the matrices in Definition 4.1 are often required to be real symmetric matrices. However, this is not necessary for the theory presented here. For the problems arising in quantum information theory, it is more natural to work with Hermitian matrices with complex entries. Moreover, many authors consider the SDP in Equation (4.1) with equality instead of inequality constraints to be the standard form. It is, however, straightforward to convert a problem with inequality constraints to one with equality constraints and vice versa using slack variables. We refer to [8, Section 1.2.3] for details. Before we continue, let us consider an example of a problem which can be cast as an SDP (see e.g. [29, Equation (2.13)]).

Example 4.2. *Let m, $n \in \mathbb{N}$ and $A \in \mathcal{M}_{m,n}$. Then, the SDP*

$$\begin{aligned} \text{maximize} \quad & \operatorname{Tr}\left(\begin{pmatrix} 0 & A \\ A^* & 0 \end{pmatrix} X\right) \\ \text{subject to} \quad & \operatorname{Tr}(X) = 1 \\ & X \geq 0, \; X \in \mathcal{M}_{m+n} \end{aligned}$$

has the solution $\|A\|_\infty$. Note that $\operatorname{Tr}(X) = 1$ *can be written as two inequality constraints.*

Semidefinite programs have a rich duality theory. Let us start by defining the *dual problem* to the SDP given in Definition 4.1.

Definition 4.3. *Let m, $d \in \mathbb{N}$ and A, $B_i \in \mathcal{M}_d^{\text{herm}}$ for all $i \in [m]$. Additionally, let $b_i \in \mathbb{R}$ for all $i \in [m]$ and $b = (b_1, \ldots, b_m)$. Then, the* dual form *of an SDP is*

$$\begin{aligned} \text{minimize} \quad & \langle y, b \rangle \\ \text{subject to} \quad & \sum_{i=1}^{m} y_i B_i - A \geq 0 \\ & y \in \mathbb{R}_+^m. \end{aligned} \tag{4.2}$$

A matrix $X \geq 0$ which satisfies the constraints in Equation (4.1) is called *feasible* for the primal problem and a vector $y \in \mathbb{R}_+^m$ which fulfills the constraints in Equation (4.2) is called feasible for the dual problem. If such a feasible element exists, we call the SDP itself feasible. If α is the optimal value for the SDP in Definition 4.1, there is not always a feasible X attaining it even if α is finite. The same is true for the dual SDP [26, p. 19f.].

Let α and β be the optimal values of the primal and the dual SDP, respectively. If the respective SDP is not feasible, we set $\alpha = -\infty$ and $\beta = \infty$. It can be shown that we always have $\beta \geq \alpha$ (see [27, Theorem 2.1], [25, Section 5.2.2]). This is called *weak duality.* In many problems of interest, we even have $\beta = \alpha$. This is called *strong duality.* A very useful criterion which implies strong duality is Slater's condition (see [27, Theorem 2.2], [25, Section 5.2.3]).

Theorem 4.4 (Slater's condition)**.** *Let m, $d \in \mathbb{N}$ and A, $B_i \in \mathcal{M}_d^{\text{herm}}$ for all $i \in [m]$. Moreover, let $b_i \in \mathbb{R}$ for all $i \in [m]$. Let α and β be the optimal solutions to the primal and dual SDP defined by A, B_i, and b_i, respectively.*

1. *If the dual SDP is feasible and there exists an $X \in \mathcal{M}_d^{\text{herm}}$, $X > 0$, such that* $\operatorname{Tr}(B_i X) < b_i$ *for all $i \in [m]$, then $\alpha = \beta$ and there exists a $y \in \mathbb{R}_+^m$ such that $\langle y, b \rangle = \beta$.*
2. *If the primal SDP is feasible and there exists a $y \in \mathbb{R}_+^m$ with strictly positive entries such that $\sum_{i=1}^m y_i B_i - A > 0$, then $\alpha = \beta$ and there exists an $X \in \mathcal{M}_d^{\text{herm}}$, $X \geq 0$, such that* $\operatorname{Tr}(AX) = \alpha$.

4.2 Verifying complete positivity

Among the many applications of SDPs to quantum information theory, there is one which is particularly useful to us in the following chapters. Let D, d, $g \in \mathbb{N}$. Assume that we are given a collection of Hermitian matrices $A_i \in \mathcal{M}_D^{\text{herm}}$, $B_j \in \mathcal{M}_d^{\text{herm}}$, where i, $j \in [g]$. Let $\mathcal{L} := \operatorname{span}\{A_i : i \in [g]\} \subseteq \mathcal{M}_D$. We would like to know whether the linear map $\Phi : \mathcal{L} \to \mathcal{M}_d$ defined by $\Phi : A_i \mapsto B_i$ for all $i \in [g]$ is completely positive. For example, this is interesting if we would like to know whether a desired transformation can be implemented by a quantum operation. Note that we could require the map to be unital by adding $A_0 = \mathbb{1}_D$, $B_0 = \mathbb{1}_d$ such that we can check whether Φ can be extended to a quantum operation in the Heisenberg picture by Theorem 3.5. For the case $\mathcal{L} = \mathcal{M}_D$, complete positivity of Φ is easy to verify, since we can check whether the Choi matrix $J(\Phi)$ is positive semidefinite (see Theorem 2.7). For $\mathcal{L} \subsetneq \mathcal{M}_D$, there is no Choi matrix, but complete positivity can be verified using an SDP as shown in [30]:

Theorem 4.5. *Let D, d, $g \in \mathbb{N}$. Moreover, let $A_i \in \mathcal{M}_D$, $B_j \in \mathcal{M}_d$, where i, $j \in [g]$. We define a linear map $\Phi : \mathrm{span}\{A_1, \dots, A_g\} \to \mathcal{M}_d$ as $\Phi : A_i \mapsto B_i$ for all $i \in [g]$. Consider the following SDP with optimal value β:*

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{g} \mathrm{Tr}\left(B_i^T H_i\right) \\ \text{subject to} \quad & \sum_{i=1}^{g} A_i \otimes H_i \geq 0 \\ & H_i \in \mathcal{M}_d^{\mathrm{herm}} \qquad \forall i \in [g]. \end{aligned} \tag{4.3}$$

Then, Φ is completely positive if and only if $\beta = 0$. Otherwise, $\beta = -\infty$.

The SDP in Theorem 4.5 is not given in standard form, but it can be converted to the dual form using a Hermitian basis of the real vector space of Hermitian operators (see e.g. [8, Equation (1.103)]). It is then possible to express the H_i in terms of this basis and subsequently optimize over vectors in $\mathbb{R}^{gd^2}$ (which we could split into the positive and the negative part).

Since the assignment $H_1 = \ldots = H_g = 0$ satisfies the constraints in Equation (4.3), the SDP is always feasible. If β is the optimal value of this SDP, then this implies $\beta \leq 0$. Furthermore, the SDP is homogeneous in the sense that $(H_1, \dots, H_g)$ being feasible implies that $\lambda(H_1, \dots, H_g)$ is feasible for all $\lambda \geq 0$. Therefore, we either have $\beta = 0$ or $\beta = -\infty$. Theorem 1 of [30] shows that Φ is completely positive on $\mathcal{L}$ if and only if $\beta = 0$. If $\mathcal{L}$ contains the identity operator, this is equivalent to the existence of a completely positive extension of Φ to $\mathcal{M}_D$ by Theorem 3.5. The SDP also returns $\beta = -\infty$ if the map Φ is not well-defined. Furthermore, it is possible to add the constraint $\|H_i\|_\infty \leq 1$ for all $i \in [g]$ to the SDP in Theorem 4.5 such that β is always finite and either $\beta = 0$ or $\beta < 0$. Using Slater's condition (Theorem 4.4), it is possible to show that in this case, the SDP exhibits strong duality [30, Proposition 2].

4.3 Algorithms to solve semidefinite programs

We conclude this chapter with a discussion of methods to solve SDPs, with an emphasis on the complexity of these methods. In general, it is not known whether SDPs are solvable in polynomial time [31] (see also [26, Section 2.6]). However, under mild assumptions about the set of feasible points of the SDP, it can be shown that it is possible to find an approximate solution to the SDP in polynomial time. For simplicity, we only discuss problems in primal form. Similar results hold for SDPs in dual form. Assume that each feasible X fulfills $\|X\|_2 \leq R$ for some $R \in \mathbb{R}$. Furthermore, assume that the set of feasible points is either empty or contains a ball of radius $r \in \mathbb{R}$ in Hilbert-Schmidt norm. Then, the *ellipsoid method* allows us to either find a feasible point X_0 which is ϵ-close to optimal for some $\epsilon > 0$ or to conclude that the problem is infeasible [32, Chapter 3]. Being ϵ-close to optimal means that $\alpha - \mathrm{Tr}\,(AX_0) \leq \epsilon$ where $\alpha \in \mathbb{R}$ is the optimal value. If R, r and ϵ only have polynomial dependence on d and m, the algorithm runs in time $\tilde{\mathcal{O}}(\max\{m, d^2\}d^6)$ [33, Section 2.2]. Here, the $\tilde{\mathcal{O}}$ notation hides factors with polylogarithmic dependence on m, d. Although the ellipsoid method is the strongest theoretical result, it is rarely used in applications, since in practice *interior point methods* are much faster [34]. The idea is to introduce a suitable barrier function into the optimization problem which penalizes closeness to the boundary of the semidefinite cone. Subsequently, we optimize this new function using Newton's method, whose rapid convergence is ensured by properties of the barrier function. Assuming that the primal and the dual problem are both strictly feasible, such methods have complexity $\tilde{\mathcal{O}}(\max\{m^3, d^2m^2, md^\omega\}\sqrt{d})$, where ω is the exponent of matrix multiplication [34, Chapter 5]. The best known upper bounds yield $\omega < 2.38$ [35]. The more involved state-of-the-art algorithm in [36, Theorem 10] even has complexity $\tilde{\mathcal{O}}(\max\{m^2, mds, d^\omega\}m)$, assuming bounds both on the primal and dual feasible set. Here, $s \in \mathbb{N}$ is the row sparsity of $A, B_1, \ldots, B_m$. Recently, there has been interest in solving SDPs on a quantum computer, since this gives an unconditional speed-up over classical methods [37].

Although for many SDPs there are algorithms which solve the problem in polynomial time, the storage requirements and the time complexity make it challenging to solve large instances of SDPs in practice. In the next chapter, we review how probabilistic methods known as sketching can be used to alleviate these issues for convex optimization problems.

5 Sketching and convex optimization

High-dimensional matrices are not only a challenge in quantum mechanics, but also in classical applications which need to handle a large amount of data. If A, $B \in \mathcal{M}_n$ and $n \in \mathbb{N}$, then operations such as computing the inverse of A or the matrix product AB are computationally very expensive, although the operations are feasible in polynomial time. Recently, the technique of linear sketching has been successfully used to accelerate such computations (see [38] for a topical survey). Following [38, Section 1], we first illustrate these techniques in Section 5.1 using the example of linear regression before discussing in Section 5.2 how sketching techniques can be used for optimization problems.

5.1 An introduction to sketching techniques

Let A be an $n \times d$ matrix such that $A^T = (a_1, \ldots, a_n)$, where $a_i \in \mathbb{R}^d$ for all $i \in [n]$. Here, $d \in \mathbb{N}$ is the number of parameters of the model at hand and $n \in \mathbb{N}$ is the number of data points. Typically, we have $n \gg d$. Moreover, let $b \in \mathbb{R}^n$. In *linear regression*, the aim is to find

$$\mathrm{argmin}_{x \in \mathbb{R}^d} \|Ax - b\|_2.$$

A solution to the above problem can be found in $\mathcal{O}(nd^2)$ time, but this is too slow for applications if n is extremely large. It is, however, possible to improve on this complexity if one is satisfied to obtain an approximate solution with high probability. That is, for $\epsilon \in (0, 1)$ we want to find $y \in \mathbb{R}^d$ such that

$$\|Ay - b\|_2 \leq (1 + \epsilon)\mathrm{min}_{x \in \mathbb{R}^d} \|Ax - b\|_2 \tag{5.1}$$

with probability at least 9/10. Note that the exact probability does not matter, since we could repeat the procedure to obtain any constant probability we desire. In [39], it has been shown that this goal can be reached by applying a random $k \times n$ matrix S from an appropriate distribution to A and b and subsequently solving the linear regression problem

$$\mathrm{argmin}_{x \in \mathbb{R}^d} \|SAx - Sb\|_2. \tag{5.2}$$

The matrix SA is called a (linear) *sketch* of A. This problem can now be solved in $\mathcal{O}(kd^2)$ time. If $k \ll n$, this is a significant improvement over solving the original problem. The advantage of this approach compared to others with the same aim is that this strategy is a *black box* reduction, since the reduced problem in Equation (5.2) is also a linear regression problem and hence of the same kind as the original one. We can thus use the same techniques to solve both problems.

For this strategy to work, it is essential to find a distribution such that for matrices S drawn according to this distribution, a solution to Equation (5.2) implies that Equation (5.1) holds with high probability. Furthermore, we need to guarantee that SA can be computed efficiently since naïve matrix multiplication of A with a dense matrix S would take $\Theta(knd)$ time, which might be slower than solving the original problem directly. It has been shown in [39] that

S can be taken to be a Fast Johnson-Lindenstrauss transform, which we discuss below. We refer to [39, Theorem 12] for the time complexity of the final algorithm and to [38, Section 1] for a discussion of subsequent improvements. Let us now review some results on Johnson-Lindenstrauss transforms which have proven very useful for sketching.

Definition 5.1 (Johnson-Lindenstrauss transform)**.** *Let k, n, $g \in \mathbb{N}$ and ϵ, $\delta \in (0,1)$. A real random $k \times n$ matrix S is a Johnson-Lindenstrauss transform (JLT) with parameters ϵ, δ, g, or an (ϵ, δ, g)-JLT, if for any subset $V \subset \mathbb{R}^n$ of cardinality at most g, it holds for all v, $w \in V$ that*

$$|\langle Sv, Sw\rangle - \langle v, w\rangle| \leq \epsilon \|v\|_2 \|w\|_2$$

with probability at least $1 - \delta$.

The study of such objects goes back to [40]. A very simple construction of a Johnson-Lindenstrauss transform is the following:

Example 5.2 ([41, Lemma 7], [38, Theorem 2.1])**.** *Let ϵ, $\delta \in (0,1)$ and let $S = \frac{1}{\sqrt{k}} R$, where R is a real $k \times n$ matrix with independent and identically distributed (i.i.d.) Gaussian random variables as entries. If $k = \Omega(\epsilon^{-2} \log(g/\delta))$, then S is an (ϵ, δ, g)-JLT.*

Note that the above example leaves room for improvement in the time needed to compute SA, since the matrix is dense. It is, however, possible to find (ϵ, δ, g)-JLTs with k as above, but only $\mathcal{O}(\epsilon^{-1} \log(g\delta^{-1}))$ non-zero entries per column [42]. This sparsity is close to optimal [43]. Another approach is to choose S not sparse but such that Sx can be computed fast for $x \in \mathbb{R}^n$. These are the Fast Johnson-Lindenstrauss transforms studied in [44], which have been used for linear regression in [39]. We refer to [38, Section 2.1] for a discussion of different JLT constructions and their optimality.

5.2 Sketching for problems in convex optimization

Now we look at applications of sketching techniques to convex optimization. Consider the following problem:

Definition 5.3 (Linear feasibility problem)**.** *Let m, $n \in \mathbb{N}$ such that $m > n$. Let moreover $b \in \mathbb{R}^m$ and let $A = (a_1, \ldots, a_n)$ be a real $m \times n$ matrix, where $a_i \in \mathbb{R}^m$ for all $i \in [n]$. The* linear feasibility problem *specified by (A, b) is to decide whether there is an $x \in \mathbb{R}^n_+$ such that $Ax = b$. If such an x exists, we say that the linear problem is feasible, otherwise that it is infeasible.*

The above problem is a feasibility problem in *linear programming*. A linear program can be seen as a special case of the SDP in Definition 4.1 where all matrices B_i and A are diagonal. Optimization can be reduced to checking feasibility using binary search.

Let $\text{cone}\{a_1, \ldots, a_n\} := \text{conv}\{\lambda a_i : i \in [n], \lambda \geq 0\}$ be the convex cone generated by the a_i. It is easy to see that the linear feasibility problem in Definition 5.3 is equivalent to deciding whether $b \in \text{cone}\{a_1, \ldots, a_n\}$. Let S be a random real $d \times m$ matrix which is an (ϵ, δ, g)-JLT. It was shown in [45, Section 3] that for appropriately chosen parameters (ϵ, δ, g), it holds with high probability that $Sb \notin \text{cone}\{Sa_1, \ldots, Sa_n\}$ if $b \notin \text{cone}\{a_1, \ldots, a_n\}$. Furthermore, $b \in \text{cone}\{a_1, \ldots, a_n\}$ always implies that $Sb \in \text{cone}\{Sa_1, \ldots, Sa_n\}$ by linearity. This means that with high probability the linear problem specified by (A, b) is feasible if and only if the linear problem specified by (SA, Sb) is feasible. Therefore, [45] proves that instead of solving the linear feasibility problem specified by (A, b), we can solve the potentially smaller linear feasibility problem specified by (SA, Sb). A major drawback of this work is that d depends on how well b is separated from $\text{cone}\{a_1, \ldots, a_n\}$ if the problem is not feasible, which is hard to determine in practice.

Another case where sketching techniques have been successfully used for convex optimization is [46]. In this work, the authors consider the convex optimization problem

$$\begin{aligned} \text{minimize} \quad & f(L[X]) \\ \text{subject to} \quad & \text{Tr}\,(X) \leq \alpha \\ & X \geq 0,\ X \in \mathcal{M}_n. \end{aligned} \tag{5.3}$$

Here, $\alpha \in \mathbb{R}_+$, $f : \mathbb{C}^g \to \mathbb{R}$ is a differentiable convex function, and $L : \mathcal{M}_n \to \mathbb{C}^g$ is the linear map

$$L[X] = \left(\text{Tr}\,(A_1^* X), \ldots, \text{Tr}\,(A_g^* X)\right) \qquad \forall X \in \mathcal{M}_n$$

with $A_i \in \mathcal{M}_n$ for all $i \in [g]$ and $n, g \in \mathbb{N}$. A famous problem of this kind is *phase retrieval* which aims at reconstructing a vector up to a global phase shift from noisy quadratic measurements [47, 48]. For $g = 1$, $f = \text{id}$ and a Hermitian matrix A_1, the problem is a semidefinite program. The work [46] provides an algorithm which solves the optimization problem (5.3) with storage $\mathcal{O}(g + rn)$ under the assumption that all solutions to problem (5.3) have rank at most r. The algorithm is based on the *conditional gradient method*. To achieve this task, the algorithm does not store $n \times n$ matrices X, but random sketches of such matrices. Let $k = 2r + 1$, $l = 4r + 3$ and let $S \in \mathcal{M}_{n,k}$, $T \in \mathcal{M}_{l,n}$ be random matrices with i.i.d. entries from the complex Gaussian distribution. The random sketches of X are $Y := XS \in \mathcal{M}_{n,k}$ and $W := TX \in \mathcal{M}_{l,n}$. Previous work [49, Section 6.4] then shows that it is possible to reconstruct a positive semidefinite rank r approximation of X which is close to optimal with high probability.

In our Article III [3], the aim is to find black box reduction for SDPs. Inspired by the work on linear feasibility problems in [45], we consider in Section 4 of Article III [3] the feasibility for a dual SDP. In contrast to [45], we require our sketch to be positive in order to obtain a smaller instance of the same problem and aim to reduce the dimension of the problem instead of the number of constraints. Let D, $m \in \mathbb{N}$. If $A, B_1, \ldots, B_m$ are D-dimensional real symmetric matrices, our aim is to determine whether there is a $y \in \mathbb{R}_+^m$ such that

$$\sum_{i=1}^{m} y_i B_i - A \geq 0. \tag{5.4}$$

Under suitable conditions on $\text{cone}\{B_1, \ldots, B_m\}$, we prove in Theorem 4.2 of Article III [3] that it is sufficient to consider the potentially smaller feasibility problem which is to determine whether there is a $y \in \mathbb{R}_+^m$ such that

$$\sum_{i=1}^{m} y_i S B_i S^T - S A S^T \geq 0. \tag{5.5}$$

Here, S is a real $d \times D$ matrix which is an (ϵ, δ, g)-JLT for an appropriate choice of parameters and $d \in \mathbb{N}$. The relation between the sketched problem in Equation (5.5) and the original problem in Equation (5.4) is similar as for the linear feasibility problem in Definition 5.3. Equation (5.5) is infeasible with high probability if Equation (5.4) is infeasible and feasibility of Equation (5.4) always implies the feasibility of Equation (5.5). If the set

$$\left\{ \sum_{i=1}^{m} y_i B_i - A : y \in \mathbb{R}_+^m \right\}$$

is well separated from the cone of positive semidefinite matrices in the case of infeasibility, we can choose $d \ll D$. Unfortunately, this is hard to check in practice. This is similar to the dependence of d on the distance of b to the cone in [45].

In Section 5 of Article III [3], we show how to approximate the optimal value of a primal SDP as in Definition 4.1 via sketching. Let the matrices A, B_i appearing in Definition 4.1 be real symmetric matrices of dimension $D \in \mathbb{N}$. In Theorems 5.3 and 5.5 of Article III [3], we show that the sketched SDP

$$\begin{aligned} \text{maximize} \quad & \operatorname{Tr}\left(SAS^TY\right) \\ \text{subject to} \quad & \operatorname{Tr}\left(SB_iS^TY\right) \leq b_i + \mu\|B_i\|_1 \qquad \forall i \in [m] \\ & Y \geq 0, Y \in \mathcal{M}_d \end{aligned} \tag{5.6}$$

provides an approximate solution to the SDP in Definition 4.1 with high probability. Here, $\mu \in \mathbb{R}_+$ is an error parameter. Again, S is a real $d \times D$ matrix which is an (ϵ, δ, g)-JLT for an appropriate choice of parameters. For the dimension d of the sketched problem to be significantly smaller than D, we require that there is a constant upper bound on $\|A\|_1$, $\|X_0\|_1$ and $\|B_i\|_1$ for all $i \in [m]$, where X_0 is an optimal solution to the original problem. A significant difference between our work and the methods used in [46] to solve problem (5.3) is that our work provides again a black box reduction. This implies that the smaller problem can be tackled with standard solvers for which reliable implementations exist. In contrast, the authors of [46] modify the conditional gradient method to obtain a new algorithm, which needs to be implemented and tested.

If our results are applicable, the sketched SDPs (5.5) and (5.6) are exponentially smaller than the original problems such that the bottleneck is to compute SAS^T and SB_iS^T. As discussed in Section 6 of Article III [3], our methods provide a speed-up if the original problem requires time $\Omega(mD^{2+\nu})$ to solve, for $\nu > 0$ (assuming that m depends only logarithmically on D). Moreover, the reduced dimension provides an exponential reduction in the storage needed to solve the problem. Although our sketching approach is simple, we prove in Theorem 3.2 and Theorem 5.1 of Article III [3] that our results cannot be significantly improved using positive linear sketches, demonstrating the limitations of our approach.

In this chapter, we have seen how we can use sketching techniques to reduce the dimensionality of classical optimization problems. In the next chapter, we consider the compression of quantum systems.

6 Compression in quantum information theory

One of the fundamental problems in classical information theory is the compression of data, which already appears in Shannon's seminal work [50]. In the realm of quantum information theory, decoherence poses an additional problem which makes it very challenging to store or transmit quantum systems. Therefore, it is especially desirable to compress the quantum information as much as possible. This chapter provides a discussion of different compression results in quantum information theory. In Sections 6.1 to 6.5, we review existing results, before we compare them in Section 6.6 to the results obtained in Core Article I [1].

6.1 Quantum source coding

Arguably the best-known compression result in quantum information theory is quantum source coding or Schumacher compression [51]. It is the quantum counterpart of Shannon's noiseless channel coding theorem [50] (see also [5, Theorem 12.4]). Let $d \in \mathbb{N}$. We consider a device which produces (not necessarily orthogonal) pure quantum states $|\psi_i\rangle\langle\psi_i| \in \mathcal{M}_d$ with probability p_i, where $i \in [g]$ for some $g \in \mathbb{N}$. This device is usually called a *quantum source*. The aim is to find a compressed description of the quantum source in the asymptotic limit of many uses of the source. Let $n \in \mathbb{N}$ and $R \in \mathbb{R}_+$. Then, the compression and decompression maps are given as quantum channels $\mathcal{C}_n : \mathcal{M}_{d^n} \to \mathcal{M}_{2^{nR}}$ and $\mathcal{D}_n : \mathcal{M}_{2^{nR}} \to \mathcal{M}_{d^n}$, respectively. Such maps are allowed to act on several copies of the quantum source at once. In the cases where nR is not an integer, we consider the floor function of this quantity instead. Let $\rho := \sum_{i=1}^{g} p_i \, |\psi_i\rangle\langle\psi_i|$ and let $|\varphi\rangle\langle\varphi| \in \mathcal{S}(\mathbb{C}^d \otimes \mathbb{C}^{d_B})$ be a purification of ρ. Moreover, let

$$T_n := \frac{1}{2}\big\| |\varphi\rangle\langle\varphi|^{\otimes n} - ((\mathcal{C}_n \circ \mathcal{D}_n) \otimes \mathrm{id}_{B^n})[|\varphi\rangle\langle\varphi|^{\otimes n}] \big\|_1,$$

where B^n is the system with Hilbert space $(\mathbb{C}^{d_B})^{\otimes n}$. The quantity $T_n \in [0,1]$ is the trace distance between $|\varphi\rangle\langle\varphi|$ and the state after compression and decompression. It quantifies how well the two states can be distinguished by a measurement [5, Section 9.2].

Let $S(\rho) = -\mathrm{Tr}\,(\rho \log_2 \rho)$ be the *von Neumann entropy* of the average state ρ. Then, the following theorem shows that the optimal compression rate R is the entropy of ρ.

Theorem 6.1 ([51, 52]). *Let $\epsilon \in (0,1)$, $\delta > 0$ and the quantum source as described above. Then, for $R = S(\rho) + \delta$, there is an $N \in \mathbb{N}$ such that $T_n < \epsilon$ for all $n > N$. Conversely, for $R = S(\rho) - \delta$ there is an $N \in \mathbb{N}$ such that $T_n > 1 - \epsilon$ for all $n > N$.*

The achievability has been shown in [51], whereas this version of the converse statement can be found in [52, Theorem 21]. Theorem 6.1 states that for large n, the information produced by the quantum source can be encoded into $S(\rho)$ qubits per use of the source and it is not possible to compress the source further. The achievability of the above theorem relies on the existence of *typical subspaces*. If $\rho = \sum_{i=1}^{d} \lambda_i \, |x_i\rangle\langle x_i|$ is the spectral decomposition of ρ, typical subspaces exist because in the asymptotic limit only a small number of sequences $\lambda_{i_1} \cdot \ldots \cdot \lambda_{i_n}$ occur with high probability, whereas most sequences are very unlikely to appear. Here, $i_j \in [d]$ for all $j \in [n]$. These typical sequences are also used to prove Shannon's noiseless channel coding theorem.

The typical subspace is then the projector onto the $|x_{i_1}\rangle\langle x_{i_1}| \otimes \ldots \otimes |x_{i_n}\rangle\langle x_{i_n}|$ corresponding to typical sequences. We refer the reader to [53, Section 15] for details. Originally, Theorem 6.1 has been proven with a slightly different, but equivalent quantity instead of T_n. See [53, Section 18.4] for a discussion of this and some generalizations of Theorem 6.1.

6.2 Measurement compression

Instead of compressing a quantum source, it is also possible to compress a quantum measurement, using similar ideas. This has been done in a series of works [54, 55, 56]. Let $E := (E_1, \ldots, E_k)$ be a d-dimensional POVM for $d,\, k \in \mathbb{N}$. In this setting, we are looking for a set of d-dimensional POVMs $F^{(i)} := (F_1^{(i)}, \ldots, F_k^{(i)})$, $i \in [g]$, $g \in \mathbb{N}$, and a probability distribution $p = (p_1, \ldots, p_g)$ such that

$$E_j = \sum_{i=1}^{g} p_i F_j^{(i)} \qquad \forall j \in [k]. \tag{6.1}$$

The aim is to find POVMs $F^{(i)}$ such that the number of non-zero effect operators in each $F^{(i)}$ is less than the one in E. This means instead of measuring E we could randomly choose an index i according to the probability distribution p and subsequently perform the corresponding measurement $F^{(i)}$ which has fewer outcomes than E. In this way, we have decomposed the measurement E into these two parts: The random choice of the measurement is classical noise, whereas the outcome of the measurement $F^{(i)}$ is the meaningful part of the quantum measurement.

As in the setting of source coding, we do not aim for an exact decomposition as in Equation (6.1), but for one with asymptotically vanishing error in the measurement statistics for some fixed $\rho \in \mathcal{S}(\mathbb{C}^d)$. Moreover, we allow the $F^{(i)}$ to act on several copies of ρ. Let $n \in \mathbb{N}$ and consider the measurement channels $\Psi_E^{\otimes n} : \mathcal{M}_{d^n} \to \mathcal{M}_{k^n}$ and $\Phi : \mathcal{M}_{d^n} \to \mathcal{M}_{k^n}$ defined as follows:

$$\Psi_E^{\otimes n}(A) := \sum_{j_1,\ldots,j_n=1}^{k} \operatorname{Tr}\left(E_{j_1,\ldots,j_n} A\right) |j_1\rangle\langle j_1| \otimes \ldots \otimes |j_n\rangle\langle j_n| \qquad \forall A \in \mathcal{M}_{d^n},$$
$$\Phi(A) := \sum_{j_1,\ldots,j_n=1}^{k} \operatorname{Tr}\left(F_{j_1,\ldots,j_n} A\right) |j_1\rangle\langle j_1| \otimes \ldots \otimes |j_n\rangle\langle j_n| \qquad \forall A \in \mathcal{M}_{d^n}.$$

Here, $E_{j_1,\ldots,j_n} := E_{j_1} \otimes \ldots \otimes E_{j_n}$, $F_{j_1,\ldots,j_n} := \sum_{i=1}^{g} p_i F^{(i)}_{j_1,\ldots,j_n}$ and $\{|j_i\rangle\}_{j_i=1}^{k}$ is the standard basis of $\mathbb{C}^k$. For an error $\epsilon > 0$, we aim for

$$\frac{1}{2}\left\|(\Phi \otimes \mathrm{id}_{B^n})(|\varphi\rangle\langle\varphi|^{\otimes n}) - (\Psi_E^{\otimes n} \otimes \mathrm{id}_{B^n})(|\varphi\rangle\langle\varphi|^{\otimes n})\right\|_1 \leq \epsilon, \tag{6.2}$$

where $|\varphi\rangle\langle\varphi| \in \mathcal{S}(\mathbb{C}^d \otimes \mathbb{C}^{d_B})$ is a purification of ρ, $d_B \in \mathbb{N}$. The main result of [56] is the following:

Theorem 6.2 ([56, Theorem 2]). *Let $d,\, k,\, n \in \mathbb{N}$. For $\rho \in \mathcal{S}(\mathbb{C}^d)$ and the d-dimensional POVM $(E_1, \ldots, E_k)$ we define a quantum source which produces states $\hat{\rho}_j$ with probability λ_j for $j \in [k]$, where*

$$\lambda_j = \operatorname{Tr}\left(\rho E_j\right), \qquad \hat{\rho}_j = \frac{1}{\lambda_j}\sqrt{\rho} E_j \sqrt{\rho}.$$

Then, there exist POVMs $F^{(i)}$ with outcomes in $[k]^n$, $i \in [g]$, each with at most m non-zero effect operators, such that for $F = \sum_{i=1}^{g} \frac{1}{g} F^{(i)}$ Equation (6.2) is satisfied for all $\epsilon > 0$ and sufficiently large n. Here,

$$m = 2^{nI(\lambda;\hat{\rho})+\mathcal{O}(\sqrt{n})} \quad \text{and} \quad g = 2^{n(S(\lambda)-I(\lambda;\hat{\rho}))+\mathcal{O}(\sqrt{n})},$$

with $I(\lambda;\hat{\rho}) = S(\rho) - \sum_{j=1}^{g} \lambda_j S(\hat{\rho}_j)$ and $S(\lambda) = -\sum_{j=1}^{g} \lambda_j \log_2 \lambda_j$.

Note that $S(\lambda)$ is the von Neumann entropy of the state with $\lambda_1, \dots, \lambda_g$ on the diagonal, which is the same as the Shannon entropy of the probability distribution λ. If $\lambda_j = 0$, the state $\hat{\rho}_j$ never occurs and is thus not relevant for the result. Therefore, we may replace it by an arbitrary state. The quantity $I(\lambda;\hat{\rho})$ is sometimes called the *Holevo χ quantity*. It is an upper bound on the accessible information when sending classical information over a quantum channel [5, Theorem 12.1]. The proof of Theorem 6.2 relies again on the existence of typical subspaces.

As in Theorem 6.1, there is a converse to Theorem 6.2 which states that ϵ goes to 1 if m, g are chosen smaller [56, Theorem 8]. In particular, Theorem 6.2 implies that the POVM E can be approximated by a POVM with at most m outcomes instead of being simulated by a random mixture of such POVM. This has been proven in the earlier works [54, 55]; see [56, Theorem 6] for a precise statement. A different perspective on measurement compression as a communication protocol is given in [57].

6.3 Sparsification of POVMs

While the above results only hold in the asymptotic limit of infinitely many uses of the measurement, the recent results in [58] show that a sparsification of POVMs is also possible for a single use of the measurement. Let $(E_1, \dots, E_k)$ be a d-dimensional POVM, $d, k \in \mathbb{N}$. We define the *distinguishability (semi-)norm* induced by this POVM as

$$\|A\|_E := \sum_{i=1}^{k} |\mathrm{Tr}\,(AE_i)| \qquad \forall A \in \mathcal{M}_d^{\mathrm{herm}}.$$

This seminorm quantifies how well the POVM can be used for distinguishing quantum states [59, Section 12.1.1]. The trace norm $\|\cdot\|_1$ is the maximum of $\|\cdot\|_E$ over all such POVMs E [5, Theorem 9.1]. The results of [58] imply that for any POVM with many outcomes there is a POVM with few outcomes which distinguishes states almost equally well.

Theorem 6.3 ([58, Theorem 5.4])**.** *Let E be any d-dimensional POVM and $\epsilon \in (0,1)$, $d \in \mathbb{N}$. Then, there is a POVM E' with k outcomes, where $k \leq C\epsilon^{-2} d^2 \log d$ such that for all $A \in \mathcal{M}_d^{\mathrm{herm}}$*

$$(1-\epsilon)\|A\|_E \leq \|A\|_{E'}.$$

The proof uses methods from asymptotic geometric analysis, namely the approximation of zonoids by zonotopes.

6.4 Model compression

Hitherto, we have discussed compression problems which either compress states or measurements, but not both at the same time. The latter is done in the setting of *model compression.*

Assume we are given data $\{p_{u,i}^{(j)}\}_{(i,j,u)\in[k_j]\times[g]\times[v]}$, where g, v, $k_j \in \mathbb{N}$ for all $j \in [g]$. We know that this data comes from a quantum measurement, i.e. there are quantum states $\rho_u \in \mathcal{S}(\mathbb{C}^D)$ and D-dimensional POVMs $(E_1^{(j)}, \ldots, E_{k_j}^{(j)})$ such that

$$p_{u,i}^{(j)} = \mathrm{Tr}\left(\rho_u E_i^{(j)}\right) \qquad \forall (i,j,u) \in [k_j] \times [g] \times [v].$$

Let $\epsilon \in (0,1)$. The task is now to find a smaller quantum system of dimension $d < D$, states $\sigma_u \in \mathcal{S}(\mathbb{C}^d)$ and d-dimensional POVMs $(F_1^{(j)}, \ldots, F_{k_j}^{(j)})$ such that

$$\left|p_{u,i}^{(j)} - \mathrm{Tr}\left(\sigma_u F_i^{(j)}\right)\right| \leq \epsilon \qquad \forall (i,j,u) \in [k_j] \times [g] \times [v].$$

Let $k_1 = \ldots = k_g = k$ for simplicity. Assume that the effect operators $E_i^{(j)}$ have rank $\mathcal{O}(1)$ in D for all $j \in [g]$, $i \in [k-1]$, i.e. there is only one effect operator with large rank in each measurement. Theorem 6 of [60] shows that in this case, d can be chosen exponentially smaller than D. Note that in this problem it is not required that there are completely positive maps which send $\rho_u \mapsto \sigma_u$ and $E_i^{(j)} \mapsto F_i^{(j)}$ and indeed the construction used in [60] is non-linear. It relies on the conjugation of the states and effect operators with a suitable (complex) Johnson-Lindenstrauss transform (see Definition 5.1 for the real case). The non-linearity is necessary for the compressed effect operators to form a POVM [60, Section IV.A]. Lower bounds on the dimension d can be found in [60, Theorem 2] and [61, Theorem 1]. The latter reference uses lower bounds from *random access codes* which are used to send classical information over a quantum channel.

6.5 Shadow tomography

Another problem in which we try to learn an efficient description of a quantum state ρ for a fixed set of measurements is *shadow tomography*, which was first studied in [62]. Assume we are given an arbitrary unknown quantum state $\rho \in \mathcal{S}(\mathbb{C}^D)$, $D \in \mathbb{N}$, and a set of binary measurements given by $E_i \in \mathcal{E}(\mathbb{C}^D)$, $i \in [g]$, $g \in \mathbb{N}$. Let furthermore ϵ, $\delta \in (0,1)$. Our aim is to output numbers $q_1, \ldots, q_g \in [0,1]$ such that

$$|q_i - \mathrm{Tr}\,(E_i \rho)\,| \leq \epsilon \qquad \forall i \in [g]$$

with success probability at least $1-\delta$. To achieve this, we are allowed to use a measurement on $\rho^{\otimes k}$, where we want to choose k as small as possible as a function of D, g, ϵ and δ [62, Problem 1]. By performing each measurement on a separate copy, we could achieve $k = \tilde{\mathcal{O}}(g/\epsilon^2)$, where $\tilde{\mathcal{O}}$ hides polynomial factors in $\log g$, $\log 1/\epsilon$. It is also possible to choose $k = \mathcal{O}(D^2/\epsilon^2)$, since these are enough copies to perform full tomography of ρ, i.e. we can obtain the matrix entries of ρ up to a small error [63, 64]. However, a full knowledge of ρ is not needed since we are only interested in a fixed set of measurements. Theorem 2 of [62] shows that it is possible to choose $k = \tilde{\mathcal{O}}(\log(1/\delta)\log(g)^4\log(D)/\epsilon^5)$, where the $\tilde{\mathcal{O}}$ notation hides polynomial factors in $\log(1/\epsilon)$, $\log\log D$ and $\log\log g$. The proof is based on an application of *postselected learning*. The procedure can be interpreted as a compression of ρ in the sense that we do not need the full information about ρ which would require a full tomography, but we can achieve the task with much less information. In the same spirit, the authors of [65] compress a classical description of a pure quantum state such that the compression allows to compute expectation values of observables with high probability.

6.6 Quantum compression relative to a set of measurements

In Core Article I [1], we consider yet another notion of compression. We consider a situation in which we want to perform at some later point a set of g measurements, $g \in \mathbb{N}$. The measurements are given as D-dimensional POVMs $E^{(j)} := (E_1^{(j)}, \ldots, E_{k_j}^{(j)})$, where D, $k_j \in \mathbb{N}$ for all $j \in [g]$. Let

$$\mathcal{O} := \left\{ E_i^{(j)} : i \in [k_j], j \in [g] \right\}$$

be the set of corresponding effect operators. Furthermore, we are given a single copy of an unknown quantum state $\rho \in \mathcal{S}(\mathbb{C}^D)$. Until we perform the measurements in $\mathcal{O}$, we need to store this state in a quantum memory. In order to use a minimum of storage space, we want to keep only the information in the state relevant for the measurements we are interested in. We are allowed to use in addition an unlimited amount of classical side information because classical information is easy to store compared to quantum information. Thus, we are interested in the compression dimension of $\mathcal{O}$ defined as follows:

Definition 6.4 ([1, Definition 4.1]). *Let $\mathcal{O} \subseteq \mathcal{M}_D^{\mathrm{herm}}$, $D \in \mathbb{N}$. The* compression dimension *of $\mathcal{O}$ is the smallest $d \in \mathbb{N}$ for which there is an $n \in \mathbb{N}$ and CPTP maps $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$, $\mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$ such that for their composition $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$, it holds that*

$$\mathrm{Tr}\,(\rho E) = \mathrm{Tr}\,(\mathcal{T}(\rho) E) \qquad \forall \rho \in \mathcal{S}(\mathbb{C}^D),\ \forall E \in \mathcal{O}. \tag{6.3}$$

If the compression dimension equals D, we refer to $\mathcal{O}$ as incompressible*.*

In contrast to quantum source coding (Section 6.1), measurement compression (Section 6.2) and shadow tomography (Section 6.5), we are given here only a single copy of an unknown quantum state. However, the setting of shadow tomography is similar in the sense that our compression procedure only has to work for a fixed set of measurements. The sparsification of a POVM reviewed in Section 6.3 also acts on a single copy, but there the aim is to reduce the number of outcomes of the measurement. In contrast to that, our aim is to reduce the dimension of the quantum system. The setting of model compression (Section 6.4) is closest to the situation considered here. The main difference is that we require the existence of both a compression and decompression channel, because our aim is to find a physical transformation implementing the compression procedure. In model compression, the lower-dimensional states and effect operators need not be connected to the original versions even by a linear map, because one wants to find a smaller model which explains the measurement outcomes. Since we allow for an arbitrary amount of classical side information, the lower bounds from model compression do not give lower bounds on the compression dimension in Definition 6.4. Therefore, our notion of compression is different from the other versions reviewed in this chapter.

In Sections 6.1 to 6.5, we always allow for a small error ϵ. Therefore, one might argue that the requirement in Equation (6.3) should also allow for a small error ϵ in the measurement statistics. We show in Theorem 5.1 of Core Article I [1] that for each $\mathcal{O}$, there is an $\epsilon > 0$ such that allowing for an error ϵ does not decrease the compression dimension. In this sense, the compression dimension is thus stable. The key insight to prove this is that each composite channel $\mathcal{T}$ with quantum dimension d can be realized using only $4 \log_2 D$ classical bits [1, Lemma 5.2].

Using the result by Arveson from Theorem 3.10, it follows immediately that an unstructured measurement with more than two elements is incompressible, since two generic operators generate the full algebra $\mathcal{M}_D$. The operator-algebraic approach taken in Core Article I [1] generalizes the proof of Theorem 3.10 to show that the compression dimension is determined by the structure of the matrix algebra $C^*(\mathcal{O})$ generated by $\mathcal{O}$. This algebra has the form given in Theorem 3.2. Let $m \in \mathbb{N}$ and let $D_i \in \mathbb{N}$, $i \in [m]$ be the dimensions of the full matrix algebras in this representation. Then, Theorems 6.1 and 7.1 of Core Article I [1] show that the compression dimension d is bounded by

$$\min_{i \in [m]} \{D_i\} \leq d \leq \max_{i \in [m]} \{D_i\}.$$

This shows that the presence of symmetries in the measurements in $\mathcal{O}$ allows for compression even if we only have access to a single copy of an unknown quantum state. In particular, we demonstrate in Section 7.2 of Core Article I [1] that two binary projective measurements can be compressed to qubit size ($d = 2$). In Section 8 of Core Article I [1], we show that $d = D_i$ for some $i \in [m]$ and provide an algorithm based on an SDP to compute the compression dimension.

This operator-algebraic approach is complemented by an independent approach based on Bézout's theorem from classical algebraic geometry and irreducible polynomials in Section 6.2 of Core Article I [1]. While the lower bounds derived there are generally weaker, they translate to the setting in which we allow the compression map to act on $\rho^{\otimes k}$ for some $k \in \mathbb{N}$ ([1, Theorem 10.2]). Note that we still require $d < D$ instead of $d/k < D$, as in source coding or measurement compression.

In this chapter, we have considered different notions of compression in quantum information theory and have explained how Core Article I [1] differs from other works. Now we turn from restrictions on the compression of quantum systems to restrictions on the amount of incompatibility in quantum measurements, studied in Core Article II [2] and Article IV [4]. In the next chapter, we consider the main tool used in these articles: free spectrahedra.

7 Free spectrahedra

This chapter gives a very brief introduction to the theory of free spectrahedra. Free spectrahedra were introduced in the field of convex optimization in order to find efficiently computable relaxations of computationally hard problems. Moreover, they have an application to quantum information theory. Thereby, this chapter introduces the necessary background for Chapter 8, where we discuss the connection between free spectrahedra and the incompatibility of quantum measurements we establish in Core Article II [2].

Let us start the introduction with the following motivational example:

Example 7.1 (Matrix cube problem [66])**.** *Let* g*,* $d \in \mathbb{N}$ *and* $E \in (\mathcal{M}_d^{\mathrm{herm}})^g$*. We define*

$$\mathcal{X}_E := \left\{ x \in \mathbb{R}^g : \sum_{i=1}^{g} x_i E_i \leq \mathbb{1}_d \right\}.$$

Then, the matrix cube problem *is to determine whether* $[-1,1]^g \subseteq \mathcal{X}_E$*.*

This problem has important applications to stability problems in engineering [66] and optimization theory (see [67, Section 1.5] for a selection of examples). It is of course possible to check the inclusion on the extreme points of $[-1,1]^g$, i.e. the vertices of the hypercube. However, since there are 2^g vertices, this is not computationally efficient. Unfortunately, the matrix cube problem is NP-hard in general, since it contains the maximization of a positive definite quadratic form as a special case [66, Section 4], which in turn contains the max-cut problem from graph theory. The latter is well-known to be NP-hard [68]. However, it is possible to find a relaxation of the above problem in terms of free spectrahedra, as we will show shortly.

Definition 7.2 (Free spectrahedron)**.** *Let* g*,* $d \in \mathbb{N}$ *and* $A \in (\mathcal{M}_d^{\mathrm{herm}})^g$*. Then,*

$$\mathcal{D}_A(n) := \left\{ X \in (\mathcal{M}_n^{\mathrm{herm}})^g : \sum_{i=1}^{g} A_i \otimes X_i \leq \mathbb{1}_d \right\}.$$

is called the free spectrahedron at level n *defined by* A*. Furthermore,*

$$\mathcal{D}_A := \bigcup_{n \in \mathbb{N}} \mathcal{D}_A(n)$$

is the free spectrahedron *defined by* A*.*

Some articles such as [67, 69] define free spectrahedra to be real, i.e. such that A, X are real symmetric in the above definition. Many results on *real free spectrahedra* can however be recovered in our setting with only minor modifications. We observe that $\mathcal{X}_E$ as in Example 7.1 is actually $\mathcal{D}_E(1)$. This is the *spectrahedron* defined by E. Before we proceed, let us consider two important examples of free spectrahedra which come from unit balls of ℓ_p-norms.

Example 7.3 (Matrix diamond [70])**.** *Let g, $n \in \mathbb{N}$. The* matrix diamond *at level n is defined as*

$$\mathcal{D}_{\diamond,g}(n) := \left\{ X \in (\mathcal{M}_n^{\text{herm}})^g : \sum_{i=1}^{g} D_i \otimes X_i \leq \mathbb{1}_{2^g n} \right\}$$
$$= \left\{ X \in (\mathcal{M}_n^{\text{herm}})^g : \sum_{i=1}^{g} \epsilon_i X_i \leq \mathbb{1}_n \ \forall \epsilon \in \{\pm 1\}^g \right\}.$$

Here, $D_i := \mathbb{1}_2^{\otimes(i-1)} \otimes \operatorname{diag}(+1,-1) \otimes \mathbb{1}_2^{\otimes(g-i)}$ for all $i \in [g]$. In particular, $\mathcal{D}_{\diamond,g}(1)$ is the unit ball of the ℓ_1-norm in $\mathbb{R}^g$, which is diamond-shaped for $g = 2$.

Example 7.4 (Matrix cube [67])**.** *Let g, $n \in \mathbb{N}$. The* matrix cube *at level n is defined as*

$$\mathcal{D}_{\square,g}(n) := \left\{ X \in (\mathcal{M}_n^{\text{herm}})^g : \sum_{i=1}^{g} C_i \otimes X_i \leq \mathbb{1}_{2gn} \right\}$$
$$= \{ X \in (\mathcal{M}_n^{\text{herm}})^g : \|X_i\|_\infty \leq 1 \ \forall i \in [g] \}.$$

Here, $C_i = |i\rangle\langle i| \oplus (-|i\rangle\langle i|)$ for all $i \in [g]$, where $\{|i\rangle\}_{i \in [g]}$ is the standard basis of $\mathbb{C}^g$. In particular, $\mathcal{D}_{\square,g}(1) = [-1,1]^g$ is the unit ball of the ℓ_∞-norm in $\mathbb{R}^g$, which is a hypercube.

Example 7.4 shows that $[-1,1]^g$ is a spectrahedron and that Example 7.1 can be rewritten as the spectrahedral inclusion problem to determine whether $\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_E(1)$. To find a relaxation of this inclusion problem, we consider the inclusion of free spectrahedra. Let $A \in (\mathcal{M}_d^{\text{herm}})^g$ and $B \in (\mathcal{M}_D^{\text{herm}})^g$ for g, d, $D \in \mathbb{N}$. We say that $\mathcal{D}_A \subseteq \mathcal{D}_B$ if $\mathcal{D}_A(n) \subseteq \mathcal{D}_B(n)$ for all $n \in \mathbb{N}$. Therefore, the free spectrahedral inclusion $\mathcal{D}_{\square,g} \subseteq \mathcal{D}_E$ implies in particular the spectrahedral inclusion $\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_E(1)$. The next proposition shows that the inclusion on the level of free spectrahedra can be verified efficiently.

Proposition 7.5 ([69, Theorem 3.5])**.** *Let $A \in (\mathcal{M}_d^{\text{herm}})^g$ and $B \in (\mathcal{M}_D^{\text{herm}})^g$ for g, d, $D \in \mathbb{N}$. Moreover, let $\mathcal{D}_A(1)$ be bounded. We consider the unital linear map*

$$\Phi : \operatorname{span}\{\mathbb{1}_d, A_1, \ldots, A_g\} \to \mathcal{M}_D$$
$$A_i \mapsto B_i \qquad \forall i \in [g].$$

Then, it holds that Φ is n-positive if and only if $\mathcal{D}_A(n) \subseteq \mathcal{D}_B(n)$, where $n \in \mathbb{N}$. In particular, $\mathcal{D}_A \subseteq \mathcal{D}_B$ if and only if Φ is completely positive.

We emphasize that Φ is n-positive only on the operator system generated by A, not on $\mathcal{M}_d$. Theorem 3.5 implies the existence of a completely positive extension of Φ to $\mathcal{M}_d$ only if Φ is completely positive. Proposition 7.5 has been proven in [69] for real free spectrahedra, but the proof holds without modification in the complex setting (see [2, Lemma IV.4]). From Proposition 7.5, it is easy to see that verifying $\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1)$ is hard in general, whereas $\mathcal{D}_A \subseteq \mathcal{D}_B$ can be checked efficiently. This is true because determining whether $\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1)$ is equivalent to determining whether Φ is positive, under the assumption that $\mathcal{D}_A(1)$ is bounded. Verifying whether a map is positive is known to be NP-hard [71]. On the other hand, under the same assumption $\mathcal{D}_A \subseteq \mathcal{D}_B$ is equivalent to Φ being completely positive. As we have seen in Chapter 4, this can be decided using the SDP in Theorem 4.5. Therefore, verifying $\mathcal{D}_{\square,g} \subseteq \mathcal{D}_E$ is a suitable relaxation of the matrix cube problem.

To use this relaxation in practice, we need to know which free spectrahedral inclusion problem we need to consider if we want to conclude that the spectrahedral inclusion $\mathcal{D}_{\square,g}(1) \subset \mathcal{D}_A(1)$ does not hold. We note that in general,

$$\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_A(1) \not\Rightarrow \mathcal{D}_{\square,g} \subseteq \mathcal{D}_A.$$

However, a similar implication holds if we shrink $\mathcal{D}_{\square,g}$, i.e.

$$\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_A(1) \implies s \cdot \mathcal{D}_{\square,g} \subseteq \mathcal{D}_A. \tag{7.1}$$

for some appropriately chosen $s \in [0,1]^g$. Here, $s \cdot \mathcal{D}_{\square,g} = \{(s_1 X_1, \ldots, s_g X_g) : X \in \mathcal{D}_{\square,g}\}$. We would like to find such s that work for a large class of A in Equation (7.1). This leads us to the definition of inclusion sets.

Definition 7.6 ([2, Definition IV.1])**.** *Let k, d, $g \in \mathbb{N}$ and $A \in \left(\mathcal{M}_d^{\mathrm{herm}}\right)^g$. The* inclusion set $\Delta_{\mathcal{D}_A}(k)$ *is defined as*

$$\Delta_{\mathcal{D}_A}(k) := \left\{ s \in \mathbb{R}_+^g : \forall B \in \left(\mathcal{M}_k^{\mathrm{herm}}\right)^g \ \mathcal{D}_A(1) \subseteq \mathcal{D}_B(1) \Rightarrow s \cdot \mathcal{D}_A \subseteq \mathcal{D}_B \right\}.$$

Here, the set $s \cdot \mathcal{D}_A := \{ (s_1 X_1, \ldots, s_g X_g) : X \in \mathcal{D}_A \}$ is the (asymmetrically) scaled *free spectrahedron.*

Often, such sets are only studied for vectors for which $s_1 = \ldots = s_g$ [67, 70, 72], but the above definition also allows for asymmetric scaling. Coming back to Example 7.1, let $s_0(1, \ldots, 1) \in \Delta_{\mathcal{D}_{\square,g}}(d)$. Then, $\mathcal{D}_{\square,g} \subseteq \mathcal{D}_E$ implies that $\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_E(1)$, by definition. If, however, $s_0(1, \ldots, 1) \cdot \mathcal{D}_{\square,g} \not\subseteq \mathcal{D}_E$, we can conclude that $\mathcal{D}_{\square,g}(1) \not\subseteq \mathcal{D}_E(1)$. In this sense, the error we make by considering the relaxation is not too large if s_0 is close to 1. Such optimal s_0 for the matrix cube have been found for real symmetric spectrahedra in [66, 67].

Now that we have seen that free spectrahedra provide useful relaxations for spectrahedral inclusion problems, one might wonder how to choose the best free spectrahedron corresponding to a given spectrahedron. Let $\mathcal{C} \subseteq \mathbb{R}^g$ be a convex set. In general, there is no unique free spectrahedron $\mathcal{D}_A$ with $\mathcal{D}_A(1) = \mathcal{C}$. It has been shown in [72, Theorem 4.1] that $\mathcal{C}$ admits a unique free spectrahedron if and only if $\mathcal{C}$ is a simplex containing 0 in its interior. For special $\mathcal{C}$, there exists a maximal free spectrahedron.

Definition 7.7 (Maximal free spectrahedron [70, Definition 4.1])**.** *Let g, $n \in \mathbb{N}$ and let $\mathcal{C} \subset \mathbb{R}^g$ be a polyhedron with 0 as an interior point. Then, the free spectrahedron defined by*

$$\mathcal{W}_{\max}(\mathcal{C})(n) := \left\{ X \in (\mathcal{M}_n^{\mathrm{herm}})^g : \sum_{i=1}^g c_i X_i \leq \alpha \mathbb{1}_n, \quad \forall c \in \mathbb{R}^g, \forall \alpha \in \mathbb{R} \ \text{ s.t. } \ \mathcal{C} \subseteq \{ x \in \mathbb{R}^g : \langle c, x \rangle \leq \alpha \} \right\}.$$

at level n is the maximal spectrahedron *corresponding to $\mathcal{C}$.*

We note that indeed $\mathcal{W}_{\max}(\mathcal{C})(1) = \mathcal{C}$, as claimed above. The $\mathcal{W}_{\max}(\mathcal{C})$ is maximal in the sense that for any $A \in (\mathcal{M}_d^{\mathrm{herm}})^g$ of arbitrary dimension $d \in \mathbb{N}$ for which $\mathcal{D}_A(1) = \mathcal{C}$, it holds that $\mathcal{D}_A \subseteq \mathcal{W}_{\max}(\mathcal{C})$ [70, Proposition 4.3]. The matrix diamond and the matrix cube are the maximal free spectrahedra for the unit balls in the ℓ_∞- and ℓ_1-norm, respectively. This can easily be seen from their definition. Although it is straightforward to construct the maximal spectrahedron for a given polyhedron, it yields the most conservative estimate in the relaxation procedure we considered (see [69, Section 5.2] for a discussion of this).

Apart from the matrix cube, bounds on the inclusion sets are only known in a few cases. The works [67, 70] consider symmetric free spectrahedra (see also [2, Section VII.A]), whereas [72] studies the inclusion sets for free spectrahedra coming from unit balls of ℓ_p-norms such as the matrix diamond. The bounds from the latter work are most useful to us in Core Article II [2], therefore we state them here. Let $g \in \mathbb{N}$ and let

$$\mathrm{QC_g} := \{s \in [0,1]^g : s_1^2 + \ldots + s_g^2 \leq 1\} \tag{7.2}$$

be the positive orthant of the Euclidean unit ball. Then [72, Theorem 6.6] shows that

$$\mathrm{QC}_g \subseteq \Delta_{\mathcal{D}_{\diamond,g}}(d) \tag{7.3}$$

for all $d \in \mathbb{N}$ (see also [2, Theorem VII.7] for a simpler proof). A close inspection of [72, Theorem 6.6] also allows us to show that

$$\Delta_{\mathcal{D}_{\diamond,g}}(d) \subseteq \mathrm{QC}_g \qquad \forall d \geq 2^{\lceil (g-1)/2 \rceil}, \tag{7.4}$$

where $\lceil \cdot \rceil : \mathbb{R} \to \mathbb{N}$ is the ceiling function.

In Core Article II [2] and Article IV [4], we show that there is a connection between the compatibility of quantum measurements and a certain free spectrahedral inclusion problem. This connection can be used to prove new bounds on the inclusion set of the matrix diamond and its generalization, the matrix jewel, defined in [4, Section 4]. The bounds from quantum information which we use are based on the study of approximate cloning, *mutually unbiased bases* (MUBs) and *symmetric informationally complete POVMs* (SIC-POVMs). These results are reviewed in Sections VI and VIII.A of Core Article II [2] and Section 6 of Article IV [4]. Their implications for the inclusion sets are discussed in Section IX.B of [2] and Section 10 of [4]. We discuss the connection to quantum incompatibility in more detail in the next chapter.

8 Compatibility of quantum measurements

In this chapter, we return to the study of measurements in quantum mechanics. We focus on an aspect which distinguishes quantum theory from classical physics, namely the existence of incompatible measurements. These correspond to observables which cannot be measured simultaneously, with position and momentum of a particle as a famous example [73, 74]. At the end of this chapter, we discuss the connection between the compatibility of quantum measurements and free spectrahedra, which were the topic of the previous chapter.

In Section 8.1, we define what it means for quantum measurements to be compatible. This leads us to the study of the compatibility regions introduced in Section 8.2, which provide a detailed picture of the amount of incompatibility available. An established way to find lower bounds on these compatibility regions is to use results from approximate quantum cloning. We review this approach in Section 8.3. Section 8.4 finally discusses the new approach from Core Article II [2] and Article IV [4] to obtain bounds on the compatibility regions which uses results from the study of free spectrahedra.

8.1 Compatible quantum measurements

As in Chapter 2, we identify measurements with POVMs. A collection of POVMs is compatible if they arise as marginals from a joint POVM (see [75] for an introduction to the topic).

Definition 8.1 (Compatible measurements)**.** *Let g, $d \in \mathbb{N}$. Let moreover $(E_1^{(i)}, \ldots, E_{k_i}^{(i)})$, $i \in [g]$, be a collection of d-dimensional POVMs, where $k_i \in \mathbb{N}$ for all $i \in [g]$. These POVMs are* jointly measurable *or* compatible *if there is another d-dimensional joint POVM $\{ R_{\mathbf{j}} \}$ with $\mathbf{j} \in [k_1] \times \ldots \times [k_g]$ such that for all $u \in [g]$ and $v \in [m_u]$,*

$$E_v^{(u)} = \sum_{\substack{\mathbf{j} \in [k_1] \times \ldots \times [k_g] \\ j_u = v}} R_{\mathbf{j}}.$$

We can alternatively define compatible measurements as those measurements which arise from a joint measurement by classical post-processing. Both definitions are equivalent [75, Section 3.1].

Proposition 8.2. *Let d, g, $k_i \in \mathbb{N}$ and $E^{(i)} \in (\mathcal{M}_d^{\mathrm{herm}})^{k_i}$, $i \in [g]$, be a collection of POVMs. These POVMs are compatible if and only if there is some $m \in \mathbb{N}$ and a POVM $M \in (\mathcal{M}_d^{\mathrm{herm}})^m$ such that*

$$E_v^{(u)} = \sum_{x=1}^{m} p_u(v|x) M_x$$

for all $v \in [k_u]$, $u \in [g]$ and some conditional probabilities $p_u(v|x)$.

For fixed u, each set $\{p_u(v|x)\}_{v \in [k_u], x \in [m]}$ defines a classical channel which processes the data collected from the measurement of M such that we obtain the measurement statistics for $E^{(u)}$. Since classical information can be duplicated, this allows us to measure all $E^{(u)}$, $u \in [g]$, at the same time. In this sense, the proposition gives an operational interpretation of the notion of compatibility.

It is known that g POVMs $E^{(i)}$, $i \in [g]$, are jointly measurable if their effect operators $E_j^{(i)}$ commute pairwise [76, Example 2.7]. If $g = 2$ and one of the measurements is a projective measurement, i.e. all effect operators of this measurement are orthogonal projections, then pairwise commutation is necessary for the two measurements to be compatible [76, Corollary 4.2]. In general, however, the condition that the measurements commute pairwise is only sufficient, not necessary, for joint measurability. We refer the reader to [75, Section 3.2] and the references therein for details.

For concrete POVMs, it is possible to check whether they are compatible using an SDP. We illustrate this for two binary measurements defined by $E, F \in \mathcal{E}(\mathbb{C}^d)$, $d \in \mathbb{N}$. The SDP is:

$$\begin{aligned} \text{minimize} \quad & \operatorname{Tr}(X) \\ \text{subject to} \quad & E \geq X \\ & F \geq X \\ & X \geq E + F - \mathbb{1}_d \\ & X \geq 0, \ X \in \mathcal{M}_d^{\text{herm}}. \end{aligned}$$

Using a basis of the Hermitian operators and using a direct sum to write all inequalities as one, this can again be brought into the dual form of an SDP (see Definition 4.3). The quantity we minimize is not important, since we are only interested in whether this problem is feasible. It has been shown in Proposition 1 of [77] that this problem is feasible if and only if $(E, \mathbb{1}_d - E)$ and $(F, \mathbb{1}_d - F)$ are jointly measurable. We refer the reader to [77] for the SDPs for an arbitrary number of outcomes and an arbitrary number of POVMs.

8.2 Compatibility regions

From entanglement theory, we know that noise can destroy the quantum properties of a system. The same is true for incompatibility. While we have seen that there are incompatible quantum measurements such as non-commuting projective measurements, a sufficient amount of noise makes any collection of POVMs compatible. A natural type of noise to consider is white noise. In this case, instead of measuring the d-dimensional POVM $E_1, \ldots, E_k$, one measures the noisy POVM with elements

$$E_i' := sE_i + (1-s)\mathbb{1}_d/k, \qquad i \in [k] \tag{8.1}$$

for some $s \in [0, 1]$. This means the new measurement $E_1', \ldots, E_k'$ corresponds to a device which performs the original measurement with probability s and with probability $1 - s$ outputs an outcome uniformly at random, independently of the state of the quantum system. The k-outcome measurement $(\mathbb{1}_d/k, \ldots, \mathbb{1}_d/k)$ is a special instance of an observable which has the same measurement statistics for every quantum state. Such observables are called *trivial observables*. It is easy to see that any trivial observable is of the form $(a_1\mathbb{1}, \ldots, a_k\mathbb{1})$ with a probability distribution $a := (a_1, \ldots, a_k)$. Different choices for the probability distribution a correspond to different noise models. Apart from the uniform distribution, another interesting choice is $a_i = \operatorname{Tr}(E_i)/d$ which depends on the POVM at hand. The advantage of this noise model is that the map

$$E_i \mapsto sE_i + (1-s)\frac{\operatorname{Tr}(E_i)}{d}\mathbb{1}_d \tag{8.2}$$

is linear for fixed $s \in [0,1]$. This type of noise has been studied in the context of *quantum steering* in order to connect the steering of noisy quantum states to the noise robustness of incompatibile quantum measurements [78].

Another parallel between incompatibility and entanglement is that both are a resource for quantum information processing tasks. In the case of incompatibility, it has been shown in [79] that only incompatible observables can violate Bell inequalities. These violations are important e.g. for quantum cryptography [80]. The resource entanglement can be quantified e.g. using entanglement of formation or the Schmidt number [17, Section XV]. In the same spirit, the authors of [81] define *incompatibility monotones* to quantify the amount of incompatibility in two binary d-dimensional measurements E and F. One approach is to consider the minimal amount of noise needed to make the measurements compatible. Proposition 2 of [81] shows that for the noise models in Equations (8.1) and (8.2), this idea indeed gives rise to the incompatibility monotones:

$$
\begin{aligned}
I_1(E,F) &:= \inf\{(1-s) : s \in [0,1],\ sE + (1-s)\mathbb{1}_d/2,\ sF + (1-s)\mathbb{1}_d/2 \text{ compatible}\},\\
I_2(E,F) &:= \inf\left\{(1-s) : s \in [0,1],\ sE + (1-s)\frac{\operatorname{Tr}(E)}{d}\mathbb{1}_d,\ sF + (1-s)\frac{\operatorname{Tr}(F)}{d}\mathbb{1}_d \text{ comp.}\right\}
\end{aligned}
$$

for $E,\ F \in \mathcal{E}(\mathbb{C}^d)$.

Since incompatibility is a resource for quantum information processing tasks which can be quantified by noise robustness, it makes sense to study in more detail how the incompatibility available depends on the number of measurements, their dimension and the number of outcomes. In particular, we would like to study the amount of incompatibility in g measurements in dimension d with outcomes $\mathbf{k} = (k_1, \ldots, k_g) \in \mathbb{N}^g$, where k_i is the number of outcomes of the i-th measurement for all $i \in [g]$. This question motivates the following definition:

Definition 8.3 ([4, Definition 3.20])**.** *Let* $\mathbf{k} \in \mathbb{N}^g$, d, $g \in \mathbb{N}$. *Then, we call*

$$\Gamma(g,d,\mathbf{k}) := \left\{ s \in [0,1]^g : s_j E^{(j)} + (1-s_j)\mathbb{1}_d/k_j \text{ compatible } \forall \text{ POVMs } E^{(j)} \in (M_d^{\text{herm}})^{k_j} \right\}$$

the balanced compatibility region *for g POVMs in d dimensions with k_j outcomes, $j \in [g]$.*

The maximum $s \in [0,1]$ such that $s(1,\ldots,1) \in \Gamma(g,d,\mathbf{k})$ would therefore give the minimal amount of incompatibility in g measurements with $\mathbf{k}$ outcomes in dimension d in the sense of [81]. The set $\Gamma(g,d,\mathbf{k})$ gives a more detailed picture of the situation in the sense that we also allow for asymmetric noise, i.e. different amounts of noise for each measurement. A similar compatibility region can be defined for the linear noise model in Equation (8.2).

The set $\Gamma(g,d,\mathbf{k})$ is easily seen to be convex [2, Proposition III.2]. Therefore, it always holds that

$$\{s \in [0,1]^g : s_1 + \ldots + s_g \le 1\} \subseteq \Gamma(g,d,\mathbf{k}) \qquad \forall (g,d,\mathbf{k}) \in \mathbb{N}^{g+2}.$$

Very similar sets were studied in [82]. There, the authors show that for QC_g as defined in Equation (7.2), it holds that $\mathrm{QC}_2 = \Gamma(2,d,2^{\times 2})$. For their proof, they use the fact that two binary measurements on a system A are incompatible if and only if we can find two measurements on a system B and a quantum state on the joint system AB such that the CHSH inequality (Equation (2.2)) is violated [77]. This means that the left hand side of Equation (2.2) has a value larger than 2.

For more than two obervables or observables with more outcomes, the one-to-one correspondence between incompatible measurements and measurements allowing for Bell inequality violations is no longer valid [83, 84]. Therefore, the proof technique used in [82] cannot be generalized. For low dimensions and few measurements, it is however possible to find upper bounds through direct computation. This was done for three measurments on a two-dimensional system in [85, 86] to show that $\Gamma(3,2,2^{\times 3}) \subseteq \mathrm{QC}_3$.

8.3 Approximate quantum cloning

In order to find lower bounds on the balanced compatibility region defined in Definition 8.3, it is possible to use results on approximate quantum cloning [87]. For simplicity, we explain the idea for two measurements. If it was possible to perfectly clone any quantum state, i.e. to find a CPTP map $\Phi : \mathcal{M}_d \to \mathcal{M}_d^{\otimes 2}$ which sends $\rho \mapsto \rho \otimes \rho$ for any $\rho \in \mathcal{S}(\mathbb{C}^d)$ and fixed $d \in \mathbb{N}$, then any collection of measurements would be compatible. To perform different measurements on the same quantum state, we could copy the state first and perform one measurement on each copy. Going to the dual picture, this would correspond to a map $\Phi^* : \mathcal{M}_d^{\otimes 2} \to \mathcal{M}_d$ such that for any number of outcomes $k_1,\ k_2 \in \mathbb{N}$,

$$R_{j_1,j_2} := \Phi^*(E^{(1)}_{j_1} \otimes E^{(2)}_{j_2}) \qquad \forall j_1 \in [k_1],\ \forall j_2 \in [k_2],$$

would be a joint measurement for any d-dimensional measurements $\{E^{(i)}_{j_i}\}_{j_i \in [k_i]}$, $i \in \{1,2\}$. This is true, since for any $u \in [k_1]$,

$$\mathrm{Tr}\left(\rho \sum_{j_2=1}^{k_2} R_{u,j_2}\right) = \mathrm{Tr}\left(\rho \Phi^*(E^{(1)}_u \otimes \mathbb{1})\right) = \mathrm{Tr}\left((\rho\otimes\rho)(E^{(1)}_u \otimes \mathbb{1})\right) = \mathrm{Tr}\left(\rho E^{(1)}_u\right) \quad \forall \rho \in \mathcal{S}(\mathbb{C}^d).$$

It is known by the *no-cloning theorem* that no perfect cloning devices for quantum states exist [7, Example 4.17], but it is possible to clone quantum states approximately.

Theorem 8.4 ([9, Theorem 7.1])**.** *Let $d,\ g \in \mathbb{N}$. For a quantum channel $\Psi : \mathcal{M}_d \to \mathcal{M}_d^{\otimes g}$, consider the quantities*

$$\mathcal{F}_{c,1}(\Psi) := \inf_{j\in[g]} \inf_{\sigma \text{ pure}} \mathrm{Tr}\left(\sigma^{(j)}\Psi(\sigma)\right),$$

where $\sigma^{(j)} = \mathbb{1}_d^{\otimes(j-1)} \otimes \sigma \otimes \mathbb{1}_d^{\otimes(g-j)} \in \mathcal{M}_d^{\otimes g}$ for all $j \in [g]$, and

$$\mathcal{F}_{c,\mathrm{all}}(\Psi) := \inf_{\sigma \text{ pure}} \mathrm{Tr}\left(\sigma^{\otimes g}\Psi(\sigma)\right).$$

These quantities are both maximized by the optimal quantum cloner

$$\Psi_{\mathrm{opt}}(\rho) := \frac{d}{d[g]} S_g(\rho \otimes \mathbb{1}) S_g. \tag{8.3}$$

Here, $d[g] = \binom{d+g-1}{g}$ is the dimension of the symmetric subspace $\vee^g\mathbb{C}^d \subseteq (\mathbb{C}^d)^{\otimes g}$ and S_g is the corresponding orthogonal projection.

The above theorem settles the question for the case that all copies of σ are required to have the same quality. Maximizing $\mathcal{F}_{c,1}(\Psi)$ means that each marginal of $\Psi(\sigma)$ has to be close to σ, whereas maximizing $\mathcal{F}_{c,\mathrm{all}}(\Psi)$ means that $\Psi(\sigma)$ has to be close to the product $\sigma^{\otimes g}$. It turns out that the map which optimizes these quantities is the same in both cases and is explicitly given by Equation (8.3). This map can again be used to construct a measurement

$$R_{j_1,j_2} := \Psi^*_{\mathrm{opt}}(E^{(1)}_{j_1} \otimes E^{(2)}_{j_2}), \qquad \forall j_1 \in [k_1],\ \forall j_2 \in [k_2],$$

which is a joint POVM for the measurements

$$\left\{ \gamma E_{j_1}^{(i)} + (1-\gamma)\frac{\operatorname{Tr}\left(E_{j_i}^{(i)}\right)}{d}\mathbb{1}_d \right\}_{j_i \in [k_i]}, \qquad i \in \{1,2\},$$

where $\gamma = (g+d)/(g(1+d))$. This can be seen from the same reasoning as above, together with the reduced states of $\Psi_{\mathrm{opt}}(\sigma)$ on each system given in [88, Equation (7)]. Doubling the dimension allows us to go from the noise model in Equation (8.2) to the one in Equation (8.1) (see [2, Proposition III.4]) such that

$$\frac{g+2d}{g(1+2d)}(1,\dots,1) \in \Gamma(g,d,\mathbf{k})$$

for all $\mathbf{k} = (k_1, \dots, k_g)$. This idea can also be generalized to asymmetric noise using results on asymmetric approximate cloning such as [89, 90, 91] to obtain lower bounds in the case where we have different levels of noise in each measurement (see [2, Section VI] and [4, Section 6.2]).

8.4 The connection to free spectrahedra

In Core Article II [2] and Article IV [4], we give a new way to prove bounds on $\Gamma(g,d,\mathbf{k})$ using a connection to the inclusion of free spectrahedra. For simplicity, we state our results only for binary measurements and refer the reader to Article IV [4] for measurements with an arbitrary number of outcomes. Our first main result is that the compatibility of binary measurements is equivalent to the inclusion of the matrix diamond into a free spectrahedron defined by the effect operators.

Theorem 8.5 ([2, Theorem V.3]). *Let $E \in \left(\mathcal{M}_d^{\mathrm{herm}}\right)^g$ and let $2E-\mathbb{1} := (2E_i-\mathbb{1}_d, \dots, 2E_g-\mathbb{1}_d)$. We have*

1. *$\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_{2E-\mathbb{1}}(1)$ if and only if $E_1, \dots, E_g$ are effect operators.*
2. *$\mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{2E-\mathbb{1}}$ if and only if $E_1, \dots, E_g$ are compatible effect operators.*
3. *$\mathcal{D}_{\diamond,g}(k) \subseteq \mathcal{D}_{2E-\mathbb{1}}(k)$ for $k \in [d]$ if and only if for any isometry $V : \mathbb{C}^k \hookrightarrow \mathbb{C}^d$, the induced compressions $V^*E_1V, \dots, V^*E_gV$ are compatible effect operators.*

This theorem shows that the different levels of the free spectrahedral inclusion problem correspond to different degrees of compatibility. The proof of the second point of Theorem 8.5 is based on the extension of completely positive maps. We include a rough sketch of the proof here: We have seen that the inclusion holds if and only if the map Φ defined in Proposition 7.5 is completely positive. By Theorem 3.5, the map Φ is completely positive if and only if there is a completely positive extension Ψ onto $\mathbb{C}^{2^g}$, since the matrices D_i defining the matrix diamond in Example 7.3 are diagonal. By Theorem 3.6, this is equivalent to the existence of a positive extension Ψ. We can then verify that such a Ψ specifies a joint POVM for $E_1, \dots, E_g$. For measurements with an arbitrary number of outcomes, the result corresponding to Theorem 8.5 is Theorem 5.2 of Article IV [4], where we have to replace the matrix diamond by the matrix jewel defined in [4, Section 4].

Our second main result shows that the inclusion set for the matrix diamond coincides with the balanced compatibility region for binary measurements.

Theorem 8.6 ([2, Theorem V.7]). *It holds that* $\Gamma(g, d, 2^{\times g}) = \Delta_{\mathcal{D}_{\diamond,g}}(d)$ *for all* d, $g \in \mathbb{N}$.

The corresponding statement for an arbitrary number of outcomes is [4, Theorem 5.3]. Theorem 8.6 allows us to translate bounds on the inclusion set of the matrix diamond into bounds on the balanced compatibility region of binary measurements and vice versa. Using the bound from [72] in Equation (7.3), we obtain in Theorem II.3 of Core Article II [2] that

$$\mathrm{QC}_g \subseteq \Gamma(g, d, 2^{\times g}) \qquad \forall g, d \in \mathbb{N}.$$

Combining this with the bound in Equation (7.4), which is also based on [72], we obtain [2, Theorem II.4]

$$\mathrm{QC}_g = \Gamma(g, d, 2^{\times g}) \qquad \forall g \geq 2,\ \forall d \geq 2^{\lceil (g-1)/2 \rceil}.$$

This is a marked improvement over the results for two measurements in [82] and three qubit measurements in [85, 86] discussed in Section 8.2. In Core Article II [2], we also derive more lower bounds which improve over the ones obtained from approximate cloning. These and other results are collected in Section IX.A of Core Article II [2] and Section 10 of Article VI [4].

Bibliography

[1] A. Bluhm, L. Rauber, and M. M. Wolf. Quantum compression relative to a set of measurements. *Annales Henri Poincaré*, 19(6):1891–1937, 2018.

[2] A. Bluhm and I. Nechita. Joint measurability of quantum effects and the matrix diamond. *Journal of Mathematical Physics*, 59:112202, 2018.

[3] A. Bluhm and D. Stilck França. Dimensionality reduction of SDPs through sketching. *Linear Algebra and its Applications*, 563:461–475, 2019.

[4] A. Bluhm and I. Nechita. Compatibility of quantum measurements and inclusion constants for the matrix jewel. *arXiv preprint arXiv:1809.04514*, 2018. Submitted to *SIAM Journal on Applied Algebra and Geometry.*

[5] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 10th anniversary edition, 2010.

[6] D. Petz. *Quantum Information Theory and Quantum Statistics.* Springer, 2008.

[7] T. Heinosaari and M. Ziman. *The Mathematical Language of Quantum Theory.* Cambridge University Press, 2011.

[8] J. Watrous. *The Theory of Quantum Information.* Cambridge University Press, 2018.

[9] M. Keyl. Fundamentals of quantum information theory. *Physics Reports*, 369(5):431–548, 2002.

[10] J. B. Conway. *A Course in Functional Analysis*, volume 96 of *Graduate Texts in Mathematics.* Springer, 1985.

[11] R. A. Horn. *Matrix Analysis.* Cambridge University Press, second edition, 2012.

[12] R. Shankar. *Principles of Quantum Mechanics.* Springer, second edition, 1994.

[13] V. Paulsen. *Completely Bounded Maps and Operator Algebras.* Cambridge University Press, 2003.

[14] R. F. Werner and M. M. Wolf. Bell inequalities and entanglement. *Quantum Information & Computation*, 1(3):1–25, 2001.

[15] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23:880–884, 1969.

[16] A. Aspect, P. Grangier, and G. Roger. Experimental tests of realistic local theories via Bell's theorem. *Physical Review Letters*, 47(7):460–463, 1981.

[17] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81:865–942, 2009.

[18] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, 1975.

[19] K. Kraus. *States, Effects and Operations*, volume 190 of *Lecture Notes in Physics*. Springer, 1983.

[20] D. R. Farenick. *Algebras of Linear Transformations*. Springer, 2001.

[21] M. W. Hirsch. *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*. Springer, 1976.

[22] M. M. Wolf. Quantum channels and operations. Lecture notes available at: `http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf`, 2012.

[23] G. Lindblad. A general no-cloning theorem. *Letters in Mathematical Physics*, 47(2):189–196, 1999.

[24] W. Arveson. Subalgebras of C^*-algebras II. *Acta Mathematica*, 128(1):271–308, 1972.

[25] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[26] B. Gärtner and J. Matoušek. *Approximation Algorithms and Semidefinite Programming*. Springer, 2012.

[27] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5:217–238, 2009.

[28] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2):022308, 2004.

[29] P. A. Parrilo. Semidefinite optimization. In *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series on Optimization, pages 3–46. SIAM, 2012.

[30] T. Heinosaari, M. A. Jivulescu, D. Reeb, and M. M. Wolf. Extending quantum operations. *Journal of Mathematical Physics*, 53:102208, 2012.

[31] L. Porkolab and L. Khachiyan. On the complexity of semidefinite programs. *Journal of Global Optimization*, 10:351–365, 1997.

[32] M. Grötschel, L. Lovász, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization*, volume 2 of *Algorithms and Combinatorics*. Springer, second edition, 1988.

[33] S. Bubeck. Convex optimization: Algorithms and complexity. *Foundations and Trends in Machine Learning*, 8(3-4):231–358, 2015.

[34] E. de Klerk. *Aspects of Semidefinite Programming*, volume 65 of *Applied Optimization*. Springer, 2002.

[35] F. Le Gall. Powers of tensors and fast matrix multiplication. In *Proceedings of the 39th International Symposium on Symbolic and Algebraic Computation*, pages 296–303, 2014.

[36] Y. T. Lee, A. Sidford, and S. C.-W. Wong. A faster cutting plane method and its implications for combinatorial and convex optimization. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, pages 1049–1065, 2015.

[37] F. G. S. L. Brandão and K. M. Svore. Quantum speed-ups for solving semidefinite programs. In *Proceedings of the 58th Annual IEEE Symposium on Foundations of Computer Science*, pages 415–426, 2017.

[38] D. P. Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1–2):1–157, 2014.

[39] T. Sarlós. Improved approximation algorithms for large matrices via random projections. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pages 143–152, 2006.

[40] W. Johnson and J. Lindenstrauss. Extensions of Lipshitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[41] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 604–613, 1998.

[42] D. M. Kane and J. Nelson. Sparser Johnson-Lindenstrauss transforms. *Journal of the ACM*, 61(1):4:1–4:23, 2014.

[43] J. Nelson and H. L. Nguyen. Sparsity lower bounds for dimensionality reducing maps. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, pages 101–110, 2013.

[44] N. Ailon and B. Chazelle. Approximate nearest neighbors and the Fast Johnson-Lindenstrauss transform. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 557–563, 2006.

[45] K. Vu, P.-L. Poirion, and L. Liberti. Random projections for linear programming. *Mathematics of Operations Research*, 43(4):1051–1071, 2018.

[46] A. Yurtsever, M. Udell, J. A. Tropp, and V. Cevher. Sketchy decisions: Convex low-rank matrix optimization with optimal storage. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pages 1188–1198, 2017.

[47] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski. Phase retrieval via matrix completion. *SIAM Review*, 57(2):225–251, 2015.

[48] A. Yurtsever, Y.-P. Hsieh, and V. Cevher. Scalable convex methods for phase retrieval. In *6th International IEEE Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pages 381–384, 2015.

[49] J. A. Tropp, A. Yurtsever, M. Udell, and V. Cevher. Practical sketching algorithms for low-rank matrix approximation. *SIAM Journal on Matrix Analysis and Applications*, 38(4):1454–1485, 2017.

[50] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 623–656, 1948.

[51] B. Schumacher. Quantum coding. *Physical Review A*, 51(4):2738–2747, 1995.

[52] M. A. Nielsen. *Quantum information theory.* PhD thesis, The University of New Mexico, 1998.

[53] M. M. Wilde. *Quantum Information Theory.* Cambridge University Press, 2017.

[54] S. Massar and S. Popescu. Amount of information obtained by a quantum measurement. *Physical Review A*, 61:062303, 2000.

[55] A. Winter and S. Massar. Compression of quantum-measurement operations. *Physical Review A*, 64(1):012311, 2001.

[56] A. Winter. "Extrinsic" and "intrinsic" data in quantum measurements: Asymptotic convex decomposition of positive operator valued measures. *Communications in Mathematical Physics*, 244(1):157–185, 2004.

[57] M. M. Wilde, P. Hayden, F. Buscemi, and M.-H. Hsieh. The information-theoretic costs of simulating quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 45(45):453001, 2012.

[58] G. Aubrun and C. Lancien. Zonoids and sparsification of quantum measurements. *Positivity*, 20(1):1–23, 2016.

[59] G. Aubrun and S. J. Szarek. *Alice and Bob meet Banach*, volume 223 of *Mathematical Surveys and Monographs.* American Mathematical Society, 2017.

[60] C. J. Stark and A. W. Harrow. Compressibility of positive semidefinite factorizations and quantum models. *IEEE Transactions on Information Theory*, 62(5):2867–2880, 2016.

[61] S. Wehner, M. Christandl, and A. C. Doherty. Lower bound on the dimension of a quantum system given measured data. *Physical Review A*, 78:062112, 2008.

[62] S. Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing*, pages 325–338, 2018.

[63] R. O'Donnell and J. Wright. Efficient quantum tomography. In *Proceedings of the 48th Annual ACM Symposium on Theory of Computing*, pages 899–912, 2016.

[64] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017.

[65] D. Gosset and J. Smolin. A compressed classical description of quantum states. *arXiv preprint arXiv:1801.05721*, 2018.

[66] A. Ben-Tal and A. Nemirovski. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization*, 12(3):811–833, 2002.

[67] J. W. Helton, I. Klep, S. McCullough, and M. Schweighofer. Dilations, linear matrix inequalities, the matrix cube problem and beta distributions. *Memoirs of the American Mathematical Society*, 257(1232), 2019.

[68] R. M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, The IBM Research Symposia Series, pages 85–103. Springer, 1972.

[69] J. W. Helton, I. Klep, and S. McCullough. The matricial relaxation of a linear matrix inequality. *Mathematical Programming*, 138(1-2):401–445, 2013.

[70] K. R. Davidson, A. Dor-On, O. M. Shalit, and B. Solel. Dilations, inclusions of matrix convex sets, and completely positive maps. *International Mathematics Research Notices*, 2017(13):4069–4130, 2017.

[71] L. Gurvits. Classical deterministic complexity of Edmonds' Problem and quantum entanglement. In *Proceedings of the 35th annual ACM Symposium on Theory of Computing*, pages 10–19, 2003.

[72] B. Passer, O. M. Shalit, and B. Solel. Minimal and maximal matrix convex sets. *Journal of Functional Analysis*, 274:3197–3253, 2018.

[73] N. Bohr. The quantum postulate and the recent development of atomic theory. *Nature*, 121(3050):580–590, 1928.

[74] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3):172–198, 1927.

[75] T. Heinosaari, T. Miyadera, and M. Ziman. An invitation to quantum incompatibility. *Journal of Physics A: Mathematical and Theoretical*, 49(12):123001, 2016.

[76] P. Lahti. Coexistence and joint measurability in quantum mechanics. *International Journal of Theoretical Physics*, 42(5):893–906, 2003.

[77] M. M. Wolf, D. Pérez-García, and C. Fernández. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Physical Review Letters*, 103:230402, 2009.

[78] R. Uola, T. Moroder, and O. Gühne. Joint measurability of generalized measurements implies classicality. *Physical Review Letters*, 113:160403, 2014.

[79] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48(5):291–295, 1982.

[80] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86:419–478, 2014.

[81] T. Heinosaari, J. Kiukas, and D. Reitzner. Noise robustness of the incompatibility of quantum measurements. *Physical Review A*, 92:022115, 2015.

[82] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens. Comparing the degrees of incompatibility inherent in probabilistic physical theories. *EPL (Europhysics Letters)*, 103(1):10002, 2013.

[83] E. Bene and T. Vértesi. Measurement incompatibility does not give rise to Bell violation in general. *New Journal of Physics*, 20(1):013021, 2018.

[84] F. Hirsch, M. T. Quintino, and N. Brunner. Quantum measurement incompatibility does not imply Bell nonlocality. *Physical Review A*, 97:012129, 2018.

[85] T. Brougham and E. Andersson. Estimating the expectation values of spin-1/2 observables with finite resources. *Physical Review A*, 76:052313, 2007.

[86] R. Pal and S. Ghosh. Approximate joint measurement of qubit observables through an Arthur–Kelly model. *Journal of Physics A: Mathematical and Theoretical*, 44(48):485303, 2011.

[87] T. Heinosaari, J. Schultz, A. Toigo, and M. Ziman. Maximally incompatible quantum observables. *Physical Review A*, 378(24&25):1695–1699, 2014.

[88] R. F. Werner. Optimal cloning of pure states. *Physical Review A*, 58(3):1827–1832, 1998.

[89] M. Studziński, P. Ćwikliński, M. Horodecki, and M. Mozrzymas. Group-representation approach to $1 \to N$ universal quantum cloning machines. *Physical Review A*, 89(5):052322, 2014.

[90] A. Kay. Optimal universal quantum cloning: Asymmetries and fidelity measures. *Quantum Information & Computation*, 16(11&12):0991–1028, 2016.

[91] A.-L. Hashagen. Universal asymmetric quantum cloning revisited. *Quantum Information & Computation*, 17(9&10):0747–0778, 2017.

A Core Articles

A.1 Quantum compression relative to a set of measurements

Quantum compression relative to a set of measurements

Andreas Bluhm, Lukas Rauber, and Michael M. Wolf

Compression is essential in information theory, classical or quantum, to make efficient use of limited resources such as bandwidth or storage space. This is especially important in quantum information theory, where decoherence makes it hard to construct large quantum memories. In this work, we investigate the compression of a quantum system to one of smaller dimension such that the outcomes of a given set of measurements are preserved. Since classical storage is much cheaper than quantum storage, we allow for an arbitrary amount of classical side information.

In Section 4 of this paper, we introduce the setup we consider. The measurements we have fixed are given as a set $\mathcal{O}$ of elements of positive operator-valued measures. i.e. positive semidefinite matrices. The compression procedure consists of a compression channel and a decompression channel. The former maps a quantum system to a bipartite system consisting of a quantum system of smaller dimension and a classical system. The decompression channel maps again to a quantum system of the initial dimension which we then measure. We demand that for all elements in $\mathcal{O}$, the measurement outcomes for the quantum system after the compression and decompression procedure are the same as for the original quantum system. The minimal dimension to which compression is possible is called the *compression dimension*.

We begin our investigation in Section 5 with proving that we can relax the assumption that the measurement statistics need to be exactly conserved. Lemma 5.2 shows that the amount of classical information we can use for compression can be bounded. This allows us to prove in Theorem 5.1 that for every set of measurements, there is an $\epsilon > 0$ such that the compression dimension does not decrease if we allow that the measurement statistics are only preserved up to an error smaller than ϵ. This means that the compression dimension is stable.

To prove lower and upper bounds on the compression dimension in Sections 6 and 7, we use an operator-algebraic and an algebro-geometric approach which are independent. The operator-algebraic approach relies on fixed-point theorems for completely positive maps and uses the structure of the matrix algebra $\mathrm{C}^*(\mathcal{O})$ generated by $\mathcal{O}$. Any matrix algebra is essentially a direct sum of full matrix algebras. In Theorems 6.1 and 7.1, we prove that the compression dimension is in between the smallest and the largest among the dimensions of the full matrix algebras appearing in this representation. Based on these results, we give an example in Section 7.2 where compression to dimension 2 is possible and prove in Lemma 6.6 that for structureless measurements, no compression is possible. In Section 8, we show that the compression dimension is equal to the dimension of one of the full matrix algebras appearing in the representation of $\mathrm{C}^*(\mathcal{O})$ and give an algorithm based on a semidefinite program to compute it.

The algebro-geometric lower bound is given in Theorem 6.8 and relies on Bézout's theorem. For this lower bound, we consider the characteristic polynomial of an arbitrary linear combination of two elements of $\mathcal{O}$. We find that the smallest degree of the irreducible factors of this polynomial is a lower bound for the compression dimension. While we recover in Lemma 6.11 that structureless measurements cannot be compressed, the lower bound obtained using these techniques is weaker than the one using operator-algebraic methods. On the other hand, the algebro-geometric approach allows us to prove in Theorem 10.2 that the lower bound is unchanged if we allow the compression channel to act on several copies of the quantum system.

I was significantly involved in finding the ideas and carrying out the scientific work of all parts of this article. Furthermore, I was in charge of writing the article with the exception of Section 2.

Permission to include:

Andreas Bluhm, Lukas Rauber, and Michael M. Wolf
Quantum compression relative to a set of measurements.
Annales Henri Poincaré, 19(6), 1891–1937 (2018).

Birkhäuser - Physics | Copyright Information: Annales Henri Poincaré

Springer www.springer.com

Physics Home > Birkhäuser > Physics

Copyright Information: Annales Henri Poincaré

Copyright Information

For Authors

Submission of a manuscript implies: that the work described has not been published before (except in form of an abstract or as part of a published lecture, review or thesis); that it is not under consideration for publication elsewhere; that its publication has been approved by all co-authors, if any, as well as – tacitly or explicitly – by the responsible authorities at the institution where the work was carried out.

Author warrants (i) that he/she is the sole owner or has been authorized by any additional copyright owner to assign the right, (ii) that the article does not infringe any third party rights and no license from or payments to a third party is required to publish the article and (iii) that the article has not been previously published or licensed. The author signs for and accepts responsibility for releasing this material on behalf of any and all co-authors. Transfer of copyright to Springer (respective to owner if other than Springer) becomes effective if and when a Copyright Transfer Statement is signed or transferred electronically by the corresponding author. After submission of the Copyright Transfer Statement signed by the corresponding author, changes of authorship or in the order of the authors listed will not be accepted by Springer.

The copyright to this article, including any graphic elements therein (e.g. illustrations, charts, moving images), is assigned for good and valuable consideration to Springer effective if and when the article is accepted for publication and to the extent assignable if assignability is restricted for by applicable law or regulations (e.g. for U.S. government or crown employees).

The copyright assignment includes without limitation the exclusive, assignable and sublicensable right, unlimited in time and territory, to reproduce, publish, distribute, transmit, make available and store the article, including abstracts thereof, in all forms of media of expression now known or developed in the future, including pre- and reprints, translations, photographic reproductions and microform. Springer may use the article in whole or in part in electronic form, such as use in databases or data networks for display, print or download to stationary or portable devices. This includes interactive and multimedia use and the right to alter the article to the extent necessary for such use.

Authors may self-archive the Author's accepted manuscript of their articles on their own websites. Authors may also deposit this version of the article in any repository, provided it is only made publicly available 12 months after official publication or later. He/she may not use the publisher's version (the final article), which is posted on SpringerLink and other Springer websites, for the purpose of self-archiving or deposit. Furthermore, the Author may only post his/her version provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Prior versions of the article published on non-commercial pre-print servers like arXiv.org can remain on these servers and/or can be updated with Author's accepted version. The final published version (in pdf or html/xml format) cannot be used for this purpose. Acknowledgement needs to be given to the final publication and a link must be inserted to the published article on Springer's website, accompanied by the text "The final publication is available at link.springer.com". Author retains the right to use his/her article for his/her further scientific career by including the final published journal article in other publications such as dissertations and postdoctoral qualifications provided acknowledgement is given to the original source of publication.

Author is requested to use the appropriate DOI for the article. Articles disseminated via link.springer.com are indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

For Readers

While the advice and information in this journal is believed to be true and accurate at the date of its publication, neither the authors, the editors, nor the publisher can accept any legal responsibility for any errors or omissions that may have been made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

All articles published in this journal are protected by copyright, which covers the exclusive rights to

Ann. Henri Poincaré 19 (2018), 1891–1937

1424-0637/18/061891-47
published online March 17, 2018
https://doi.org/10.1007/s00023-018-0660-z

Annales Henri Poincaré

CrossMark

Quantum Compression Relative to a Set of Measurements

Andreas Bluhm, Lukas Rauber and Michael M. Wolf

Abstract. In this work, we investigate the possibility of compressing a quantum system to one of smaller dimension in a way that preserves the measurement statistics of a given set of observables. In this process, we allow for an arbitrary amount of classical side information. We find that the latter can be bounded, which implies that the minimal compression dimension is stable in the sense that it cannot be decreased by allowing for small errors. Various bounds on the minimal compression dimension are proven, and an SDP-based algorithm for its computation is provided. The results are based on two independent approaches: an operator algebraic method using a fixed-point result by Arveson and an algebro-geometric method that relies on irreducible polynomials and Bézout's theorem. The latter approach allows lifting the results from the single-copy level to the case of multiple copies and from completely positive to merely positive maps.

1. Introduction

Compression of information is essential in order to make efficient use of limited storage space or bandwidth. This is even more true if we work with quantum information for which decoherence is an existential threat that makes reliable storage or transmission an extraordinary difficult task.

In this work, we consider the situation in which an unknown quantum state has to be stored for some time before one out of a set of measurements is performed. We assume that this set is known beforehand and we investigate to what extent the required storage space, measured in terms of its Hilbert space dimension, can be reduced depending on the set of measurements. Intuitively, in this setup only the information relevant for the given set of measurements has to be preserved. So if this set is not too large and sufficiently benign, this might allow for compression that is either lossless or only introduces small errors in the measurement statistics. Since classical storage space is cheap compared to quantum storage, we allow for an arbitrary amount of classical side information in this process. One may envision the considered situation as part of a larger protocol, where one party has to wait for additional input that then determines the measurement to be performed. A different scenario where our analysis could be applicable is in protocols with a bounded storage assumption.

Before going into details, let us review some of the different notions of compression that appear in quantum information theory and see how they relate to or differ from the setup analyzed in this paper. A classical task is *quantum source coding* [30]. Here, one is given an ensemble of pure quantum states and a quantum source that prepares elements from this set with a given probability. The aim is to encode a string of these states into one of smaller length so that the original message can be retrieved up to a small error. The compression rate for which this is possible is famously bounded by the von Neumann entropy of the state describing the source, and asymptotically the error can be made arbitrarily small. Hence, in this setup compression works irrespective of the measurement or operation that is eventually performed on the system.

Another version of compression can be found in [37]. Given a quantum state and a positive-operator-valued measure (POVM), the task is to find another POVM acting on many copies of the state whose outcomes have fewer entropy. The new POVM is required to be close to the original one . This amounts to reducing the number of POVM elements compared to the POVM which consists of tensor powers of the original one. The compression rate can again be bounded in terms of the entropy of the state and properties of the original POVM. These results are proven in the asymptotic setting of many copies of a given quantum state, but the results in [1] show that similar compression is also possible in a non-asymptotic setup.

Instead of compressing either states or POVMs, one could also be interested in compressing both, which is the setting of *model compression*. Given a set of states and POVMs, the task is to find new states and POVMs in a Hilbert space of smaller dimension, such that the measurement statistics are unchanged, possibly up to a small error. The original and new elements need not be connected by a physical transformation. In [31], this was shown to be possible if all effect operators except one per POVM have low rank. Here, the compression is a nonlinear map. In the same vein, lower bounds in terms of the entropy of measurement outcomes have been proven in [35], based on random access codes.

Compared to the first two notions of compression discussed above, the setup of our paper starts with the single-copy scenario (rather than with the asymptotic case) and aims at minimizing the system size under the constraint that after decompression only the statistics of a given set of observables have to be preserved. In this respect, our setup is similar to model compression. Contrary to the setting of model compression, however, we demand both compression and decompression to be achieved by physical transformations. Moreover, we allow for an arbitrary amount of classical information, which is considered to be for free.

2. Main Results

In this section, we will briefly outline the framework together with our main results. More detailed formulations and further results will be provided in subsequent sections. The starting point of our analysis is a set of measurements described by positive-operator-valued measures (POVMs). This means that one can assign a positive operator—a so-called *effect operator*—to every measurable subset of outcomes. Let $\mathcal{O}$ be the collection of all effect operators that belong to the considered measurements. If the underlying Hilbert space has dimension D, then $\mathcal{O}$ is a subset of the set $\mathcal{M}_D$ of complex $D \times D$ matrices. The type of compression we are interested in is given by a *compression map* $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$ and a *decompression map* $\mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$. Both maps are completely positive, trace preserving and such that for every density operator $\rho \in \mathcal{M}_D$ and every effect $E \in \mathcal{O}$ it holds that

$$\mathrm{Tr}\left((\mathcal{D} \circ \mathcal{C})[\rho]E\right) = \mathrm{Tr}\left(\rho E\right). \tag{1}$$

That is, we require the measurement statistics after compression and decompression to be exactly preserved. Here, $n \in \mathbb{N}$ quantifies the amount of classical information and $d \in \mathbb{N}$ is the intermediate Hilbert space dimension that we want to minimize. For a given $\mathcal{O}$, the minimal such dimension will be called its *compression dimension*. If this equals D, we call $\mathcal{O}$ *incompressible*.

Our first finding (Lemma 5.2) is that the amount of classical information can without loss of generality be restricted to $4 \log D$ bits. More precisely, if a map $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_D$ can be realized as $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$ for given n, d, then it can be realized in this way with $n \leq D^4$ and d unchanged. This fact, together with a compactness argument, then enables us to prove (Theorem 5.1) that the compression dimension is stable in the following sense: For every set of measurements there is an $\epsilon > 0$ such that even if deviations from Eq. (1) up to ϵ are allowed, the compression dimension cannot be decreased. In other words, allowing for errors does not change the picture as long as these are small enough. In light of this, the remaining part of the work then considers exact compression.

We prove bounds on the compression dimension following two different approaches: an operator algebraic and an algebro-geometric approach, to which we will for brevity refer to as algebraic and geometric, respectively. The algebraic path is based on the C*-algebra $\mathrm{C}^*\left(\mathcal{O}\right)$ generated by $\mathcal{O}$. Being finite dimensional, it is, up to an isomorphism, always of the form

$$\mathrm{C}^*\left(\mathcal{O}\right) \simeq \bigoplus_i \mathcal{M}_{D_i}.$$

Theorem. (Algebraic bounds on the compression dimension) *Let d be the compression dimension of $\mathcal{O}$ and $\{D_i\}$ be the dimensions of the matrix algebras* *occurring in the representation of the* C^**-algebra* $\mathrm{C}^*(\mathcal{O})$*. Then, it holds that* $\min_i\{D_i\} \le d \le \max_i\{D_i\}$*.*

This is the content of Theorems 6.1 and 7.1. If $\mathcal{O}$ for instance contains the effect operators of two binary von Neumann measurements, then these bounds generically coincide and are equal to $d = 2$ if D is even (Sect. 7.2). For structureless $\mathcal{O}$ with more than two elements, however, Lemma 6.6 shows that the foregoing theorem implies that $d = D$, so that $\mathcal{O}$ is incompressible.

The bounds in the foregoing theorem are tight in the sense that they cannot be improved solely on the basis of $\mathrm{C}^*(\mathcal{O})$ (unless the algebra consists only of one large block and one or two blocks of dimension one, in which case $d = \max_i\{D_i\}$, cf. Corollary 8.5). In particular, there are cases where the compression dimension is substantially smaller than the algebra generated by $\mathcal{O}$. In more abstract terms, the C^*-algebra is too coarse and we need to resort to the operator system that is generated by $\mathcal{O}$. In doing so, the following is shown in Sect. 8. The complexity of the corresponding algorithm is analyzed in Appendix E.

Theorem. (Algorithm for the minimal compression dimension) *The compression dimension of* $\mathcal{O}$ *is given by one of the matrix dimensions* D_i *that occur in the representation of the* C^**-algebra* $\mathrm{C}^*(\mathcal{O})$*. It can be computed by an algorithm that is based on a semidefinite program.*

The proof of correctness for this algorithm implies that the amount of classical side information needed is upper-bounded by the number of matrix algebras occurring in the representation of $\mathrm{C}^*(O)$. This bound is sharper than the one on the classical side information needed for arbitrary maps of the form $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$.

The geometric approach leads to the following lower bound (Theorem 6.8):

Theorem. (Algebro-geometric lower bound on the compression dimension) *Let* E_1*,* E_2 *be in the real linear span of* $\mathcal{O}$ *and define the real polynomial* $p(x, z) := \det[x\mathbb{1} - E_1 - zE_2]$*. The smallest of the degrees of the irreducible factors over the reals of* p *is a lower bound on the compression dimension of* $\mathcal{O}$*.*

Again, if E_1 and E_2 are generic, structureless effect operators, then this lower bound is equal to D (Lemma 6.11). As such, the geometric lower bound turns out to be weaker than the algebraic one. However, it becomes more powerful if the setup is extended. For example, if several copies of the state ρ are provided, the geometric argument is still valid and the lower bound remains unchanged (cf. Theorem 10.2). The same is true if we allow for positive (de-)compression maps that are not necessarily completely positive (cf. Sect. 9). Irrespective of the method, all our results still hold if we are only interested in preserving the expectation values of the measurements instead of the full statistics. This is true, because the elements in $\mathcal{O}$ need not be positive but only Hermitian.

Along the way, we prove some results that might be of independent interest. This includes in particular results on (Schwarz-) positive maps.

3. Preliminaries

In this section, we will review some concepts and notations from quantum information theory and classical algebraic geometry. Let $\mathcal{M}_{m,n}$ for n, $m \in \mathbb{N}$ denote the complex $m \times n$ matrices, which we concisely write as $\mathcal{M}_n$ for $m = n$. The set of Hermitian $n \times n$ matrices will be written $\mathcal{M}_n^{\text{herm}}$, and the set of real symmetric ones $\mathcal{M}_n^{\text{sym}}$. For a set $\mathcal{O} \subset \mathcal{M}_n$, we will denote by $\mathrm{C}^*(\mathcal{O})$ the complex C^*-algebra generated by $\mathcal{O}$ and the identity matrix $\mathbb{1}$. We also need the unitary group on $\mathbb{C}^d$, $d \in \mathbb{N}$, which we write as $\mathcal{U}(d)$. By $\|\cdot\|_\infty$, we denote the operator norm, whereas $\|A\|_p$, $p \in \mathbb{N}$, is the Schatten p-norm for $A \in \mathcal{M}_d$. If $|\phi\rangle \in \mathbb{C}^d$, $\||\phi\rangle\|_2$ is its Euclidean norm. For brevity, we will often refer to the set $\{ 1, \ldots, n \}$ as $[n]$.

We will work exclusively in finite-dimensional settings with Hilbert space $\mathcal{H} \simeq \mathbb{C}^d$ for some $d \in \mathbb{N}$ so that the bounded linear operators are represented by $d \times d$ matrices with complex entries. The set of states/density operators is defined as $\mathcal{S}(\mathbb{C}^d) := \{ \rho \in \mathcal{M}_d : \operatorname{Tr}(\rho) = 1, \rho \geq 0 \}$. Any pure state on a bipartite system $\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$, d_A, $d_B \in \mathbb{N}$, can be expressed in terms of its Schmidt decomposition. This means that for any pure state $|\psi\rangle \in \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$ there are orthonormal sets $\{ |e_i\rangle \}_{i=1}^k \subset \mathbb{C}^{d_A}$ and $\{ |f_j\rangle \}_{j=1}^k \subset \mathbb{C}^{d_B}$ such that

$$|\psi\rangle = \sum_{i=1}^{k} \sqrt{\lambda_i}\, |e_i\rangle \otimes |f_i\rangle$$

for some $\lambda_i > 0$ for all $i \in [k]$ and such that $\sum_{i=1}^k \lambda_i = 1$. Here, $k \in \mathbb{N}$ is the Schmidt rank of $|\psi\rangle$ [18, Proposition 2.2.1]. This concept can be extended to mixed states [32, Definition 1]:

Definition 3.1. (*Schmidt number*) A mixed state $\rho \in \mathcal{S}(\mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B})$ has Schmidt number k if for any decomposition $\{ p_i \geq 0, |\psi_i\rangle \}_{i=1}^n$, $n \in \mathbb{N}$, with

$$\rho = \sum_{i=1}^{n} p_i\, |\psi_i\rangle\langle\psi_i|$$

at least one of the pure states $|\psi_i\rangle \in \mathbb{C}^{d_A} \otimes \mathbb{C}^{d_B}$, $i \in [n]$, has Schmidt rank k and there exists a decomposition into pure states such that every pure state has Schmidt rank at most k.

The concept of measurement will be expressed through the set of effect operators $\mathcal{E}(\mathbb{C}^d) := \{ E \in \mathcal{M}_d^{\text{herm}} : 0 \leq E \leq \mathbb{1} \}$. Let Σ be the set of measurement outcomes, which we assume to be countable for simplicity. A set of effect operators $\{ E_s \}_{s \in \Sigma}$, $E_s \in \mathcal{E}(\mathbb{C}^d)$ for all $s \in \Sigma$ characterizes a positive-operator-valued measure (POVM) if

$$\sum_{s \in \Sigma} E_s = \mathbb{1}$$

(cf. [18, Section 2.1.4]).

We describe transformations on physical systems by completely positive maps. Let $D \in \mathbb{N}$. Recall that a linear map $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_d$ is called m-positive if $\mathcal{T} \otimes \mathrm{id}_m : \mathcal{M}_D \otimes \mathcal{M}_m \to \mathcal{M}_d \otimes \mathcal{M}_m$ is positive, where id_m is the identity map on $\mathcal{M}_m$. $\mathcal{T}$ is completely positive if it is positive for all $m \in \mathbb{N}$. This is equivalent to $\mathcal{T}$ having the form $\mathcal{T}(A) = \sum_{i=1}^{k} V_i^* A V_i$, where $V_i \in \mathcal{M}_{D,d}$ are the Kraus operators [25, Section 8.2.3]. If the map is additionally trace preserving, we will call this a quantum channel or a CPTP map. For $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_d$ completely positive, the map $\mathcal{T}^* : \mathcal{M}_d \to \mathcal{M}_D$ will be the dual map with respect to the Hilbert–Schmidt inner product. If $\mathcal{T}$ is trace preserving, the dual channel $\mathcal{T}^*$ is unital and furthermore it is completely positive if and only if its dual map is. We will denote by $|\Omega\rangle$ a maximally entangled state on $\mathbb{C}^{D^2}$,

$$|\Omega\rangle := \frac{1}{\sqrt{D}} \sum_{j=1}^{D} |j\rangle \otimes |j\rangle ,$$

where $\{ |j\rangle \}_{j=1}^{D}$ is an orthonormal basis of $\mathbb{C}^D$. A convenient way to check complete positivity of a linear map $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_d$ is to compute its Choi matrix [8]

$$\tau = \mathcal{T} \otimes \mathrm{id}_D(|\Omega\rangle\langle\Omega|).$$

It is known that $\mathcal{T}$ is completely positive if and only if τ is positive. One type of completely positive map which we will encounter frequently is the map $\Theta_A : \mathcal{M}_D \to \mathcal{M}_d$, defined as

$$\Theta_A(B) := A^* B A \quad \forall B \in \mathcal{M}_D$$

for fixed $A \in \mathcal{M}_{D,d}$.

Apart from completely positive maps, we will also need the notion of Schwarz maps. These are the unital positive linear maps for which the Schwarz inequality

$$\mathcal{T}(A^*)\mathcal{T}(A) \leq \mathcal{T}(A^* A) \tag{2}$$

holds true. Note that every unital 2-positive map (and hence also every unital completely positive map) fulfills the Schwarz inequality [27, Exercise 3.4].

Furthermore, we will need some notation to work with polynomials. Let $\mathbb{R}[x_1, \ldots, x_n]$, $n \in \mathbb{N}$, be the ring of polynomials in n-variables with real coefficients. In this work, we will only be concerned with irreducibility over the reals. Let

$$\mathbf{H}^d(n) = \left\{ f \in \mathbb{R}[x_1, \ldots, x_n] : f(\lambda x_1, \ldots, \lambda x_n) = \lambda^d f(x_1, \ldots, x_n) \right\}$$

be the space of homogeneous polynomials in n variables of degree d, $d \in \mathbb{N}$. We will identify a polynomial with the vector of its coefficients when convenient. The set of homogeneous polynomials in n variables and of any degree will be denoted by $\mathbf{H}(n) = \bigcup_{d \in \mathbb{N}} \mathbf{H}^d(n)$. We recall which homogeneous polynomials are called hyperbolic:

Definition 3.2. (*Hyperbolic polynomials*) Let $p \in \mathbf{H}^d(n)$. It is called hyperbolic with respect to the vector $e \in \mathbb{R}^n$ if $p(e) \neq 0$ and if for all vectors $w \in \mathbb{R}^n$ the univariate polynomial $t \mapsto p(w - te)$ has only real roots.

We will write $\mathcal{Z}(f)$ for the real zero set of the polynomials contained in the ideal generated by f. For $f \in \mathbb{R}[x, y]$ ($f \in \mathbf{H}(3)$), this set will be called an algebraic curve in real (projective) space. We can always switch between homogeneous and affine coordinates by homogenization, introducing an additional variable and setting this additional variable to 1, respectively (cf. [4, §3]). To conclude this section, let us finish by stating a classical result in algebraic geometry about the number of intersections of two algebraic curves (cf. [4, Theorem 11.10]).

Lemma 3.3. (Bézout's theorem) *Let $f \in \mathbf{H}^m(3)$, $g \in \mathbf{H}^n(3)$ such that they have no common factors of positive degree over the real numbers. Then, the curves $\mathcal{Z}(f)$ and $\mathcal{Z}(g)$ intersect at most $m \cdot n$ times, counting multiplicities, in the real projective plane.*

Then, of course, it is also true that $\mathcal{Z}(f(\cdot, 1, \cdot))$ and $\mathcal{Z}(g(\cdot, 1, \cdot))$ intersect in at most $m \cdot n$ points, since going to the projective plane only adds intersection points at infinity (points with $y = 0$).

4. Setup

In most of this work, we will consider the following situation: We would like to perform at some later point a set of s measurements, $s \in \mathbb{N}$, each with countably many outcomes $\{ a_i^k \}_{i=1}^{m_k}$, $m_k \in \mathbb{N} \cup \{ \infty \}$, where the index k denotes the kth measurement. That is, upon preparation $\rho \in \mathcal{S}(\mathbb{C}^D)$ we obtain outcome a_i^k with probability $\mathrm{Tr}\left(\rho E_i^k\right)$ for all $i \in [m_k]$, $k \in [s]$. Here, $E_i^k \in \mathcal{E}(\mathbb{C}^D)$ is the effect operator associated with outcome a_i^k and the effect operators $\{ E_i^k \}_{i=1}^{m_k}$ belonging to the same measurement form a POVM. Let us define the set of these effect operators

$$\mathcal{O} = \left\{ E_i^k : i \in [m_k], k \in [s] \right\}.$$

Note that assuming the outcomes to be countable simplifies notation, but our setup can easily be adapted to real measurement outcomes, for example. In that case, each effect corresponds to a measurable set of outcomes. See [16, Section 3.1.4] for details.

We are given an unknown quantum state $\rho \in \mathcal{S}(\mathbb{C}^D)$ that we want to store. In order to use a minimum of storage space, we want to keep only the information in the state relevant for the measurements that give rise to $\mathcal{O}$. Motivated by the fact that classical information is cheap to store compared to quantum information, we aim to minimize the dimension of the quantum system while allowing for an arbitrarily large amount of classical side information. Therefore, we are looking for a quantum compression channel $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$ and a quantum decompression channel $\mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$ such that for their composition $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$, the outcomes of the specified observables occur with the same probability as for the original state:

$$\mathrm{Tr}\left(\rho E\right) = \mathrm{Tr}\left(\mathcal{T}(\rho) E\right) = \mathrm{Tr}\left(\rho \mathcal{T}^*(E)\right) \quad \forall \rho \in \mathcal{S}(\mathbb{C}^D), \quad \forall E \in \mathcal{O}.$$

The channels $\mathcal{C}$ and $\mathcal{D}$ can be seen as an instrument and a parameter-dependent operation, respectively (cf. [18, Section 3.2.5]). Now, we can define our notion of compression.

Definition 4.1. (*Compression of observables*) Let $\mathcal{O}$ be a set of Hermitian operators in $\mathcal{M}_D$. The *compression dimension* of $\mathcal{O}$ is the smallest $d \in \mathbb{N}$ for which there is an $n \in \mathbb{N}$, a CPTP map $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$ and a CPTP map $\mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$ such that for their composition $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$, the constraints

$$\operatorname{Tr}(\rho E) = \operatorname{Tr}(\mathcal{T}(\rho)E) = \operatorname{Tr}(\rho\mathcal{T}^*(E)) \quad \forall \rho \in \mathcal{S}(\mathbb{C}^D), \quad \forall E \in \mathcal{O} \tag{3}$$

are satisfied. If the compression dimension equals D, $\mathcal{O}$ is said to be *incompressible.*

Note that the constraints are linear; hence, the relevant object is the linear subspace spanned by the effect operators, not the effect operators themselves. As the dual channel $\mathcal{T}^*$ is unital, we can add the identity to $\mathcal{O}$ without loss of generality. Then, the linear subspace contains the identity operator and is therefore an operator system. Let us denote the Hermitian part of this operator system by

$$\mathcal{L}(\mathcal{O}) := \operatorname{span}_{\mathbb{R}} \{ \mathcal{O}; \mathbb{1} \} .$$

This also implies that it is irrelevant whether the effect operators belong to the same observable or to different ones, although these are two different physical situations. Therefore, we will henceforth only assume that $\mathcal{O} \subset \mathcal{M}_D^{\text{herm}}$ instead of requiring the elements in $\mathcal{O}$ to be positive or even effect operators.

5. Approximate Compression

In Sect. 4, we have demanded the measurement statistics to be exactly conserved. This may seem very restrictive, but we will see shortly that it can be relaxed without changing the picture. The aim of this section is to show that the inexact case in which we demand

$$|\operatorname{Tr}(\rho E) - \operatorname{Tr}(\mathcal{T}(\rho)E)| \le \epsilon \quad \forall \rho \in \mathcal{S}(\mathbb{C}^D), \quad \forall E \in \mathcal{O},$$

instead of Eq. (3) reduces to the exact case ($\epsilon = 0$) for ϵ small enough. This is the content of the following theorem:

Theorem 5.1. (Stability of compression dimension) *Let $\mathcal{O} \subset \mathcal{M}_D^{\text{herm}}$ be a compact set with compression dimension d. Then, there is an $\epsilon > 0$ such that for any $d' < d$, $d' \in \mathbb{N}$ and any CPTP maps $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_{d'} \otimes \mathbb{C}^n$, $\mathcal{D} : \mathcal{M}_{d'} \otimes \mathbb{C}^n \to \mathcal{M}_D$ with $n \in \mathbb{N}$ there is a state $\rho \in \mathcal{S}(\mathbb{C}^D)$ and an operator $E \in \mathcal{O}$ for which*

$$|\operatorname{Tr}(\rho E) - \operatorname{Tr}((\mathcal{D} \circ \mathcal{C})[\rho]E)| \ge \epsilon.$$

The compression dimension is therefore stable under small errors. To prove the statement, we will need the following lemma, which shows that $4 \log D$ bits of classical side information suffice for compression.

Lemma 5.2. (Bound on classical information) *Let $\mathcal{C}$, $\mathcal{D}$ be two CPTP maps, $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$, $\mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$, $n \in \mathbb{N}$ and $d \le D$. We define $\mathcal{T} := \mathcal{D} \circ \mathcal{C}$. Then, there are two CPTP maps $\widetilde{\mathcal{C}} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^{n_0}$, $\widetilde{\mathcal{D}} : \mathcal{M}_d \otimes \mathbb{C}^{n_0} \to \mathcal{M}_D$ with $n_0 \in \mathbb{N}$, $n_0 \le D^4$ such that $\mathcal{T} = \widetilde{\mathcal{D}} \circ \widetilde{\mathcal{C}}$.*

Proof. Note that $\mathcal{M}_d \otimes \mathbb{C}^n \simeq \bigoplus_{i=1}^n \mathcal{M}_d$ has a block structure. Let P_i be the projection onto the ith block. Then, $\mathcal{T}_i := \mathcal{D} \circ \Theta_{P_i} \circ \mathcal{C}$ is again a completely positive map, although not necessarily trace preserving. The Choi matrix can thus be written

$$(\mathcal{T} \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|) = \sum_{i=1}^{n} (\mathcal{T}_i \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|).$$

We will argue that the Choi matrix has Schmidt number at most d (see Definition 3.1). By the introduction of an isometry $V_i : \mathbb{C}^d \hookrightarrow \mathbb{C}^{nd}$ such that $V_i V_i^* = P_i$, we can decompose $\mathcal{T}_i = \mathcal{D}_i \circ \mathcal{C}_i$ with $\mathcal{C}_i : \mathcal{M}_D \to \mathcal{M}_d$ where $\mathcal{C}_i = \Theta_{V_i} \circ \mathcal{C}$ and $\mathcal{D}_i : \mathcal{M}_d \to \mathcal{M}_D$ where $\mathcal{D}_i = \mathcal{D} \circ \Theta_{V_i^*}$. Therefore, it is easy to see that $(\mathcal{C}_i \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$ has Schmidt number at most d. Embedding $\mathcal{M}_d \hookrightarrow \mathcal{M}_D$, we can regard $\mathcal{D}_i$ as a map from $\mathcal{M}_D$ to itself. Since it only acts on one part of the bipartite system, $\mathcal{D}_i \otimes \mathrm{id}$ is a local operation. It is well known that such operations cannot increase the Schmidt number [32, Proposition 1]. Hence, $(\mathcal{T}_i \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$ has Schmidt number at most d and the same holds for $(\mathcal{T} \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|)$. An alternative way to see this is to note that the Kraus operators of $\mathcal{C}_i$ and $\mathcal{D}_i$ give a decomposition of $\mathcal{T}_i$ into Kraus operators of rank at most d.

Now, consider

$$\mathcal{S}_d = \left\{ |\psi\rangle\langle\psi| \in \mathcal{S}(\mathbb{C}^D \otimes \mathbb{C}^D) : |\psi\rangle \text{ has Schmidt rank } \leq d \right\}.$$

The set of states on $\mathbb{C}^D \otimes \mathbb{C}^D$ with Schmidt number at most d can then be written as the convex hull of $\mathcal{S}_d$. By Carathéodory's theorem, for every $\rho \in \mathcal{S}(\mathbb{C}^D \otimes \mathbb{C}^D)$ of Schmidt number at most d there are D^4 elements of $\mathcal{S}_d$ such that ρ can be written as a convex combination of these elements. We only need D^4 instead of D^4+1 elements, since $\mathcal{S}(\mathbb{C}^D \otimes \mathbb{C}^D)$ is contained in an affine subspace of dimension $D^4 - 1$. That means

$$(\mathcal{T} \otimes \mathrm{id})(|\Omega\rangle\langle\Omega|) = \sum_{i=1}^{D^4} p_i \, |\psi_i\rangle\langle\psi_i| \quad |\psi_i\rangle\langle\psi_i| \in \mathcal{S}_d, p_i \geq 0, \sum_{i=1}^{D^4} p_i = 1.$$

Each $p_i \, |\psi_i\rangle\langle\psi_i|$ can be regarded as Choi matrix of a completely positive map $\widetilde{\mathcal{T}}_i$. We would like to decompose these maps into $\widetilde{\mathcal{C}}_i : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^{D^4}$, $\widetilde{\mathcal{D}}_i : \mathcal{M}_d \otimes \mathbb{C}^{D^4} \to \mathcal{M}_D$, $\widetilde{\mathcal{T}}_i = \widetilde{\mathcal{D}}_i \circ \widetilde{\mathcal{C}}_i$. We note that since the Schmidt rank of $|\psi_i\rangle$ is at most d, we can write it as

$$|\psi_i\rangle = (X_i \otimes \mathbb{1}) \, |\Omega\rangle \quad X_i \in \mathcal{M}_D,$$

where X_i has rank at most d. We can take, e.g., $X_i = \sqrt{D \mathrm{Tr}_2 \left(|\psi_i\rangle\langle\psi_i|\right)} W$, where $W \in \mathcal{U}(D)$ and $\mathrm{Tr}_2 (\cdot)$ denotes the partial trace over the second system. Then, we can find $A_i \in \mathcal{M}_{D,d}$, $B_i \in \mathcal{M}_{d,D}$ such that $X_i = A_i B_i$ [15, Theorem 0.4.6 e)]. For A_i, we can use the polar decomposition such that $A_i = R_i Q_i$ with $Q_i \in \mathcal{M}_d$, $Q_i \geq 0$ and $R_i \in \mathcal{M}_{D,d}$ such that R_i has orthonormal columns, which means that R_i is an isometry [15, Theorem 7.3.1 c)]. Choose $\widetilde{\mathcal{C}}_i : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^{D^4}$ as

$$\widetilde{\mathcal{C}}_i := p_i \Theta_{(Q_i B_i)^*} \otimes |i\rangle\langle i|$$

and $\widetilde{\mathcal{D}}_i : \mathcal{M}_d \otimes \mathbb{C}^{D^4} \to \mathcal{M}_D$ as

$$\widetilde{\mathcal{D}}_i := \Theta_{R_i^*} \otimes \langle i|\cdot|i\rangle .$$

Then, we can define $\widetilde{\mathcal{C}} := \sum_{i=1}^{D^4} \widetilde{\mathcal{C}}_i$ and $\widetilde{\mathcal{D}} := \sum_{i=1}^{D^4} \widetilde{\mathcal{D}}_i$, where $\{\,|i\rangle\,\}_{i=1}^{D^4}$ is an orthonormal basis of $\mathbb{C}^{D^4}$. The maps $\widetilde{\mathcal{C}}$ and $\widetilde{\mathcal{D}}$ are CPTP with $\widetilde{\mathcal{D}} \circ \widetilde{\mathcal{C}} = \mathcal{T}$. □

Now, we want to argue that taking the infimum over channels which arise from compression and decompression maps amounts to taking the infimum over a compact set. Define

$$\begin{aligned}\mathcal{CH}_d := \ & \{\mathcal{T}^* : \mathcal{M}_D \to \mathcal{M}_D | \mathcal{T} \text{ CPTP };\exists \mathcal{C},\mathcal{D} \text{ s.t. } \mathcal{D}\circ\mathcal{C} = \mathcal{T},\\ & \mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n,\ \mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D, n \in \mathbb{N};\ \mathcal{C},\mathcal{D} \text{ CPTP}\}\end{aligned}$$

and

$$\begin{aligned}\widetilde{\mathcal{CH}}_d := \ & \{(\mathcal{C}^*,\mathcal{D}^*) | \mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^{D^4},\ \mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^{D^4} \to \mathcal{M}_D;\\ & \mathcal{C},\mathcal{D} \text{ CPTP}\}.\end{aligned}$$

Lemma 5.3. *$\widetilde{\mathcal{CH}}_d$ is a compact subset of the space $\mathcal{X} := \mathcal{B}(\mathcal{M}_d \otimes \mathbb{C}^{D^4}, \mathcal{M}_D) \times \mathcal{B}(\mathcal{M}_D, \mathcal{M}_d \otimes \mathbb{C}^{D^4})$ equipped with some norm $\|\cdot\|_{\mathcal{X}}$. Here, $\mathcal{B}(\mathcal{H},\mathcal{K})$ is the vector space of bounded linear operators from $\mathcal{H}$ to $\mathcal{K}$.*

Proof. Define

$$\mathcal{Y}_1 := \left\{ \mathcal{D}^* : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^{D^4} : \mathcal{D} \text{ CPTP} \right\}.$$

This set is both closed and bounded (by the Russo–Dye theorem). Since $\mathcal{B}(\mathcal{M}_D, \mathcal{M}_d \otimes \mathbb{C}^{D^4})$ is a finite-dimensional normed space, $\mathcal{Y}_1$ is compact. By the same reasoning,

$$\mathcal{Y}_2 := \left\{ \mathcal{C}^* : \mathcal{M}_d \otimes \mathbb{C}^{D^4} \to \mathcal{M}_D : \mathcal{C} \text{ CPTP} \right\}$$

is compact. It can easily be seen that $\widetilde{\mathcal{CH}}_d \simeq \mathcal{Y}_1 \times \mathcal{Y}_2$. Since products of compact sets are compact again in the product topology and all our spaces are finite dimensional, the assertion follows. □

Now, we can finally prove the main result of this section.

Proof of Theorem 5.1. Consider

$$\epsilon_{d'} := \inf_{\mathcal{T}^* \in \mathcal{CH}_{d'}} \max_{E \in \mathcal{O}} \|E - \mathcal{T}^*(E)\|_\infty.$$

By Lemma 5.2, we can equivalently write

$$\epsilon_{d'} := \inf_{(\mathcal{C}^*,\mathcal{D}^*) \in \widetilde{\mathcal{CH}}_{d'}} \max_{E \in \mathcal{O}} \|E - (\mathcal{C}^* \circ \mathcal{D}^*)(E)\|_\infty. \tag{4}$$

In Lemma 5.3, we have shown that $\widetilde{\mathcal{CH}}_{d'}$ is a compact set. Note that $\mathcal{R} \mapsto \max_{E\in\mathcal{O}} \|\mathcal{R}(E)\|_\infty$ is a seminorm for any linear map $\mathcal{R} : \mathcal{M}_D \to \mathcal{M}_D$ and seminorms on finite-dimensional vector spaces are continuous. It is thus easy to see that $f : (\widetilde{\mathcal{CH}}_{d'}, \|\cdot\|_{\mathcal{X}}) \to \mathbb{R}$, $f(\mathcal{C}^*, \mathcal{D}^*) = \max_{E\in\mathcal{O}} \|E - (\mathcal{C}^* \circ \mathcal{D}^*)(E)\|_\infty$ is continuous. Therefore, the infimum in Eq. (4) is attained and we can write

$$\epsilon_{d'} := \min_{(\mathcal{C}^*,\mathcal{D}^*)\in\widetilde{\mathcal{CH}}_{d'}} \max_{E\in\mathcal{O}} \|E - (\mathcal{C}^* \circ \mathcal{D}^*)(E)\|_\infty .$$

Let $\epsilon := \min_{d'\in[d-1]} \epsilon_{d'}$. As the compression dimension is d, we know that $\epsilon > 0$. This implies that for any $\mathcal{T}^* \in \mathcal{CH}_{d'}$, $d' \in [d-1]$, there is an $E \in \mathcal{O}$ such that

$$\max_{\rho\in\mathcal{S}(\mathbb{C}^D)} |\mathrm{Tr}\,(\rho E) - \mathrm{Tr}\,(\mathcal{T}(\rho)E)\,| \geq \epsilon.$$

□

6. Lower Bounds

6.1. Algebraic Arguments

In this section, we will prove and discuss a lower bound on the compression dimension using techniques from operator algebras. This lower bound will depend on the structure of the algebra which is generated by the measurements we would like to perform. Note that any finite-dimensional C*-subalgebra of $\mathcal{M}_D$ containing the identity has the form [10, Theorem 5.6]

$$U^* \left(\bigoplus_{i=1}^{s} \mathcal{M}_{D_i} \otimes \mathbb{1}_{m_i} \right) U$$

with $\sum_{i=1}^s D_i m_i = D$, $U \in \mathcal{U}(D)$. The following theorem will be the main result of this section.

Theorem 6.1. (Operator algebraic lower bound on compression dimension) *Let $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$ and*

$$\mathrm{C}^*\,(\mathcal{O}) = U^* \left(\bigoplus_{i=1}^{s} \mathcal{M}_{D_i} \otimes \mathbb{1}_{m_i} \right) U,$$

where $\sum_{i=1}^s D_i m_i = D$ and $U \in \mathcal{U}(D)$. Then, $\min_{i\in[s]} D_i$ is a lower bound on the compression dimension of $\mathcal{O}$. In particular, if $\mathrm{C}^*\,(\mathcal{O}) = \mathcal{M}_D$, then $\mathcal{O}$ is incompressible.

The proof of this statement goes back to an idea of Arveson [2, p. 288]. In his paper, he proved the following:

Lemma 6.2. *Let Φ be a unital completely positive map of a matrix algebra $\mathcal{M}_D$ onto itself whose fixed points algebraically generate the full matrix algebra. Then, Φ is the identity map.*

In Arveson's work, Lemma 6.2 follows from a more general statement about boundary representations (cf. [2, Theorem 2.1.1]). The proof of Theorem 6.1 uses Arveson's idea and extends it to more general situations, connecting it to the compression of quantum measurements. We start by proving a lemma which is essentially Lemma 1 on p. 285 f. in [2]. For this, we recall the definition of the support projection of a unital completely positive map.

Let $\mathcal{R}$ be such a map on a matrix $*$-algebra $\mathcal{A}$. Then, the support projection of $\mathcal{R}$ is the minimal orthogonal projection $P \in \mathcal{A}$ such that $\mathcal{R}(P) = \mathbb{1}$. An equivalent definition as well as basic properties of the support projection can be found in Appendix A (Lemma A.4).

Lemma 6.3. *Let $\mathcal{R}$ be a unital completely positive linear map on a finite-dimensional C^*-algebra $\mathcal{A} \subset \mathcal{M}_D$, $D \in \mathbb{N}$, such that $\mathcal{R} \circ \mathcal{R} = \mathcal{R}$. Let P be the support projection of $\mathcal{R}$. Then, P commutes with the fixed points of $\mathcal{R}$.*

Proof. Since for positive maps $\mathcal{R}(A)^* = \mathcal{R}(A^*)$ for all $A \in \mathcal{A}$, this implies that the set of fixed points is closed under involution. Thus, proving $PAP = AP$ for all fixed points A is enough, since it implies $AP = PA$ for self-adjoint elements and arbitrary fixed points can be decomposed into self-adjoint components. It is even sufficient to prove $PA^*PAP = PA^*AP$, since for any vector $|\phi\rangle \in \mathbb{C}^D$ it holds that

$$\begin{aligned}\|(\mathbb{1} - P)AP\,|\phi\rangle\|_2^2 &= \|AP\,|\phi\rangle\|_2^2 - \|PAP\,|\phi\rangle\|_2^2 \\ &= \langle\phi|PA^*AP|\phi\rangle - \langle\phi|PA^*PAP|\phi\rangle\,.\end{aligned}$$

By the polarization identity, this extends to all matrix elements. For the first equality, we used that $\mathbb{1} - P$ is an orthogonal projection. Now, let $A \in \mathcal{A}$ be a fixed point of $\mathcal{R}$. Then, $A^*A \le \mathcal{R}(A^*PA)$ follows from the Schwarz inequality, $\mathcal{R}(A) = \mathcal{R}(PA)$ and from the fact that A is a fixed point. Multiplying by P from both sides and using $A^*PA \le A^*A$, this gives

$$PA^*PAP \le PA^*AP \le P\mathcal{R}(A^*PA)P \tag{5}$$

This can be rewritten as $P\mathcal{R}(A^*PA)P - PA^*PAP \ge 0$. The support projection P fulfills the equation

$$\mathcal{R}(A) = \mathcal{R}(PAP) \quad \forall A \in \mathcal{A}$$

and $\mathcal{R}|_{P\mathcal{A}P}$ is faithful, i.e.,

$$\mathcal{R}(A) = 0 \leftrightarrow PAP = 0 \quad \forall A \in \mathcal{A}_+. \tag{6}$$

Here, $\mathcal{A}_+$ are the positive elements of the algebra. This implies that

$$P\mathcal{R}(A^*PA)P - PA^*PAP = 0$$

holds since $\mathcal{R}$ was assumed to be idempotent. The statement then follows from Eq. (5). □

We will also need a simple proposition which allows us to consider simpler algebras. From a physicist's point of view, the $*$-isomorphism π takes care of the right choice of measurement basis and the elimination of duplicate blocks in the structure of the operators in $\mathcal{O}$.

Proposition 6.4. *Let $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$ be such that*

$$\mathrm{C}^*(\mathcal{O}) = U^* \left(\bigoplus_{i=1}^{s} \mathcal{M}_{D_i} \otimes \mathbb{1}_{m_i} \right) U$$

with $\sum_{i=1}^{s} D_i m_i = D$ *and* $U \in \mathcal{U}(D)$*. Then, there exist unital CP maps* $\pi : \mathcal{M}_D \to \mathcal{M}_{\sum_{i=1}^{s} D_i}$ *and* $\pi^{-1} : \mathcal{M}_{\sum_{i=1}^{s} D_i} \to \mathcal{M}_D$ *such that*

$$\pi(\mathrm{C}^*(\mathcal{O})) = \bigoplus_{i=1}^{s} \mathcal{M}_{D_i} =: \mathcal{A}$$

and $\pi|_{\mathrm{C}^*(\mathcal{O})}$ *is a* $*$*-isomorphism with inverse* $\pi^{-1}|_{\mathcal{A}}$*. Moreover,* $\mathcal{O}$ *can be compressed to dimension* d *if and only if* $\pi(\mathcal{O})$ *can be compressed to dimension* d*.*

Proof. Let $A \in \mathrm{C}^*(\mathcal{O})$. Then, it has the form

$$A = U^* \left(\bigoplus_{i=1}^{s} A_i \otimes \mathbb{1}_{m_i} \right) U,$$

where $A_i \in \mathcal{M}_{D_i}$. It is easy to see that $\tilde{\pi} : \mathrm{C}^*(\mathcal{O}) \to \mathcal{A}$,

$$\tilde{\pi}(A) = \bigoplus_{i=1}^{s} A_i,$$

is a $*$-isomorphism. Note that both $\tilde{\pi}$ and its inverse $\tilde{\pi}^{-1}$ are unital completely positive maps. Let $\mathcal{E}_1 : \mathcal{M}_D \to \mathrm{C}^*(\mathcal{O})$ and $\mathcal{E}_2 : \mathcal{M}_{\sum_{i=1}^{s} D_i} \to \mathcal{A}$ be conditional expectations onto the respective subalgebras. These maps are known to be completely positive and unital. Then, $\pi = \tilde{\pi} \circ \mathcal{E}_1$ and $\pi^{-1} = \tilde{\pi}^{-1} \circ \mathcal{E}_2$ are the desired maps. Let $\mathcal{C}^* : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$, $\mathcal{D}^* : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$ be a dual compression and decompression map for $\mathcal{O}$, respectively. For the constraints in Eq. (3) to hold, $\mathcal{O}$ must be in the fixed-point set of $\mathcal{T}^* = \mathcal{C}^* \circ \mathcal{D}^*$. Then, $\pi \circ \mathcal{C}^*$ and $\mathcal{D}^* \circ \pi^{-1}$ are again dual channels and achieve compression to dimension d for $\pi(\mathcal{O})$, because $\pi(\mathcal{O})$ is contained in the fixed-point set for the composition of these maps. Conversely, let $\widetilde{\mathcal{C}}^* : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_{\sum_{i=1}^{s} D_i}$, $\widetilde{\mathcal{D}}^* : \mathcal{M}_{\sum_{i=1}^{s} D_i} \to \mathcal{M}_d \otimes \mathbb{C}^n$ be a dual compression and decompression map for $\pi(\mathcal{O})$, respectively. Then, by a similar argument, $\pi^{-1} \circ \widetilde{\mathcal{C}}^*$ and $\widetilde{\mathcal{D}}^* \circ \pi$ achieve compression to dimension d for $\mathcal{O}$. □

With these preparations, we can prove the main result of this section.

Proof of Theorem 6.1. By Proposition 6.4, we can assume without loss of generality that the algebra is of the form $\bigoplus_{i=1}^{s} \mathcal{M}_{D_i}$ for $\sum_{i=1}^{s} D_i = D$, because $\mathrm{C}^*(\mathcal{O})$ must be $*$-isomorphic to such an algebra. We already noted that for the constraints in Eq. (3) to hold, $\mathcal{O}$ must be in the fixed-point set of $\mathcal{T}^* = \mathcal{C}^* \circ \mathcal{D}^*$.

Now, we note that there is an idempotent map with the same fixed points as $\mathcal{T}^*$. We can, for example, consider the Cesàro-mean

$$\mathcal{T}^*_\infty = \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} (\mathcal{T}^*)^n.$$

It is known that $\mathcal{T}^*_\infty$ has the same fixed points as $\mathcal{T}^*$ and is unital, idempotent and also completely positive (cf. Lemma A.1). Moreover, $\mathcal{T}^*_\infty \circ \mathcal{T}^* = \mathcal{T}^*_\infty$ holds.

Now, we prove that

$$\mathcal{F} := \{ A \in \mathcal{M}_D : P\mathcal{T}^*(A)P = PAP;\ [P, A] = 0 \}$$

is a $*$-algebra, where P is the support projection of $\mathcal{T}_\infty^*$. We note that $\mathcal{F}$ is an operator system as $\mathcal{T}_\infty^*$ is a unital positive linear map and that P commutes with $\mathrm{C}^*(\mathcal{F})$. Thus, we only need to show that $\mathcal{F}$ is closed under multiplication. Using the Schwarz inequality and the fact that P is an orthogonal projection, it follows for $A \in \mathcal{F}$ that

$$\begin{aligned} PA^*AP &= PA^*PAP \\ &= P\mathcal{T}^*(A^*)P\mathcal{T}^*(A)P \\ &\leq P\mathcal{T}^*(A^*)\mathcal{T}^*(A)P \\ &\leq P\mathcal{T}^*(A^*A)P. \end{aligned}$$

Hence, we see that

$$P[\mathcal{T}^*(A^*A) - A^*A]P \geq 0.$$

Finally, we show that equality holds here. Applying $\mathcal{T}_\infty^*$ to this and using both $\mathcal{T}_\infty^*(PBP) = \mathcal{T}_\infty^*(B)$ for all $B \in \mathcal{M}_D$ and $\mathcal{T}_\infty^* \circ \mathcal{T}^* = \mathcal{T}_\infty^*$, we infer

$$\mathcal{T}_\infty^*(P\,[\mathcal{T}^*(A^*A) - A^*A]\,P) = 0.$$

This implies by faithfulness of $\mathcal{T}_\infty^*|_{P\mathcal{M}_DP}$ that

$$P\,[\mathcal{T}^*(A^*A) - A^*A]\,P = 0.$$

Thus, $A \in \mathcal{F}$ implies $A^*A \in \mathcal{F}$ and the fact that $\mathcal{F}$ is a $*$-algebra then follows from the polarization identity

$$\begin{aligned} B^*A = \ & \frac{1}{4}[(A+B)^*(A+B) - (A-B)^*(A-B) + \mathrm{i}(A+\mathrm{i}B)^*(A+\mathrm{i}B) \\ & - \mathrm{i}(A-\mathrm{i}B)^*(A-\mathrm{i}B)]. \end{aligned}$$

The second main ingredient of the proof is the fact that the support projection P of $\mathcal{T}_\infty^*$ commutes with the fixed points of the map as shown in Lemma 6.3. Then, P also commutes with every element of the C^*-algebra generated by the fixed points of $\mathcal{T}^*$. Thus, it commutes especially with $\mathrm{C}^*(\mathcal{O})$. Therefore, $\mathrm{C}^*(\mathcal{O}) \subset \mathcal{F}$ and

$$P[\mathcal{T}(A) - A]P = 0 \quad \forall A \in \mathrm{C}^*(\mathcal{O}).$$

We can now use the structure of $\mathrm{C}^*(\mathcal{O})$. By Schur's lemma, we can conclude that

$$P = \bigoplus_{i\in[s]} \chi_{\mathcal{I}}(i)\mathbb{1}_{\mathcal{M}_{D_i}}$$

for some $\mathcal{I} \subset [s]$, where $\chi_{\mathcal{I}}$ is the indicator function of the set $\mathcal{I}$. Let $V_i : \mathbb{C}^{D_i} \hookrightarrow \mathbb{C}^D$ for $i \in [s]$ be an isometry such that $V_iV_i^*$ is the projection onto the ith block. As $\theta_{V_i^*}(B) \in \mathrm{C}^*(\mathcal{O})$ for all $B \in \mathcal{M}_{D_i}$, we have shown that

$$(\Theta_{V_i} \circ \mathcal{T}^* \circ \Theta_{V_i^*})(A) = A \quad \forall A \in \mathcal{M}_{D_i}, i \in \mathcal{I}.$$

Thus, $\Theta_{V_i} \circ \mathcal{T}^* \circ \Theta_{V_i^*} = \mathrm{id}\,\forall i \in \mathcal{I}$ holds. By the definition of the support projection, we infer further that

$$\mathcal{T}_\infty^*((\mathbb{1} - P)A) = \mathcal{T}_\infty^*(A(\mathbb{1} - P)) = 0$$

for all $A \in \mathcal{M}_D$, hence especially $0 \oplus \mathcal{M}_{D_i} \oplus 0 \in \ker \mathcal{T}_\infty^*\, \forall i \in [s] \setminus \mathcal{I}$.

It could, however, be possible to enlarge the intermediate space, but to use classical side information to compress the quantum component of the system nonetheless. The following shows that this cannot happen. We identify the intermediate space $\mathcal{M}_d \otimes \mathbb{C}^n$ with $\bigoplus_{i=1}^n \mathcal{M}_d$. Let Q_i be the orthogonal projection onto the ith block, $i \in [n]$. Then,

$$\mathcal{T}_{ij} = \Theta_{V_j^*} \circ \mathcal{D} \circ \Theta_{Q_i} \circ \mathcal{C} \circ \Theta_{V_j}$$

is again a completely positive map and $\sum_{i=1}^n \mathcal{T}_{ij} = \mathrm{id}$ for every $j \in \mathcal{I}$. Looking at the Choi matrices for $\mathcal{T}_{ij}$, we can see that each needs to be proportional to $|\Omega\rangle\langle\Omega|$, because each Choi matrix is positive semidefinite and their sum is a rank-one projection. We infer that $\mathcal{T}_{ij}$ must be proportional to the identity channel, i.e., $\mathcal{T}_{ij} = p_i \mathrm{id}$, $p_i \geq 0$, $\sum_{i=1}^n p_i = 1$. This is a well-known result in quantum information (no information without disturbance, see [16, Section 5.2.2]). From the rank–nullity theorem, we conclude that $d \geq D_j$ for all $j \in \mathcal{I}$. As the set of fixed points of $\mathcal{T}^*$ is non-empty, we know that $\mathcal{I}$ has to be non-empty as well. From there, the lower bound on d follows. □

The following corollary follows immediately from the proof of Theorem 6.1.

Corollary 6.5. (Fixed points of Schwarz maps) *Let $\mathcal{T}^* : \mathcal{M}_D \to \mathcal{M}_D$ be a Schwarz map and $\mathcal{O}$ a set of fixed points of $\mathcal{T}^*$ such that*

$$\mathrm{C}^*(\mathcal{O}) = \bigoplus_{i=1}^{s} \mathcal{M}_{D_i}$$

and $\sum_{i=1}^s D_i = D$ and let $V_i : \mathbb{C}^{D_i} \hookrightarrow \mathbb{C}^D$ be an isometry such that $V_i V_i^*$ is the projection onto the ith block for $i \in [s]$. Then, there is an index set $\mathcal{I} \subset [s]$ such that $\Theta_{V_i} \circ \mathcal{T}^* \circ \Theta_{V_i^*} = \mathrm{id}$ for all $i \in \mathcal{I}$ and $0 \oplus \mathcal{M}_{D_i} \oplus 0 \in \ker \mathcal{T}_\infty^*$ for all $i \in [s] \setminus \mathcal{I}$, where $\mathcal{T}_\infty^*$ is the Cesàro-mean of $\mathcal{T}^*$. Moreover, $d \geq \max_{i \in \mathcal{I}} D_i$.

To conclude this section, we will prove that two matrices generically generate the full matrix algebra. This shows that a set of unstructured effect operators is typically incompressible. More precisely, we show that the set of pairs of Hermitian matrices which do not generate the full matrix algebra has measure zero.

Lemma 6.6. *Let $\mathcal{N} = \left\{ (A, B) \in \mathcal{M}_D^{\mathrm{herm}} \times \mathcal{M}_D^{\mathrm{herm}} : \mathrm{C}^*(\{ A, B \}) \subsetneq \mathcal{M}_D \right\}$. Then, the set $\mathcal{N}$ has Lebesgue measure zero on $\mathcal{M}_D^{\mathrm{herm}} \times \mathcal{M}_D^{\mathrm{herm}}$.*

Proof. By Burnside's theorem (cf. [22]), it is clear that $\mathcal{N}$ is contained in the set of tuples of matrices which have a non-trivial common invariant subspace. This requirement can be formulated as the zero set of a polynomial as we will see. From [11, Theorem 2.2], we know that if A, $B \in \mathcal{M}_D^{\mathrm{herm}}$ have a common invariant subspace of dimension k, then also

$$P_k(A, B) := \det \left[\sum_{i,j=1}^{D-1} \left[C_k(A)^i, C_k(B)^j \right]^* \left[C_k(A)^i, C_k(B)^j \right] \right] = 0$$

where $C_k(A)$ is the kth compound matrix of A, i.e., the matrix with entries $\det(A[\alpha|\beta])$ and α, β sequences of strictly increasing integers contained in $[n]$, $A[\alpha|\beta]$ the submatrix of A in rows α and columns β. The entries of $C_k(A)$ are arranged in lexicographical order. Multiplying the P_k, we obtain a polynomial $P := \prod_{k=1}^{n-1} P_k$ in the real and imaginary parts of the entries of A, B which contains $\mathcal{N}$ in its zero set. Since P is not identically zero, its zero set and therefore $\mathcal{N}$ must have measure zero. □

We could also consider $\mathcal{M}_D$ instead of $\mathcal{M}_D^{\text{herm}}$ and the statement would still hold. However, in the setting of (operator systems generated by) quantum observables, it is more natural to assume that the matrices involved are Hermitian.

6.2. Geometric Arguments

To give a different perspective on the problem, we will prove in this section again that compression in the setup of Sect. 4 is impossible in general, this time using basic techniques from algebraic geometry. This will be useful later to obtain results in situations in which we cannot apply the techniques of Sect. 6.1 (see Sect. 10). We emphasize again that we are interested in irreducibility over the reals. The following lemma is the main technical result of this section.

Lemma 6.7. *Let A, $B \in \mathcal{M}_D^{\text{herm}}$ such that $p(x,z) := \det[x\mathbb{1} - A - zB]$ is a polynomial of degree D with a decomposition into irreducible factors*

$$p(x,z) = \prod_{i=1}^{s} p_i(x,z)^{m_i} \quad m_i \in \mathbb{N},$$

where $\deg p_i = D_i$ and $\sum_{i=1}^{s} m_i D_i = D$. Moreover, let $W \subset \mathbb{R}$ be open and non-empty and let C, $F \in \mathcal{M}_d^{\text{herm}}$ be such that

$$\|C + tF\|_\infty = \|A + tB\|_\infty \quad \forall t \in W.$$

Then, this implies that $d \geq \min_{i\in[s]} D_i$.

From this statement follows in particular that $d \geq D$ if $p(x,z)$ is an irreducible polynomial. This lemma can be used to prove lower bounds on the compression dimension.

Theorem 6.8. (Lower bound on compression dimension (geometric)) *Let $\mathcal{O} \subset \mathcal{M}_D^{\text{herm}}$ be a set of Hermitian operators, E_1, $E_2 \in \mathcal{L}(\mathcal{O})$ and*

$$p(x,z) := \det[x\mathbb{1} - E_1 - zE_2].$$

Then, the smallest among the degrees of the irreducible factors of p is a lower bound on the compression dimension of $\mathcal{O}$. In particular, if p is irreducible over the reals, then $\mathcal{O}$ is incompressible.

Proof. First, we have that $\mathcal{T}^*$ is a contraction by the Russo–Dye theorem, since $\mathcal{T}^*$ is a positive unital map. The same is true for the dual channels $\mathcal{D}^*$ and $\mathcal{C}^*$. If we require Eq. (3) to hold, then $\mathcal{L}(\mathcal{O})$ has to be in the fixed-point space of $\mathcal{T}^*$ as seen before. By the fixed-point property, the quantity $\|E_1 + tE_2\|_\infty$ has to be preserved under $\mathcal{T}^*$ for all $t \in \mathbb{R}$. Here, we have taken the modulus and then the maximum over all states in Eq. (3). Since both $\mathcal{C}^*$ and $\mathcal{D}^*$ are contractions as well, this implies that

$$\|E_1 + tE_2\|_\infty = \|\mathcal{D}^*(E_1) + t\mathcal{D}^*(E_2)\|_\infty \quad \forall t \in \mathbb{R}.$$

The assertion then follows from Lemma 6.7. □

In fact, we can strengthen Theorem 6.8 in the case when $\mathrm{C}^*(\mathcal{O})$ is a proper subalgebra and we have more information on its block structure. This is captured by the next corollary.

Corollary 6.9. *Let $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$ be such that*

$$\mathrm{C}^*(\mathcal{O}) = U^* \left(\bigoplus_{i=1}^{s} \mathcal{M}_{D_i} \otimes \mathbb{1}_{m_i} \right) U$$

with $\sum_{i=1}^s D_i m_i = D$ and $U \in \mathcal{U}(D)$. Then, the minimal compression dimension is lower-bounded by D_{j_0} if there are E_1, $E_2 \in \mathcal{O}$, $j_0 \in [s]$ and an open set $V \subset \mathbb{R}$ such that

$$\|E_1 + tE_2\|_\infty = \|E_1^{j_0} + tE_2^{j_0}\|_\infty$$

for all $t \in V$ and $E_1^{j_0}$, $E_2^{j_0}$ are such that $\det\left[x\mathbb{1} - E_1^{j_0} - zE_2^{j_0}\right]$ is irreducible over the reals. Here, we have used that for all $E \in \mathcal{O}$ we can write

$$E = U^* \left(\bigoplus_{j=1}^{s} E^j \otimes \mathbb{1}_{m_j} \right) U$$

for $E^j \in \mathcal{M}_{D_j}$, $j \in [s]$.

Proof. As in the proof of Theorem 6.8, we obtain

$$\|E_1 + tE_2\|_\infty = \|\mathcal{D}^*(E_1) + t\mathcal{D}^*(E_2)\|_\infty \quad \forall t \in \mathbb{R}. \tag{7}$$

The definition of $\mathcal{D}$ requires

$$\mathcal{D}^*(\mathcal{L}(E_1, E_2)) \subset \bigoplus_{i=1}^{n} \mathcal{M}_d \simeq \mathcal{M}_d \otimes \mathbb{C}^n.$$

Assume therefore that $\mathcal{D}^*(E_1) + t\mathcal{D}^*(E_2) = \bigoplus_{i=1}^n \left(F_1^i + tF_2^i\right)$, $F_j^i \in \mathcal{M}_d$ for all $i \in [n]$, $j \in [2]$ and $t \in \mathbb{R}$. Since

$$\|\mathcal{D}^*(E_1) + t\mathcal{D}^*(E_2)\|_\infty = \max_{i \in [n]} \left\|F_1^i + tF_2^i\right\|_\infty$$

for a fixed t, we can assume that there is an open set $W \subset V$ such that

$$\|\mathcal{D}^*(E_1) + s\mathcal{D}^*(E_2)\|_\infty = \left\|F_1^{k_0} + sF_2^{k_0}\right\|_\infty \quad \forall s \in W \tag{8}$$

for some $k_0 \in [n]$. This is true since for two blocks either

$$\left\|F_1^1 + tF_2^1\right\|_\infty = \left\|F_1^2 + tF_2^2\right\|_\infty$$

for all $t \in V$ or there is a $t_0 \in V$ such that

$$\left\|F_1^1 + t_0F_2^1\right\|_\infty > \left\|F_1^2 + t_0F_2^2\right\|_\infty.$$

In the latter case, we can find an open neighborhood W of t_0 such that

$$\left\|F_1^1 + t_0 F_2^1\right\|_\infty > \left\|F_1^2 + t_0 F_2^2\right\|_\infty$$

for all $t \in W$ by continuity of the operator norm with respect to t. This can be extended to more blocks by induction in the block number and possibly further shrinking W.

By assumption, Eqs. (7) and (8) then imply

$$\left\|E_1^{j_0} + tE_2^{j_0}\right\|_\infty = \left\|F_1^{k_0} + tF_2^{k_0}\right\|_\infty \quad \forall t \in W.$$

The assertion $d \geq D_{j_0}$ then follows from Lemma 6.7. □

The condition $\|E_1 + tE_2\|_\infty = \|E_1^{j_0} + tE_2^{j_0}\|_\infty$ might look artificial, but can easily be checked. We just have to find a $t_0 \in \mathbb{R}$ which is not a crossing point and check which block has the largest operator norm in some open neighborhood of t_0. If furthermore $\det[x\mathbb{1} - E_1^{j_0} - zE_2^{j_0}]$ is irreducible (this might be hard to check), we can apply the above corollary to find a lower bound on d. Note that the condition also implies that the j_0th block is not redundant (cf. discussion in Sect. 8), since Lemma 6.7 guarantees that smaller blocks have smaller operator norm for some $t \in U$. By contractivity, it then follows that there is no unital completely positive map $\Phi : \mathcal{M}_{\sum_{j=1}^s D_j} \to \mathcal{M}_{D_{j_0}}$ such that

$$E_k^{j_0} = \Phi\left(\bigoplus_{i=1}^s \chi_{\mathcal{I}}(i) E_k^i\right) \quad \forall k \in [2]$$

and $\mathcal{I}$ such that $D_i < D_{j_0}\, \forall i \in \mathcal{I}$ and $\chi_{\mathcal{I}}$ is the indicator function of $\mathcal{I}$. We still have to prove Lemma 6.7, which we will do now.

Proof of Lemma 6.7. First note that $A + tB$ has only real eigenvalues for $t \in \mathbb{R}$. Thus, for any fixed t, the characteristic polynomial has D real solutions counting multiplicities. Without loss of generality, let $U \subset W$ be a non-empty open set such that $\|A + tB\|_\infty$ is the maximal eigenvalue $\lambda_{max}(t)$ of $A+tB$ for all $t \in U$ and the same holds for $C + tF$. We denote the maximal eigenvalue of the latter matrix by $\mu_{max}(t)$. This is possible, since there are only finitely many level crossings in any finite interval (cf. [17, p. 124]). Moreover, if the minimal eigenvalue of $A + tB$ has larger modulus, we can consider $-(A + tB)$ instead which clearly has the same operator norm and the same is possible for $C + tF$. Then,

$$V := \{\, (x, z) : x = \lambda_{max}(z), z \in U \,\}$$

is a subset of $\mathcal{Z}(p)$ with infinitely many points since U is open in $\mathbb{R}$. Let

$$q(x, z) := \det[x\mathbb{1} - C - zD]$$

which is a polynomial of degree d. Assume $d < D_i$ for all $i \in [s]$. Since the p_i are irreducible by assumption, p_i and q have no common factors for any $i \in [s]$. Therefore, by Bézout's theorem and since

$$\mathcal{Z}(p) \cap \mathcal{Z}(q) = \bigcup_{i \in [s]} (\mathcal{Z}(p_i) \cap \mathcal{Z}(q)),$$

the zero sets of the two polynomials have at most $\sum_{i=1}^{s} d \cdot D_i$ points in common (cf. Lemma 3.3). Thus, $\mathcal{Z}(q)$ especially cannot contain V, which implies $\|A + tB\|_\infty \neq \|C + tF\|_\infty$ for infinitely many $t \in U$, since

$$\{ (x, z) : x = \mu_{max}(z), z \in U \} \subset \mathcal{Z}(q).$$

□

Let us make the following remark concerning our use of Bézout's theorem. Commonly, the theorem is formulated as an equality (counting multiplicities) over an algebraically closed field such as $\mathbb{C}$. Since real polynomials are coprime over the reals if and only if they are coprime over the complex numbers (cf. [4, Theorem 11.9]), the complex version of Bézout's theorem implies an upper bound on the number of intersections of real coprime polynomials over the reals which we used here (cf. [4, Theorem 11.10]).

The last question we have to answer in this section is the existence of irreducible polynomials of any degree which arise from a determinant of $D \times D$ matrices. We would also like to know how common these are. This will also show that there are effect operators which give rise to irreducible polynomials. For this, we do not require the matrices A, $B \in \mathcal{M}_D^{\text{herm}}$ to be positive, because we can convert them into effect operators. For any $A \in \mathcal{M}_D^{\text{herm}}$, there is a $\lambda \in \mathbb{R}$ such that $A + \lambda\mathbb{1} \geq 0$ and we can scale this expression by a positive scalar such that it becomes smaller than the identity operator. This way, we can find nonzero effect operators E_1, E_2 such that $A, B \in \mathcal{L}(\{ E_1, E_2 \})$ and E_1, E_2 are fixed points if and only if A, B are. Furthermore, $\det[x\mathbb{1} - E_1 - zE_2]$ is irreducible if and only if $\det[x\mathbb{1} - A - zB]$ is irreducible for linearly independent A, $B \in \mathcal{L}(E_1, E_2)$, since a (non-singular) coordinate transformation does not change reducibility properties of the polynomial (cf. [4, discussion before Theorem 4.5]). The key ingredient to show existence of the required polynomials is the Lax conjecture which was proven in [21, Conjecture 4]. We give it here for convenience.

Theorem 6.10. (Lax conjecture) *A polynomial $p \in \mathbf{H}^D(3)$ is hyperbolic with respect to the vector $e := (1, 0, 0)$ and satisfies $p(e) = 1$ if and only if there exist matrices $A, B \in \mathcal{M}_D^{\text{sym}}$ such that p is given by*

$$p(x, y, z) = \det\left[x\mathbb{1} + yA + zB\right].$$

The result that A, B can be chosen real symmetric is even stronger than needed for our purposes.

Lemma 6.11. *For any $D \in \mathbb{N}$, there is an irreducible homogeneous polynomial and A, $B \in \mathcal{M}_D^{\text{sym}}$ such that*

$$p(x, y, z) = \det[x\mathbb{1} + yA + zB].$$

Moreover, these elements are generic in the set of homogeneous polynomials normalized to $p(e) = 1$ for $e := (1, 0, 0)$.

Proof. By the Lax conjecture, it suffices to show that there are homogeneous polynomials of any degree which are both hyperbolic with respect to e and irreducible. The case $D = 1$ is trivial, since there are no reducible elements and all polynomials are hyperbolic. Hence, assume $D > 1$. It is known that the set of reducible elements in this case does not contain any open subset in the Euclidean topology (see Lemma B.1 for a proof). Since the set of hyperbolic polynomials with respect to a fixed point e has non-empty interior in this topology by [26] (cf. Appendix B to see that this is not affected by normalization), it especially contains an open set, hence it cannot be fully contained in the set of reducible elements. Therefore, there must be elements which are both hyperbolic and irreducible. Lemma B.1 also states that the set of normalized reducible polynomials has measure zero, hence its intersection with the set of normalized hyperbolic polynomials has measure zero as well. □

Theorem 6.8 states that compression is not possible if the polynomial

$$p(x, y, z) = \det[x\mathbb{1} - yA - zB] \tag{9}$$

is irreducible, where $A, B \in \mathcal{L}(E_1, E_2)$. Lemma 6.11 therefore implies that effect operators which cannot be compressed are the generic case, i.e., the set of $p(x, y, z)$ corresponding to effect operators which admit compression has Lebesgue measure zero in the space of normalized homogeneous polynomials in three variables of fixed degree D. This follows because p has to be hyperbolic to admit a determinantal representation as in Eq. (9), even if we allow for Hermitian matrices. Furthermore, p needs to be reducible to possibly admit a compression by the above. Unfortunately, irreducibility over the reals is difficult to check.

So far, we have only shown existence of such $p(x, y, z)$. We can also give an explicit example of such a polynomial in every dimension (with Hermitian matrices).

Proposition 6.12. *Let*

$$A := \frac{1}{2}\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & \ddots & & \vdots \\ \vdots & & \ddots & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix}, \quad B := \frac{1}{2}\begin{bmatrix} 0 & \mathrm{i} & \dots & \mathrm{i} \\ -\mathrm{i} & \ddots & & \vdots \\ \vdots & & \ddots & \mathrm{i} \\ -\mathrm{i} & \dots & -\mathrm{i} & 0 \end{bmatrix},$$

$A, B \in \mathcal{M}_D$, $D \geq 1$. Then, the polynomial $p(x, z) := \det[x\mathbb{1} + A + zB]$ is irreducible.

Proof. $D = 1$ is trivial, thus assume $D \geq 2$. Reparameterizing with $\tilde{z} := z - \mathrm{i}$, we obtain

$$\tilde{p}(x, \tilde{z}) := \det\left[\tilde{A}(x) + \tilde{z}B\right]$$

with

$$\tilde{A}(x) := \begin{bmatrix} x & 0 & \dots & 0 \\ 1 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 1 & \dots & 1 & x \end{bmatrix}.$$

We know that $\tilde{p}(x,0) = x^D$. In order to prove that $\tilde{p}$ is irreducible, we show first that it cannot be decomposed as $\tilde{p} = q \cdot r$ with $q,\ r \in \mathbb{C}[x,\tilde{z}]$ with $q(0,0) = 0 = r(0,0)$. Since the constant terms of both q and r must be zero, the expansion of $\tilde{p}(x,\epsilon)$ to first order in ϵ would have at least one root $x = 0$ if such a decomposition existed (cf. Lemma C.5). We can expand

$$\tilde{p}(x,\epsilon) := x^D - \frac{\epsilon \mathrm{i}}{2}\left[Dx^{D-1} + (x-1)^D - x^D\right] + \mathcal{O}(|\epsilon|^2)$$

(cf. Lemma C.4). However,

$$\tilde{p}(0,\epsilon) = (-1)^{D+1}\epsilon\mathrm{i}/2 + \mathcal{O}(|\epsilon|^2),$$

i.e., the term linear in ϵ does not vanish. Therefore, without loss of generality, $r(0,0)$ is nonzero. This implies that r is constant, since x cannot divide $r(x,0)$, which implies that $q(x,0)$ must have degree D and $\tilde{p}$ is of degree D. Thus, $\tilde{p}$ is irreducible (even over the complex numbers) and hence the same holds for p. □

6.3. Comparing the Arguments

Before we continue, let us compare the two techniques used to prove that compression is not possible in general. We will see that the algebraic method shows incompressibility for a larger class of effect operators. However, we will see in Sect. 10 that the geometric argument can be used in situations where the algebraic argument is not applicable.

If $\mathrm{C}^*(\mathcal{O})$ is only a subalgebra of $\mathcal{M}_D$, then $\mathcal{L}(\mathcal{O}) \subset U^*(\mathcal{M}_{D_1} \oplus \mathcal{M}_{D_2})U$, with $D_1,\ D_2 \in \mathbb{N}$ and $D_1 + D_2 = D$. Let $A,\ B \in \mathcal{L}(\mathcal{O})$. By the above, they have the form $A = U^*(A_1 \oplus A_2)U$, $B = U^*(B_1 \oplus B_2)U$ with $A_i,\ B_i \in \mathcal{M}_{D_i}$, $i \in [2]$. Hence,

$$\begin{aligned}\det[x\mathbb{1} - A - zB] &= \det[x\mathbb{1}_{D_1} - A_1 - zB_1]\det[x\mathbb{1}_{D_2} - A_2 - zB_2] \\ &= p_1(x,z)p_2(x,z)\end{aligned}$$

with p_1, p_2 real polynomials of degree strictly less than D. Therefore, we know that $\mathrm{C}^*(\mathcal{O}) \subsetneq \mathcal{M}_D$ implies that $\det[x\mathbb{1} - A - zB]$ for $A,\ B \in \mathcal{L}(\mathcal{O})$ is not irreducible over the reals. We could suppose that also the converse holds, namely that for $A,\ B$ such that the above determinant is a reducible polynomial, $\mathrm{C}^*(\{A, B\})$ must be a proper subalgebra of $\mathcal{M}_D$ (note that the C^*-algebra does not depend on which generators were used as long as $\mathcal{L}(\{A, B\}) = \mathcal{L}(\mathcal{O})$). Alas, this is not the case, as the following counterexample shows:

Example 6.13. Let $p \in \mathbf{H}^3(3)$ be defined as

$$p(x,y,z) := (x - 1/2y)(x^2 - y^2 - z^2).$$

This is clearly reducible over the reals. However, p admits a monic determinantal representation

$$p(x,y,z) = \det[x\mathbb{1} + yA + zB]$$

such that $\mathrm{C}^*(\{A, B\}) = \mathcal{M}_3$.

Proof. By unitary invariance of the determinant, we can assume that A is diagonal. It is easy to verify that $p(x, y, z)$ is hyperbolic with respect to $(1, 0, 0)$ and that $p(1, 0, 0) = 1$, such that we can choose A to be real (cf. [21]). We can therefore compare coefficients directly and solve a system of equations for the matrix coefficients which is reasonably small. One possible determinantal representation is given by

$$A = \begin{bmatrix} -1 & & \\ & -1/2 & \\ & & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0 & -1/2 & 0 \\ -1/2 & 0 & -\sqrt{3}/2 \\ 0 & -\sqrt{3}/2 & 0 \end{bmatrix}.$$

The matrix B has eigenvalues $-1, 1, 0$ with corresponding eigenvectors $(1, 2, \sqrt{3})$, $(1, -2, \sqrt{3})$ and $(-\sqrt{3}, 0, 1)$. Note that both matrices have non-degenerate spectrum. By Burnside's theorem (cf. [22] for the exact statement and a simple proof), the generators of any proper subalgebra of $\mathcal{M}_D$ must have a common invariant subspace other than 0 or $\mathbb{C}^D$. Since the eigenvectors of A and B are pairwise linearly independent, there are no common invariant subspaces of dimension one. As only the eigenvector of B corresponding to eigenvalue 0 is in any of the two-dimensional subspaces spanned by the pairs of eigenvectors of A, there are no common two-dimensional invariant subspaces, either. By Burnside's theorem thus $\mathrm{C}^*(\{A, B\}) = \mathcal{M}_3$. □

Note that from [33], we know that the determinantal representation of (irreducible smooth) algebraic curves of degree 2 is unique up to equivalence, whereas in degree 3, there are infinitely many (not necessarily real symmetric) determinantal representations. Hence, it was natural to look for counterexamples of this degree.

7. Upper Bounds

7.1. Compression to Maximal Block Size

We will show now that using classical side information we can at least compress to the dimension of the largest block. Note that the proof of the lemma yields explicit coding and decoding channels.

Theorem 7.1. (Upper bound on the compression dimension) *Let $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$ be such that*

$$\mathrm{C}^*(\mathcal{O}) = U^* \left(\bigoplus_{i=1}^{s} \mathcal{M}_{D_i} \otimes \mathbb{1}_{m_i} \right) U \tag{10}$$

where $\sum_{i=1}^{s} D_i m_i = D$ and $U \in \mathcal{U}(D)$. Then, $\max_{j \in [s]} D_j$ is an upper bound on the minimal compression dimension.

Proof. By Proposition 6.4, we can assume that

$$\mathrm{C}^*(\mathcal{O}) = \bigoplus_{i=1}^{s} \mathcal{M}_{D_i}$$

with $\sum_{i=1}^{s} D_i = D$. Without loss of generality, let $D_1 \geq D_j \, \forall j \in [s]$. Let $V_j : \mathbb{C}^{D_j} \hookrightarrow \mathbb{C}^D$ be an isometry such that $V_j V_j^* = P_j$ is the projection onto the jth block. In the same vein, let $W_j : \mathbb{C}^{D_j} \hookrightarrow \mathbb{C}^{D_1}$ be an isometry such that $W_j W_j^* = Q_j$ is the projection onto M_{D_j}, i.e. $Q_j = \mathbb{1}_{D_j} \oplus 0$. We define $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_{D_1} \otimes \mathbb{C}^s$ as

$$\mathcal{C}(\rho) = \sum_{j=1}^{s} W_j V_j^* \rho V_j W_j^* \otimes |j\rangle\langle j| \,, \tag{11}$$

where $\{ \, |j\rangle \, \}_{j=1}^{s}$ is an orthonormal basis of $\mathbb{C}^s$. This map is obviously completely positive, since it is given in Kraus decomposition. It is also trace preserving, because

$$\begin{aligned} \mathrm{Tr}\,(\mathcal{C}(\rho)) &= \sum_{j=1}^{s} \mathrm{Tr}\,\big(W_j V_j^* \rho V_j W_j^* \otimes |j\rangle\langle j|\big) \\ &= \sum_{j=1}^{s} \mathrm{Tr}\,(P_j \rho) = \mathrm{Tr}\,(\rho)\,. \end{aligned}$$

For $\mathcal{D}$, it is easier to define the dual map. We will need the following maps $\mathcal{R}_j : \mathcal{M}_{D_j} \to \mathcal{M}_{D_1}$ given by

$$A \mapsto A \oplus \mathrm{Tr}\,(A\eta_j)\, \mathbb{1}_{D_1 - D_j} \quad \eta_j \in \mathcal{S}(\mathbb{C}^{D_j}).$$

The choice of η_j is somewhat arbitrary and is needed to ensure linearity. This map is completely positive, since it is a composition of $A \mapsto A \otimes \mathbb{1}_2$ and the direct sum of the identity map and the map $A \mapsto \mathrm{Tr}\,(A\eta_j)\, \mathbb{1}_{D_1 - D_j}$, all of which are completely positive and unital. With this, we define the dual channel $\mathcal{D}^* : \mathcal{M}_D \to \mathcal{M}_{D_1} \otimes \mathbb{C}^s$ as

$$A \mapsto \sum_{j=1}^{s} \mathcal{R}_j(V_j^* A V_j) \otimes |j\rangle\langle j|\,. \tag{12}$$

This map is unital since $\mathcal{R}_j$ is. To show correctness, we need to verify that $\mathrm{Tr}\,(\rho E) = \mathrm{Tr}\,(\mathcal{C}(\rho)\mathcal{D}^*(E))$ for all $\rho \in \mathcal{S}(\mathbb{C}^D)$, $E \in \mathcal{O}$. We compute for such ρ, E

$$\begin{aligned} \mathrm{Tr}\,(\mathcal{C}(\rho)\mathcal{D}^*(E)) &= \sum_{j=1}^{s} \mathrm{Tr}\,\big([V_j^* \rho V_j V_j^* E V_j] \oplus 0\big) \\ &= \sum_{j=1}^{s} \mathrm{Tr}\,(\rho P_j E P_j)\,, \end{aligned}$$

where we used $W_j V_j^* \rho V_j W_j^* = V_j^* \rho V_j \oplus 0$ in the first equation. The last line is equal to $\mathrm{Tr}\,(\rho E)$ since E is block diagonal.

To obtain compression and decompression maps for the original algebra $\mathrm{C}^*(\mathcal{O})$ in Eq. (10), we can use the $*$-isomorphism given in Proposition 6.4 and define $\widetilde{\mathcal{D}}^* := \mathcal{D}^* \circ \pi$, $\widetilde{\mathcal{C}}^* := \pi^{-1} \circ \mathcal{C}^*$, where $\mathcal{C}$, $\mathcal{D}$ are the maps constructed above. $\square$

We have given an explicit way to compress a subalgebra to the size of its largest block. So far, it is, however, unclear if compression to the largest block is indeed the best we can do or if d can be chosen smaller. Before we will pursue this, we will apply the above theorem in two concrete situations. First, we prove that for $\dim \mathcal{L}(\mathcal{O}) < 3$, the set of effect operators $\mathcal{O}$ is trivially compressible.

Proposition 7.2. (Compression of a single binary measurement) *Let $\mathcal{O} = \{ E, \mathbb{1} - E \}$ be a set of effect operators, where $E \in \mathcal{E}(\mathbb{C}^D)$. Then, the compression dimension is* 1.

Proof. As E is an effect operator, we can diagonalize E to show that $\mathrm{C}^*(\mathcal{O})$ is $*$-isomorphic to $\bigoplus_{i=1}^{s} \mathbb{C}$, $s \leq D$. The assertion follows from Theorem 7.1. □

We will now continue to use Theorem 7.1 to discuss the important example of two von Neumann measurements with two outcomes each.

7.2. Compressibility for Two Binary Von Neumann Measurements

We have shown that compressibility strongly depends on the algebra generated by the desired effect operators. In this section, we will show that in the case of two bipartite projective measurements, we can compress to qubits ($d = 2$) using classical side information. The idea is that two projections generate an algebra which has a block structure of 2×2-matrices. This will use a finite-dimensional version of Halmos' two projections theorem (cf. [13, Theorem 2], [6, Theorem 1.1]).

Suppose we are given two orthogonal projections P and Q acting on a $\mathbb{C}^D$ with $\operatorname{Ran} P = M$, $\operatorname{Ran} Q = N$. Then, $\mathbb{C}^D$ can be decomposed as

$$\mathbb{C}^D = (M \cap N) \oplus (M \cap N^\perp) \oplus (M^\perp \cap N) \oplus (M^\perp \cap N^\perp) \oplus M_0 \oplus M_1.$$

The spaces M_0 and M_1 are defined through the decomposition of $\mathbb{C}^D$ into M and $M^\perp$,

$$M = (M \cap N) \oplus (M \cap N^\perp) \oplus M_0$$
$$M^\perp = (M^\perp \cap N) \oplus (M^\perp \cap N^\perp) \oplus M_1,$$

and their dimensions have to agree in order for them to be non-empty. We will use the abbreviation

$$(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \alpha_1 \mathbb{1}_{M \cap N} \oplus \alpha_2 \mathbb{1}_{M \cap N^\perp} \oplus \alpha_3 \mathbb{1}_{M^\perp \cap N} \oplus \alpha_4 \mathbb{1}_{M^\perp \cap N^\perp}.$$

If one of these subspaces is $\{ 0 \}$, we will just ignore this contribution irrespective of α_j. Note that this is the generic case. With this, we have the following theorem which is [6, Corollary 2.2]:

Lemma 7.3. *If one of the spaces M_0 and M_1 is non-trivial, then these two spaces have the same dimension $r \in \mathbb{N}$ and there exists a unitary matrix*

$V \in \mathcal{M}_D$ such that

$$VPV^* = (1,1,0,0) \oplus \operatorname{diag} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}_{j=1}^{r},$$

$$VQV^* = (1,0,1,0) \oplus \operatorname{diag} \begin{bmatrix} 1-\mu_j & \sqrt{\mu_j(1-\mu_j)} \\ \sqrt{\mu_j(1-\mu_j)} & \mu_j \end{bmatrix}_{j=1}^{r},$$

where $0 \leq \mu_j \leq 1$ for all $j \in [r]$.

This theorem is attributed to [36, Section 2], but similar questions concerning pairs of projections have already been studied by Camille Jordan in the nineteenth century. See [6, Remark 1.3] for a discussion of related results. Thus, the algebra generated by two projections and the identity operator consists essentially of block diagonal matrices with 2×2-blocks. For three projections, such a form can no longer be proven, since there are cases in which three projections generate the full matrix algebra (cf. concluding remarks of [6]). Hence, we cannot guarantee compression to be possible for more than two bipartite von Neumann measurements.

Proposition 7.4. (Compression of two binary projective measurements) *Let $\mathcal{O} = \{P, \mathbb{1} - P, Q, \mathbb{1} - Q\} \subset \mathcal{M}_D$ be a set of effect operators and P, Q two distinct orthogonal projections. Then, the compression dimension for the set of these effect operators is upper-bounded by $d = 2$.*

Proof. Let P, $Q \in \mathcal{M}_D$ be two distinct orthogonal projections. Lemma 7.3 provides a unitary operator V such that

$$Q = V^*((1,1,0,0) \oplus Q_5 \oplus \ldots \oplus Q_k)V,$$
$$P = V^*((1,0,1,0) \oplus P_5 \oplus \ldots \oplus P_k)V,$$

where Q_i, $P_j \in \mathcal{M}_2$ for i, $j \in \{5, \ldots, k\}$. We are therefore in the situation of Theorem 7.1 with $D_i = 1$ for $i \in [4]$ and $D_j = 2$ for $j \in [k] \setminus [4]$ (identifying $\{(\alpha_1, \ldots, \alpha_4) : \alpha_i \in \mathbb{C}, i \in [4]\}$ with $\mathbb{C}^4$, thus eliminating redundancies). Theorem 7.1 gives us a coding map $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_2 \otimes \mathbb{C}^k$ and a decoding map $\mathcal{D} : \mathcal{M}_2 \otimes \mathbb{C}^k \to \mathcal{M}_D$ which satisfies the constraints in Eq. (3). □

8. Computing the Compression Dimension

Hitherto, we have only seen that the dimension of the largest block is attainable for compression (Theorem 7.1), whereas the dimension of the smallest block is a lower bound on the compression dimension (Theorem 6.1), which is not necessarily attainable. In this section, we give an algorithm which allows us to compute the minimal dimension we can compress to using classical side information. We will assume that the operators in $\mathcal{O}$ are already given in block diagonal form. Whether two given Hermitian operators have a common block diagonal structure can be checked using the algorithm in [11, Section 4]. Algorithms to bring a finite-dimensional C*-algebra into block diagonal form can be found, e.g., in [24]. We analyze the latter algorithm in Appendix E. Assume that we are given a set of Hermitian operators $\mathcal{O}$. By Proposition 6.4, we can assume that $\mathrm{C}^*(\mathcal{O}) = \bigoplus_{i=1}^s \mathcal{M}_{D_i}$ with $\sum_{i=1}^s D_i = D$. The question of finding the minimal dimension which we can compress to amounts to determining which blocks are redundant, as will be proven below (cf. Theorem 8.2). Let us define what we mean by redundant.

Definition 8.1. (*Redundancy*) Let $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$ be such that

$$\mathrm{C}^*(\mathcal{O}) = \bigoplus_{i=1}^s \mathcal{M}_{D_i}$$

with $\sum_{i=1}^s D_i = D$. We will call the ith block *redundant* if the compression dimension is smaller than D_i.

We claim that checking redundancy can be phrased as an interpolation problem. Let D_1 be a block of maximal dimension (it does not matter which one we take if several of them have the same dimension, since all are redundant if one of them is). Then, we ask whether there is a completely positive map $\Phi_1 : \mathcal{M}_D \to \mathcal{M}_{D_1}$ such that

$$\Phi_1\left(\begin{bmatrix} 0 & & & \\ & E^2 & & \\ & & \ddots & \\ & & & E^s \end{bmatrix}\right) = E^1 \quad \forall E \in \mathcal{O} \cup \{ \mathbb{1} \},$$

where $E^i \in \mathcal{M}_{D_i}$ for all $i \in [s]$. This is a problem which can be solved using a semidefinite program (SDP, cf. [7]) as shown in [14]. Without loss of generality, we can assume that $\mathcal{O} = \{ \mathbb{1}, E_2, \ldots, E_k \}$, $k \in \mathbb{N}$. If this is not the case, substitute $\mathcal{O}$ by a set of linearly independent Hermitian operators including the identity where $k = \dim \mathcal{L}(\mathcal{O})$. The SDP is the following:

$$\text{Minimize} \quad \sum_{i=1}^k \operatorname{Tr}\left((E_i^1)^T H_i\right)$$

$$\text{Subject to} \quad \sum_{i=1}^k \begin{bmatrix} 0 & & & \\ & E_i^2 & & \\ & & \ddots & \\ & & & E_i^s \end{bmatrix} \otimes H_i \geq 0 \quad H_i \in \mathcal{M}_{D_1}, \quad \forall i \in [k].$$

We will refer to an algorithm which solves this problem as InterpolationSDP with parameters $E_1, \ldots, E_k$ and j, where j denotes the block which appears in the minimization (in the above case $j = 1$). This SDP has either $-\infty$ or 0 as solution; the latter solution confirms that there is a Φ_1 as specified above. If such a Φ_1 cannot be found, D_1 is the minimal dimension we can compress to, otherwise we proceed to the next block. Then, we can repeat the procedure with the remaining blocks until we either encounter one block which is not redundant or we are left with only one block. This algorithm is formalized in pseudocode in Algorithm 1.

To see that there are actually Hermitian operators which give rise to redundant blocks such that the dimension we can compress to is strictly less

Algorithm 1 Compute minimal compression dimension

Require: List of $E_i = \{ E_i^1, \dots, E_i^s \}, E_i^j \in \mathcal{M}_{D_j}, i \in [k], j \in [s]$, where $E_1 = \mathbb{1}$; List $\{ D_1, \dots, D_s \}$ s. t. $D_1 \geq \dots \geq D_s$.

$j := 1$
Dmax := D_1
while $j < s$ **do**
  $h \leftarrow$ InterpolationSDP$(E_1, \dots, E_k)(j)$ ▷ 0 if block redundant, $-\infty$ otherwise
  $j \leftarrow j + 1$
  if h = 0 **then**
    Dmax $\leftarrow D_j$
    $E_i^j \leftarrow 0 \, \forall i \in [k]$ ▷ Set largest nonzero block to zero
  else
    $j \leftarrow s$ ▷ Terminates computation
  end if
end while
return Dmax ▷ Dimension of largest non-redundant block

than the maximal block dimension, we refer to the end of this section. We proceed with a proof that the dimension computed by Algorithm 1 is indeed the minimal one.

Theorem 8.2. (Correctness of the algorithm) *The dimension computed by Algorithm 1 is the compression dimension.*

Proof. We proceed in three steps. First, we see that d is the dimension of the largest block on which an optimal compression map acts as the identity. Then, we see that all larger blocks can be interpolated. Last, we see that no other blocks have a solution to the interpolation problem. Assume that we have found a map $\mathcal{T}^*$ such that d is the compression dimension. Then, this map has to be the identity on some blocks by Corollary 6.5. Let $\mathcal{I} \subset [s]$ be the index set of the blocks for which this is the case. Again by Corollary 6.5, we can conclude that $d \geq \max_{i \in \mathcal{I}} D_i =: D_{\max}$. We have to show that $D_{\max}$ can be attained to complete the first step, whereby $d = D_{\max}$. Since $\mathcal{T}^*$ and $\mathcal{T}_\infty^*$ have the same fixed-point set, we can use $\mathcal{T}_\infty^*$ to construct another compression map. Note that all blocks with $i \notin \mathcal{I}$ lie in the kernel of $\mathcal{T}_\infty^*$ by Corollary 6.5. Thus, for all $j \in [s] \setminus \mathcal{I}$ there must be a completely positive map $\Phi_j : \mathcal{M}_D \to \mathcal{M}_{D_j}$ such that

$$\Phi_j \left(\bigoplus_{i=1}^{s} \chi_{\mathcal{I}}(i) E^i \right) = E^j \quad \forall E \in \mathcal{O} \cup \{ \mathbb{1} \}, \tag{13}$$

where $\chi_{\mathcal{I}}$ is the indicator function of the set $\mathcal{I}$. Hence, we can give the following compression scheme which attains $D_{\max}$. For the decompression map $\mathcal{D}$, we can almost use the map given in the proof of Theorem 7.1 with $d = D_{\max}$, but requiring the sum in Eq. (12) to run only over $\mathcal{I}$. Without loss of generality, we can assume that $\mathcal{I}$ is the set of the first $|\mathcal{I}|$ entries, such that $n = |\mathcal{I}|$ is the dimension needed for the classical side information. Let $V_i : \mathbb{C}^{D_i} \hookrightarrow \mathbb{C}^D$ be an isometry such that $V_i V_i^* = P_i$ is the projection onto the ith block $\forall i \in [s]$ and $W_j : \mathbb{C}^{D_j} \hookrightarrow \mathbb{C}^{D_{\max}}$ an isometry such that $W_j W_j^* = Q_j$ is the projection onto $\mathcal{M}_{D_j} \, \forall j \in [s]$. Then, we can define the dual compression map

$\mathcal{C}^* : \mathcal{M}_{D_{\max}} \otimes \mathbb{C}^n \to \mathcal{M}_D$ as

$$\mathcal{C}^*(A) := \mathcal{T}_\infty^* \left(\sum_{j \in \mathcal{I}} V_j \left(W_j^* \otimes \langle j| \right) A \left(W_j \otimes |j\rangle \right) V_j^* \right).$$

This map can easily be seen to be completely positive, because $\mathcal{T}_\infty^*$ is. Correctness follows from the construction in Theorem 7.1 since the missing blocks are all in the kernel of $\mathcal{T}_\infty^*$. The same holds for unitality. Hence, $d = D_{\max}$ since otherwise the map just defined would allow for an even better decompression, which contradicts that d is minimal. This shows that all redundant blocks have a solution to the interpolation problem and completes the second step.

If we could find a set $\mathcal{J} \subset [s]$ and completely positive maps such that Eq. (13) holds for this $\mathcal{J}$ instead of $\mathcal{I}$ and such that $D'_{\max} = \max_{j \in J} D_j < d$, we could construct a dual channel attaining better compression. To see this, define a map $\mathcal{R}^*$ which is Φ_j on blocks $j \in [s] \setminus \mathcal{J}$ and the identity on all other blocks. Then, we could substitute $\mathcal{J}$ for $\mathcal{I}$, $D'_{\max}$ for $D_{\max}$ and $\mathcal{R}^*$ for $\mathcal{T}_\infty^*$ in the above construction to obtain a map with compression dimension $D'_{\max}$, which contradicts minimality of d. This shows that a block admits a solution to the interpolation problem if and only if it is redundant and completes the last step. □

For a discussion of the complexity of the proposed algorithm, we refer to Appendix E. Instead, we will show now that unless the algebra has a very specific structure, any block can be redundant. We start with two lemmas investigating the matrix $*$-algebra generated by the image of a unital CP map.

Lemma 8.3. *Let $\mathcal{A}$ be a unital matrix $*$-algebra that contains $\mathcal{M}_2$ or $\mathbb{C}^3$ as a subalgebra, which we denote by $\mathcal{A}'$. For every unital matrix $*$-algebra $\mathcal{B}$, there is a unital CP map $\Phi : \mathcal{A} \to \mathcal{B}$ and positive rank-one elements $A_1, A_2, A_3 \in \mathcal{A}'$ such that*

1. $\mathrm{C}^*(\{A_1, A_2, A_3\}) = \mathcal{A}'$,
2. $\mathrm{C}^*(\{\Phi(A_1), \Phi(A_2), \Phi(A_3)\}) = \mathcal{B}$, *and*
3. *each* $\Phi(A_i)$, $i \in [3]$ *is positive definite.*

Proof. Assume that $\mathcal{B}$ is not $*$-isomorphic to a subalgebra of $\mathcal{A}'$ (otherwise the statement is trivial). Without loss of generality, let $\mathcal{B} = \bigoplus_{i=1}^s \mathcal{M}_{D_i}$, as $\mathcal{B}$ is $*$-isomorphic to such an algebra and any $*$-isomorphism is a unital CP map. Choose X_i, $Y_i \in \mathcal{M}_{D_i}$ such that $0 < X_i, Y_i < \mathbb{1}/2$, $\mathrm{C}^*(\{ X_i, Y_i, \mathbb{1} \}) = \mathcal{M}_{D_i}$ and such that $B_1 := \bigoplus_{i=1}^s X_i$ and $B_2 := \bigoplus_{i=1}^s Y_i$ both have non-degenerate spectrum. This is possible, because invoking Lemma 6.6 lets us choose generic Hermitian $\tilde{X}_i$, $\tilde{Y}_i$ such that these generate the respective algebras. The non-singular transformation $X_i \mapsto \lambda_{x,i}(X_i + \mu_{x,i}\mathbb{1})$ with $\lambda_{x,i}$, $\mu_{x,i} > 0$ and $Y_i \mapsto \lambda_{y,i}(Y_i + \mu_{y,i}\mathbb{1})$ with $\lambda_{y,i}$, $\mu_{y,i} > 0$ allow us to choose the elements positive definite and not too large with an appropriate choice of parameters.

Now, set $B_3 := \mathbb{1} - B_1 - B_2$. Note that $B_3 > 0$ and $\mathrm{C}^*(\{ B_1, B_2, \mathbb{1} \}) = \mathcal{B}$ hold by the above construction, which in particular implies that the B_i are linearly independent if $\mathcal{B}$ is non-commutative. Choose a set of linearly independent, positive rank-one operators A_i, $i \in [3]$ such that $\sum_{j=1}^{3} A_j = \mathbb{1}$ and $\mathrm{Tr}\,(A_j) = c$, $c > 0$. For $\mathcal{A}' = \mathbb{C}^3$, we can pick an ONB and for $\mathcal{A}' = \mathcal{M}_2$ the operators $A_i = 2/3\,|a_i\rangle\langle a_i|$ with $|a_i\rangle = \cos\theta_i\,|0\rangle + \sin\theta_i\,|1\rangle$ and $\theta_i = i2\pi/3$, $i \in [3]$. Note that both these choices generate $\mathcal{A}'$ as a C*-algebra.

We define $\tilde{\Phi} : \mathcal{A}' \to \mathcal{B}$ as

$$\tilde{\Phi}(Z) = \frac{1}{c}\sum_{i=1}^{3} \mathrm{Tr}\,(A_i Z)\, B_j.$$

This is clearly a unital CP map. If $\mathcal{B}$ is commutative, then it follows that $D_i = 1$ for all $i \in [s]$ and it is easy to see that the assertion of the lemma holds if we extend $\tilde{\Phi}$ to a map Φ on $\mathcal{A}$. If $\mathcal{B}$ is not commutative, we claim that $\dim(\mathrm{Ran}\;(\tilde{\Phi})) = 3$, even if the preimage is restricted to the linear span of the A_i's. This can be seen as follows. Assume that $\tilde{\Phi}(A_1) = \lambda\tilde{\Phi}(A_2) + \mu\tilde{\Phi}(A_3)$ for some λ, $\mu \in \mathbb{C}$. By linear independence of the B_j, this implies

$$\mathrm{Tr}\,(A_j(A_1 - \lambda A_2 - \mu A_3)) = 0 \quad \forall j \in [3].$$

The above implies, however, that $\mathrm{Tr}\left(|A_1 - \lambda A_2 - \mu A_3|^2\right) = 0$ and hence $A_1 - \lambda A_2 - \mu A_3 = 0$. This is a contradiction due to the linear independence of the A_i, which proves the claim that $\dim(\mathrm{Ran}\;(\tilde{\Phi})) = 3$.

Therefore, $\tilde{\Phi}$ maps span $\{ A_1, A_2, A_3 \}$ onto span $\{ B_1, B_2, \mathbb{1} \}$. Let Φ be the extension of $\tilde{\Phi}$ to $\mathcal{A}$. So we have finally proven claim (2) of the Lemma since

$$\mathcal{B} \supset \mathrm{C}^*\,(\Phi(\mathcal{A})) \supset \mathrm{C}^*\,(\{\Phi(A_1), \Phi(A_2), \Phi(A_3)) = \mathrm{C}^*\,(\{\,\mathbb{1}, B_1, B_2\,\}) = \mathcal{B}.$$

□

Lemma 8.4. *Let $\mathcal{A}$ be $*$-isomorphic to $\mathbb{C}^1$ or $\mathbb{C}^2$ and let $\mathcal{B}$ be non-commutative. Then, there is no unital CP map $\Phi : \mathcal{A} \to \mathcal{B}$ such that $\mathrm{C}^*\,(\Phi(\mathcal{A})) = \mathcal{B}$.*

Proof. If $\mathcal{A} = \mathbb{C}^1$, then $\mathrm{C}^*\,(\Phi(\mathcal{A}))$ is clearly commutative due to linearity of the map. The algebra is also commutative for $\mathcal{A} = \mathbb{C}^2$ due to unitality of Φ. □

As a corollary to these two lemmas, we can now investigate the possible redundancies of blocks in the representation of a finite-dimensional C*-algebra.

Corollary 8.5. (Tightness of algebraic bounds) *Let $\mathcal{A} = \bigoplus_{i=1}^{s} \mathcal{M}_{D_i}$.*

1. *If $\mathcal{A}$ contains three $\mathcal{M}_1$-blocks in its block structure, then there is a set of effect operators $\mathcal{W} \subset \mathcal{A}$ with compression dimension $d = 1$ and s.t. $\mathrm{C}^*\,(\mathcal{W}) = \mathcal{A}$.*
2. *If $\mathcal{A}$ contains $\mathcal{M}_\delta$ for some $\delta \geq 2$ in its block structure, then we can find a set of effect operators $\mathcal{W} \subset \mathcal{A}$ with compression dimension $d = \delta$ and such that $\mathrm{C}^*\,(\mathcal{W}) = \mathcal{A}$.*
3. *Let $\delta := \max_{i\in[s]} D_i$. If $\mathcal{A} \setminus \mathcal{M}_\delta$ does neither contain $\mathcal{M}_2$ nor $\mathbb{C}_3$ as subalgebra, then every $\mathcal{W}$ with $\mathrm{C}^*\,(\mathcal{W}) = \mathcal{A}$ has compression dimension δ.*

Proof. Claim (1) and claim (2) for $\delta = 2$ follow directly from Lemma 8.3 when choosing $\mathcal{A}'$ as the considered subalgebra $\mathcal{M}_2$ or $\mathbb{C}^3$ and $\mathcal{B} := \mathcal{A} \setminus \mathcal{A}'$: we define $\mathcal{W} = \{ A_i \oplus \Phi(A_i) : i \in [3] \}$ where A_i, Φ are as in the lemma. From the representation theory of matrix $*$-algebras, it follows that $\mathrm{C}^*(\mathcal{W}) = \mathcal{A}$, since we constructed the map such that $\Phi(A_i) > 0$. This excludes that the block generated by the A_i has multiplicity greater than 1 in $\mathrm{C}^*(\mathcal{W})$. The assertions then follows from Theorem 8.2.

Claim (2) with $\delta > 2$ is a simple consequence of the assertion for $\delta = 2$ by using an isometric embedding of $\mathcal{M}_2$ into $\mathcal{M}_\delta$. The set $\mathcal{W}$ is then obtained by taking the above (embedded) construction and adding sufficiently many elements of the form $A \oplus 0 \in \mathcal{M}_\delta \oplus (\mathcal{A} \setminus \mathcal{M}_\delta)$ so that the C^*-algebra that they generate is the entire block $\mathcal{M}_\delta \oplus 0$ (and not only the embedded $\mathcal{M}_2$ subalgebra). We can choose one of these elements such that all blocks of dimension less than δ have operator norm strictly less than the block of dimension δ. This guarantees that the compression dimension is not smaller than δ because of the contractivity of unital positive maps.

Claim (3) follows directly from Lemma 8.4 and Theorem 8.2. □

Note that we can extend the above corollary to general matrix $*$-algebras by invoking Proposition 6.4. We have therefore shown that unless we are in the last case of the corollary, our upper and lower bounds on the compression dimension are tight.

9. Generalizations

9.1. Measurements and Expectation Values

When we presented the setup in Sect. 4, we were interested in preserving the measurement statistics, i.e., the probabilities of each outcome of a fixed set of measurements. Subsequently, we realized that it makes no difference whether we assume $\mathcal{O} \subset \mathcal{E}(\mathbb{C}^D)$ or $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$, since only the operator system generated by $\mathcal{O}$ was important. This implies, however, that instead of (approximately) preserving the probabilities for each outcome, we could also only aim to preserve the expectation values of a set $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$ of observables and all results of this paper still apply. In particular, for generic $A, B \in \mathcal{M}_D^{\mathrm{herm}}$ such that $\mathrm{C}^*(\{ A, B \}) = \mathcal{M}_D$, Theorem 6.1 still states that those observables are incompressible.

Another possible modification of our setup would be to ask only for measurements E' on the compressed state $\mathcal{C}(\rho)$ which return the original statistics, but without imposing that $E' = \mathcal{D}^*(E)$ for some $E \in \mathcal{O}$ and some channel $\mathcal{D}$. This relaxation, however, can easily be seen to allow for more powerful compression in certain cases. Let $\mathcal{O} = \{ E_1, E_2, E_3 \}$, where the elements form a POVM and E_1 and E_2 are generic. In the alternative setup, we can see that it is possible to compress to $d = 1$ using

$$\mathcal{C}(\rho) = \sum_{i=1}^{3} \mathrm{Tr}\,(E_i \rho)\, |i\rangle\langle i|$$

and choosing $E_i' = |i\rangle\langle i|$. However, Theorem 6.1 implies that $\mathcal{O}$ is incompressible in the original setup. The explanation for this difference is that there is no channel $\mathcal{D}$ which allows to map the elements of $\mathcal{O}$ to projections. Therefore, we see that this modification changes the problem significantly and we leave it for future work.

9.2. Positive and Schwarz Maps

The aim of this section is to explore how much we can relax the requirements on the compression and decompression channel. We still consider the setup of Sect. 4, but now we require $\mathcal{C}$, $\mathcal{D}$ only to be positive instead of completely positive. Since the argument at the beginning of the proof of Theorem 6.8 uses only that the maps involved are positive and trace preserving to apply the Russo–Dye theorem, the results of Theorem 6.8 carry over to this setting. Note that the results obtained in the algebraic setting do not carry over to arbitrary positive maps, since for Arveson's result it is important that the map is a Schwarz map (see remark before [38, Example 5.3]). Complete positivity, however, is not needed; a trace-preserving positive map whose dual is also a Schwarz map is enough. See Lemma A.3 for a proof that the Cesàro-mean of a Schwarz map is again a Schwarz map. Using Lemma D.1 instead of the corresponding well-known result for completely positive maps, we can extend Theorem 6.1 and Lemma 6.2 to Schwarz maps. From a physicist's perspective, exchanging $\mathcal{D}$ for a positive instead of a completely positive map can be interpreted as measuring different effect operators and inferring from them the statistics with respect to the original effect operators. Note that this is still less general than the modified setup discussed in Sect. 9.1. Since only completely positive maps are considered meaningful evolutions of a physical system, we have proven the theorems under these stronger conditions.

9.3. Completely Positive Maps on Operator Spaces

Most of our analysis has been carried out in the Heisenberg picture. The dual maps $\mathcal{T}^*$, $\mathcal{C}^*$ and $\mathcal{D}^*$ have been assumed to be completely positive on the full matrix algebra. However, one could argue that only complete positivity on the operator system $\mathcal{L}'(\mathcal{O})$ generated by $\mathcal{O}$ is required. By Arveson's extension theorem [27, Theorem 7.5] (or [27, Theorem 6.2], since we only need the finite-dimensional version) any completely positive map $\mathcal{T}^* : \mathcal{L}'(\mathcal{O}) \to \mathcal{M}_D$ can be extended to a completely positive map on $\mathcal{M}_D$. Hence, as long as we consider the setup relevant for quantum information, we need not distinguish whether $\mathcal{T}^*$ is completely positive on the full matrix algebra or on the operator system. For positive maps, this is no longer true in general (see [27, remark after Corollary 7.6]).

9.4. Finitely Many States

This section will briefly address the question of what can be proven if instead of all states $\mathcal{S}(\mathbb{C}^D)$ we only want to measure effect operators on a subset $\mathcal{S}_\mathcal{I} = \{ \rho_i : i \in \mathcal{I} \}$ for some states $\rho_i \in \mathcal{S}(\mathbb{C}^D)$ and some index set $\mathcal{I} \subset \mathbb{R}$. We note that the situation in Sect. 4 is not changed if $\mathcal{M}_D = \mathrm{span}_\mathbb{C} \{ \mathcal{S}_\mathcal{I} \}$, since again only the operator space spanned by the states matters, not the states themselves. We could thus exchange the set of all states for the set of pure states and our results in the above sections still hold.

Consider next the situation in which we allow only for states from $\mathcal{S}_{\mathcal{I}}$, but this time we want to measure a set of effect operators $\widetilde{\mathcal{O}}$ such that $\mathcal{L}(\widetilde{\mathcal{O}}) = \mathcal{M}_D^{\mathrm{herm}}$. For example, $\widetilde{\mathcal{O}} = \mathcal{E}(\mathbb{C}^D)$. This is the converse situation of what we considered before. Although we cannot apply the techniques used so far in this situation, this setup is actually significantly simpler. Let us adapt our definition of compressibility to this new setting.

Definition 9.1. (*Compression of states*) Let $\mathcal{S}_{\mathcal{I}}$ be a set of states in $\mathcal{M}_D$. The *compression dimension* of $\mathcal{S}_{\mathcal{I}}$ is the smallest $d \in \mathbb{N}$ for which there is an $n \in \mathbb{N}$, a CPTP map $\mathcal{C} : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$ and a CPTP map $\mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$ such that for their composition $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$, the constraints

$$\operatorname{Tr}(\rho A) = \operatorname{Tr}(\mathcal{T}(\rho) A) = \operatorname{Tr}(\rho \mathcal{T}^*(A)) \quad \forall \rho \in \mathcal{S}_{\mathcal{I}}, \quad \forall A \in \mathcal{M}_D^{\mathrm{herm}} \tag{14}$$

are satisfied. If the compression dimension equals D, $\mathcal{S}_{\mathcal{I}}$ is said to be *incompressible.*

Then, we can give a lower bound on the compression dimension in this setup.

Theorem 9.2. (Lower bound for states) *Let $\mathcal{S}_{\mathcal{I}}$ be a set of states and*

$$\mathrm{C}^*(\mathcal{S}_{\mathcal{I}}) = W \left[0 \oplus \bigoplus_{k=1}^{s} (\mathcal{M}_{D'_k} \otimes \mathbb{1}_{m'_k}) \right] W^*$$

with $D_0 + \sum_{k=1}^{s} m_k D'_k = D$ and $W \in \mathcal{U}(D)$. Then, the compression dimension is $\max_{k \in [s']} D'_k$. In particular, if $\mathrm{C}^*(\mathcal{S}_{\mathcal{I}}) = \mathcal{M}_D$, then $\mathcal{S}_{\mathcal{I}}$ is incompressible.

Before we can proof this, we need to prove a lemma.

Lemma 9.3. *Let $\mathcal{C}$, $\mathcal{D}^* : \mathcal{M}_D \to \mathcal{M}_d \otimes \mathbb{C}^n$ be linear positive maps and let $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$. If the fixed-point set of $\mathcal{T}$ has the form $\mathcal{M}_{D'} \otimes \rho$, $\rho \in \mathcal{S}(\mathbb{C}^m)$ such that $m \cdot D' = D$, then $d \geq D'$.*

Proof. We define $\iota_\rho : \mathcal{M}_{D'} \to \mathcal{M}_D$ by $\iota_\rho(A) = A \otimes \rho$ for all $A \in \mathcal{M}_{D'}$. This defines a completely positive map. We can also define $\widetilde{\mathcal{T}} : \mathcal{M}_{D'} \to \mathcal{M}_{D'}$

$$\widetilde{\mathcal{T}} := \operatorname{Tr}_{\mathbb{C}^m} \circ \mathcal{T} \circ \iota_\rho.$$

Here, we have made the identification $\mathbb{C}^D \simeq \mathbb{C}^{D'} \otimes \mathbb{C}^m$. By our assumption on the fixed point set of $\mathcal{T}$, we know that $\widetilde{\mathcal{T}}$ is the identity map. By the same argument as in the proof of Theorem 6.1, $d \geq D'$ follows from Lemma D.1. □

Proof of Theorem 9.2. Since we require Eq. (14) to hold, we see that $\mathcal{S}_{\mathcal{I}} \subset \mathcal{F}_{\mathcal{T}}$ for $\mathcal{F}_{\mathcal{T}}$ the fixed-point set of $\mathcal{T}$. For $\mathcal{T}$ a completely positive and trace-preserving map, it is known that the fixed-point set has the structure

$$\mathcal{F}_{\mathcal{T}} = U \left(0 \oplus \bigoplus_{k=1}^{s} \mathcal{M}_{D_k} \otimes \rho_k \right) U^* \tag{15}$$

with $U \in \mathcal{U}(D)$, $\rho_k \in \mathcal{S}(\mathbb{C}^{m_k})$ and $D = D_0 + \sum_{k=1}^{s} m_k D_k$, where D_0 is the dimension of the zero block. The ρ_k can be assumed to be diagonal (we can absorb the unitaries diagonalizing them into U), hence

$$\mathrm{C}^*\left(\mathcal{F}_{\mathcal{T}}\right) \subset U\left(0 \oplus \bigoplus_{k=1}^{s} \bigoplus_{j=1}^{m_k} \mathcal{M}_{D_k}\right) U^*.$$

These blocks can be considered independently. Let $V_k : \mathbb{C}^{D_k \cdot m_k} \hookrightarrow \mathbb{C}^D$ be an isometry such that $V_k V_k^*$ is the projection onto the kth block in the outer direct sum. Then, define $\mathcal{T}_k : \mathcal{M}_{m_k \cdot D_k} \to \mathcal{M}_{m_k \cdot D_k}$ by

$$\mathcal{T}_k := \Theta_{V_k} \circ \Theta_U \circ \mathcal{T} \circ \Theta_{U^*} \circ \Theta_{V_k^*}.$$

By construction, the fixed-point set of $\mathcal{T}_k$ is $\mathcal{M}_{D_k} \otimes \rho_k$. The map factorizes into $\mathcal{C}_k : \mathcal{M}_{m_k \cdot D_k} \to \mathcal{M}_d \otimes \mathbb{C}^n$ with

$$\mathcal{C}_k = \mathcal{C} \circ \Theta_{U^*} \circ \Theta_{V_k^*}$$

and $\mathcal{D}_k : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_{m_k \cdot D_k}$ with

$$\mathcal{D}_k = \Theta_{V_k} \circ \Theta_U \circ \mathcal{D}.$$

Lemma 9.3 shows that $d \geq D_k$. Since this holds for all $k \in [s]$, it follows that $d \geq \max_{k \in [s]} D_k$. Furthermore, for all $i \in [s']$ there must be a $k \in [s]$ such that $D_k \geq D_i'$, otherwise the structure of $\mathrm{C}^*\left(S_{\mathcal{I}}\right)$ could not be as assumed. Hence, also $d \geq \max_{k \in [s']} D_k'$. That this bound is achievable can be seen through a slight modification of the construction in Theorem 7.1. The maps can be chosen the same (assuming $\mathcal{S}_{\mathcal{I}}$ already to be in block diagonal form), but the isometries need to be chosen such that they respect the block structure of the states instead of the block structure of the operators in $\mathcal{O}$. Here, we treat the zero block as a direct sum of D_0 1-dimensional blocks, which do not affect the compression dimension. □

The same result also holds in a more general setting. For the theorem to hold, $\mathcal{T}$ needs not be completely positive. Since Eq. (15) is also valid if $\mathcal{T}$ is a positive, trace preserving, linear map such that the dual map satisfies the Schwarz inequality, the above theorem also holds for $\mathcal{C}, \mathcal{D}$ such that $\mathcal{T}$ satisfies these weaker conditions (see [38, Theorem 6.14]).

We have shown that unlike in the converse situation, there are no redundant blocks. Whether better bounds can be shown for finite sets of both states and effect operators beyond the results from [31,35] remains an open problem.

10. Several Copies of the Same State

In this section, we consider the following modifications compared to Sect. 4. We are given a set of Hermitian operators as before which we denote by $\mathcal{O}$. Instead of only one state, we consider finitely many copies of the same state (provided, e.g., by identical preparations). Hence, we consider a quantum channel $\mathcal{C} : \mathcal{M}_{mD} \to \mathcal{M}_d \otimes \mathbb{C}^n$. Compression in this setting is defined as follows:

Definition 10.1. (*Compression of observables using copies*) Let $\mathcal{O}$ be a set of Hermitian operators in $\mathcal{M}_D$ and $m \in \mathbb{N}$ the number of copies available. The *compression dimension* of $\mathcal{O}$ is the smallest $d \in \mathbb{N}$ for which there is an $n \in \mathbb{N}$, a CPTP map $\mathcal{C} : \mathcal{M}_{mD} \to \mathcal{M}_d \otimes \mathbb{C}^n$ and a CPTP map $\mathcal{D} : \mathcal{M}_d \otimes \mathbb{C}^n \to \mathcal{M}_D$ such that for their composition $\mathcal{T} = \mathcal{D} \circ \mathcal{C}$, the constraints

$$\operatorname{Tr}(\rho E) = \operatorname{Tr}\left(\mathcal{T}(\rho^{\otimes m})E\right) = \operatorname{Tr}\left(\rho^{\otimes m}\mathcal{T}^*(E)\right) \quad \forall \rho \in \mathcal{S}(\mathbb{C}^D), \quad \forall E \in \mathcal{O} \quad (16)$$

are satisfied. If the compression dimension equals D, $\mathcal{O}$ is said to be *incompressible.*

We prove now that taking copies of the state does not affect compressibility in the geometric picture.

Theorem 10.2. (Lower bounds on compression dimension for finitely many copies) *Let $\mathcal{O} \subset \mathcal{M}_D^{\mathrm{herm}}$ a set of Hermitian operators, E_1, $E_2 \in \mathcal{L}(\mathcal{O})$ and*

$$p(x, z) := \det[x\mathbb{1} - E_1 - zE_2].$$

Then, the smallest among the degrees of the irreducible factors of p is a lower bound on the compression dimension of $\mathcal{O}$. In particular, if p is irreducible over the reals, then $\mathcal{O}$ is incompressible.

Proof. Maximizing Eq. (16) over $\rho \in \mathcal{S}(\mathbb{C}^D)$, we obtain

$$\|A\|_\infty = \max_{\rho \in \mathcal{S}(\mathbb{C}^D)} \left|\operatorname{Tr}\left(\rho^{\otimes m}\mathcal{T}^*(A)\right)\right| \quad \forall A \in \mathcal{L}(\mathcal{O}). \quad (17)$$

The right-hand side of the above is clearly upper-bounded by $\|\mathcal{T}^*(A)\|_\infty$. Since $\mathcal{T}^*$ is unital, it is a contraction by the Russo–Dye theorem and

$$\|\mathcal{T}^*(A)\|_\infty \leq \|A\|_\infty$$

from which equality follows together with Eq. (17). Thus, we are able to apply the techniques from Sect. 6.2. Since $\mathcal{C}^*$ and $\mathcal{D}^*$ are unital as well, we obtain again

$$\|E_1 + tE_2\|_\infty = \|\mathcal{D}^*(E_1) + t\mathcal{D}^*(E_2)\|_\infty \quad \forall t \in \mathbb{R}.$$

The assertion then directly follows from Lemma 6.7. □

Note that in this case, we have to make use of the geometric arguments since we cannot infer from Eq. (16) that $E \in \mathcal{O}$ have to be fixed points of the dual channel.

Appendix A. Cesàro-mean and the Support Projection

This section exposes some facts which are needed in Sect. 6.1. In the following lemma, we collect some well-known facts about the Cesàro-mean (cf. [20] and [38, Chapter 6]). We recall the definition of the transfer matrix corresponding to the projection onto the fixed points of $\mathcal{R}$,

$$\hat{\mathcal{R}}_\infty = \sum_{\{\, k : \lambda_k = 1 \,\}} P_k, \quad (18)$$

where P_{λ_k} is the projection onto the (one dimensional) Jordan block associated with the eigenvalue λ_k of $\mathcal{R}$. $\mathcal{R}_\infty$ is the channel associated with this transfer matrix. Recall that the Cesàro-mean of $\mathcal{R}$, if it exists, is

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathcal{R}^n.$$

Lemma A.1. *Let $\mathcal{R}$ be a unital m-positive map on $\mathcal{M}_D$ with $m \in \mathbb{N}_0$. Then, $\mathcal{R}_\infty$ can be written as the Cesàro-mean of $\mathcal{R}$, it is unital and m-positive, $\mathcal{R}_\infty$ is idempotent, $\mathcal{R}_\infty \circ \mathcal{R} = \mathcal{R}_\infty = \mathcal{R} \circ \mathcal{R}_\infty$ and $\mathcal{R}_\infty$ has the same fixed-point set as $\mathcal{R}$.*

Proof. The spectral radius of $\mathcal{R}$ is equal to 1 by [38, Proposition 6.1]. Note furthermore that [38, Proposition 6.2] implies that the Jordan blocks belonging to eigenvalues of modulus 1 are one dimensional. By the same argument as in [38, Proposition 6.3], the first assertion then follows. From there, unitality and m-positivity directly follow. Looking at its transfer matrix, $\mathcal{R}_\infty$ is clearly idempotent, i.e.,

$$\mathcal{R}_\infty \circ \mathcal{R}_\infty = \mathcal{R}_\infty.$$

$\mathcal{R}_\infty = \mathcal{R} \circ \mathcal{R}_\infty$ holds since for every A in the range of $\mathcal{R}_\infty$, we know that $\mathcal{R}(A) = A$. Furthermore,

$$\mathcal{R}_\infty \circ \mathcal{R} = \mathcal{R}_\infty \tag{19}$$

follows by multiplication of the respective transfer matrices and using that Jordan blocks for eigenvalues of modulus 1 are one dimensional. Obviously, for $A \in \mathcal{M}_D$ such that $\mathcal{R}(A) = A$, also $\mathcal{R}_\infty(A) = A$ holds; therefore, the fixed-point sets are equal by the definition of $\mathcal{R}_\infty$. □

We also need the fact that the Cesàro-mean of a Schwarz map is again a Schwarz map. To prove this, we will need a lemma.

Lemma A.2. *Let $\mathcal{T}, \mathcal{R}$ be two Schwarz maps on $\mathcal{M}_D$. Then, $\mathcal{T} \circ \mathcal{R}$ is a Schwarz map as well. Furthermore $\lambda\mathcal{T} + (1-\lambda)\mathrm{id}$ is a Schwarz map for all $\lambda \in [0,1]$.*

Proof. Applying the Schwarz inequality twice, we obtain

$$(\mathcal{T} \circ \mathcal{R})(A)(\mathcal{T} \circ \mathcal{R})(A^*) \leq \mathcal{T}(\mathcal{R}(A)\mathcal{R}(A)^*) \leq \mathcal{T} \circ \mathcal{R}(AA^*),$$

where we used positivity of both $\mathcal{T}$ and $\mathcal{R}$. For the second assertion, we compute

$$\begin{aligned}&(\lambda\mathcal{T} + (1-\lambda)\mathrm{id})(AA^*) - (\lambda\mathcal{T} + (1-\lambda)\mathrm{id})(A)(\lambda\mathcal{T} + (1-\lambda)\mathrm{id})(A^*)\\ &\quad = \lambda(1-\lambda)\left[\mathcal{T}(AA^*) + AA^* - A\mathcal{T}(A^*) - \mathcal{T}(A)A^*\right]\\ &\quad + \lambda^2\left[\mathcal{T}(AA^*) - \mathcal{T}(A)\mathcal{T}(A^*)\right].\end{aligned}$$

We have to show that the above expression is positive. The second term is positive by the Schwarz inequality. We can reformulate the first term as

$$\mathcal{T}(AA^*) + AA^* - A\mathcal{T}(A^*) - \mathcal{T}(A)A^* = \begin{bmatrix}\mathbb{1} & A\end{bmatrix} \begin{bmatrix}\mathcal{T}(AA^*) & -\mathcal{T}(A)\\ -\mathcal{T}(A^*) & \mathbb{1}\end{bmatrix} \begin{bmatrix}\mathbb{1}\\ A^*\end{bmatrix}.$$

The operator matrix can be shown to be positive semidefinite using the Schur complement, so the right-hand side of the above is positive as well. □

Lemma A.3. *Let $\mathcal{T} : \mathcal{M}_D \to \mathcal{M}_D$ be a Schwarz map. Then, the Cesàro-mean of $\mathcal{T}$ is a Schwarz map as well.*

Proof. The statement for

$$\frac{1}{N}\sum_{n=1}^{N}\mathcal{T}^n$$

follows by induction. Using Lemma A.2, we infer that

$$\frac{1}{2}\left[\mathcal{T}^2+\mathcal{T}\right]=\frac{1}{2}\left[\mathcal{T}+\mathrm{id}\right]\circ\mathcal{T}$$

is a Schwarz map. By the same lemma, it follows that

$$\frac{1}{N+1}\sum_{n=1}^{N+1}\mathcal{T}^n=\left(\frac{1}{N+1}\mathrm{id}+\left(1-\frac{1}{N+1}\right)\frac{1}{N}\sum_{n=1}^{N}\mathcal{T}^n\right)\circ\mathcal{T}$$

is a Schwarz map using the induction hypothesis. The statement follows taking the limit $N \to \infty$. □

The rest of this section focuses on the support projection. We are only concerned with matrix algebras, so we assume $\mathcal{A} \subset \mathcal{M}_D$ to be a finite-dimensional C*-algebra and let $\mathcal{A}_+$ denote the positive elements in this algebra. For the case of von Neumann algebras of arbitrary dimensions, see, e.g., [9] or [5, III.2.2.25]. The support projection of a Schwarz map is not to be confused with the support projection of its transfer matrix. They are in general not the same. First we define the set

$$\mathcal{N}=\{\,A\in\mathcal{A}:\mathcal{R}(A^*A)=0\,\}$$

for some Schwarz map $\mathcal{R} : \mathcal{A} \to \mathcal{A}$. This set contains projections, as we shall see. Using the spectral decomposition, we may write $A^*A=\sum_{i=1}^{n}\sigma_i^2P_i$, where $\sigma_i > 0$, $i \in [n]$ are the distinct singular values of A and $P_i \in \mathcal{A}$ the corresponding spectral projections. Then,

$$\mathcal{R}(A^*A)=\sum_{i=1}^{n}\sigma_i^2\mathcal{R}(P_i).$$

The sum is zero if and only if $\mathcal{R}(P_i) = 0$ for all $i \in [n]$ and thus also the support projection of A, $V_A=\sum_{i=1}^{n}P_i$, is in $\mathcal{N}$. By the lattice structure of the set of projections, there is a unique maximal projection in $\mathcal{N}$. We will denote this projection by Q. Using the existence of such a Q, we get

$$\|\mathcal{R}(AQ)\|_\infty^2\le\|\mathcal{R}(QA^*AQ)\|_\infty=0\quad\forall A\in\mathcal{A},$$
$$\|\mathcal{R}(QA)\|_\infty^2\le\|\mathcal{R}(QAA^*Q)\|_\infty=0\quad\forall A\in\mathcal{A},$$

where we used the C*-property, the fact that positive maps are hermiticity preserving, the Schwarz inequality Eq. (2) and

$$\|\mathcal{R}(QBQ)\|_\infty\le\|B\|_\infty\|\mathcal{R}(Q)\|_\infty=0\quad\forall B\in\mathcal{A}_+.$$

This implies

$$\mathcal{R}(AQ)=\mathcal{R}(QA)=0\quad\forall A\in\mathcal{A}.$$

Hence, we can define the support projection as $P := \mathbb{1} - Q$. By the above property of Q, it fulfills

$$\mathcal{R}(A) = \mathcal{R}(PA) = \mathcal{R}(AP) = \mathcal{R}(PAP) \quad \forall A \in \mathcal{A}.$$

The following lemma collects the properties of the support projection which we use.

Lemma A.4. *Let $\mathcal{R} : \mathcal{A} \to \mathcal{A}$ be a Schwarz map. Then, for its support projection P, we have that*

$$\mathcal{R}(A) = \mathcal{R}(PA) = \mathcal{R}(AP) = \mathcal{R}(PAP) \quad \forall A \in \mathcal{A}$$

and $\mathcal{R}|_{P\mathcal{A}P}$ is faithful.

Proof. We only need to check the last claim, since we have already shown the rest. Being faithful on $P\mathcal{A}P$ means that the implication

$$\mathcal{R}(A) = 0 \to PAP = 0 \quad \forall A \in \mathcal{A}_+ \tag{20}$$

is true. This can be seen to hold as follows: Assume $\mathcal{R}(B) = 0$ for some $B \in \mathcal{A}_+$. Then, there is an $A \in \mathcal{A}$ such that $B = A^*A$, because B is positive. For this A, we know that $A \in \mathcal{N}$ and $QA^*AQ = A^*A$ by the definition of Q. Hence, also

$$B = A^*A = QA^*AQ = QBQ.$$

However, as $P = \mathbb{1} - Q$, this gives

$$PBP = 0$$

as claimed. □

Note that for general (non-positive) $B \in \mathcal{A}$, the implication in Eq. (20) is no longer true.

Appendix B. Existence of Both Irreducible and Hyperbolic Polynomials of Any Degree

The aim of this subsection is to show that there exist homogeneous polynomials of any degree which are both irreducible and hyperbolic. This is used in Sect. 6.2. This well-known fact from algebraic geometry will be proven here for convenience. It is clear that there are irreducible polynomials of any degree since $p(x, y, z) = x^d + y^d - z^d$ is irreducible for any $d \in \mathbb{N}$. Furthermore, it has been shown in [26] that the set of polynomials $p \in \mathbf{H}^d(n)$ hyperbolic with respect to a fixed point $e \in \mathbb{R}^n$ has non-empty interior in $\mathbf{H}^d(n)$ (see also [12, Theorem 2.1]). It is, however, not clear a priori that there are elements which fulfill both properties, since the $p(x, y, z)$ given above are not hyperbolic, as can be checked easily. The idea now is to prove that the set of reducible polynomials in $\mathbf{H}^d(n)$ does not contain any open subset, which would then mean that the set of irreducible and the set of hyperbolic elements in this space have non-empty intersection. The argument proceeds by dimension counting. We restrict to the case $n = 3$ for simplicity. Since we will be interested in normalized polynomials (i.e., $p(e) = 1$ for $e = (1, 0, 0)$), let $\mathbf{H}^d_N(3) \subset \mathbf{H}^d(3)$

be the affine subspace of such polynomials, where normalization decreases the dimension by one. We will identify $\mathbf{H}_N^d(3) \simeq \mathbb{R}^{\dim \mathbf{H}^d(3)-1}$, since we are only interested in the topology and measure on this affine space. Redoing the argument by Nuij shows that the set of normalized hyperbolic polynomials has non-empty interior in $\mathbf{H}_N^d(3)$ as well, since it basically only uses that the simple roots of an univariate polynomial depend continuously on the coefficients of the polynomial.

Lemma B.1. *The set of reducible elements over the reals in* $\mathbf{H}_N^d(3)$, $d \in \mathbb{N}$, $d > 2$ *does not contain any subset which is open in Euclidean topology. Moreover, this set has Lebesgue measure zero.*

Proof. Let $p \in \mathbf{H}_N^d(3)$ be a reducible element. Then, there are $q \in \mathbf{H}_N^k(3)$, $r \in \mathbf{H}_N^{d-k}(3)$, $k \in [d-1]$, such that $p = q \cdot r$. The fact that these polynomials can be chosen normalized follows since for $q(e) = c \neq 0$, necessarily $r(e) = 1/c$ by normalization of p and the polynomials can be multiplied by c and $1/c$, respectively, to obtain a decomposition into normalized elements. Hence, we define a mapping

$$\begin{aligned}\Phi : \mathbf{H}_N^k(3) \times \mathbf{H}_N^{d-k}(3) &\to \mathbf{H}_N^d(3)\\ (q, r) &\mapsto q \cdot r.\end{aligned}$$

$\mathbf{H}_N^k(3)$ is a semi-algebraic set (e.g., as $\mathcal{Z}(x_0 - 1)$, where x_0 is the coefficient belonging to x^k) with dimension $\binom{3+k-1}{k} - 1$ (by [3, Proposition 2.8.1], since dividing out the ideal $(x_0 - 1)$ decreases the dimension by one). Moreover, Φ is a semi-algebraic mapping (see [3, Definition 2.2.5]), since its graph can be expressed as

$$\Bigg\{(q, r, p) \in \mathbb{R}^{\dim(\mathbf{H}_N^k(3))} \times \mathbb{R}^{\dim(\mathbf{H}_N^{d-k}(3))} \times \mathbb{R}^{\dim(\mathbf{H}_N^d(3))} :$$

$$\sum_{\substack{i_l + j_l = m_l\\ l \in [3]}} q_i r_j - p_m = 0; m_1 \in [d-1],\ m_2, m_3 \in [d],\ m_1 + m_2 + m_3 = d\Bigg\}.$$

Here, $i, j, m \in \mathbb{N}_0^3$ are multi-indices such that $|i| = k$, $|j| = d-k$ and $|m| = d$. This is a finite collection of polynomial equalities which have to be fulfilled; thus, it is a semi-algebraic set. We have written q_i to be the coefficient belonging to $x^{i_1} y^{i_2} z^{i_3}$ of the polynomial q for clarity (same for p, r). Note that $q_{(k,0,0)} = r_{(d-k,0,0)} = p_{(d,0,0)} = 1$ has been fixed beforehand by normalization. By [3, Proposition 2.2.7], we know that the image of Φ for a fixed k is also a semi-algebraic set; likewise, this holds for the set of reducible elements in $\mathbf{H}_N^d(3)$, since it is a finite union of semi-algebraic sets. Now, we come back to the dimensions of the sets involved. By [3, Proposition 2.8.5 (ii)], the domain of Φ has dimension

$$\binom{3+k-1}{k} + \binom{3+d-k-1}{d-k} - 2.$$

Further,

$$\binom{3+d-1}{d} - \binom{3+k-1}{k} - \binom{3+d-k-1}{d-k} + 1 = (d-k)k,$$

which is greater equal $d-1$ for $k \in [d-1]$ and hence strictly positive for $d > 1$. Hence, the set of reducible elements has dimension strictly smaller than the dimension of $\mathbf{H}_N^d(3)$, $d > 1$, by [3, Proposition 2.8.5 (i)] and [3, Proposition 2.8.8]. This implies that it cannot contain any open $U \subset \mathbf{H}_N^d(3)$, since $\mathcal{I}(U) = \{ 0 \}$ necessarily, but there is at least one non-trivial polynomial vanishing on the set of reducible elements (otherwise this set would have full dimension by [3, Definition 2.8.1]), which would also vanish on any subset of these. By [19, Proposition A.1], any semi-algebraic $\mathcal{B} \subset \mathbb{R}^m$ of dimension less than m has zero m-dimensional Hausdorff measure and hence also zero Lebesgue measure, since those only differ by a constant factor on $\mathbb{R}^m$. □

Appendix C. Matrix Computations

This section contains some elementary computations needed to show the irreducibility of the polynomial in Proposition 6.12. Let the matrices $\tilde{A}(x)$, $B \in \mathcal{M}_D$, $D \geq 2$ be defined as follows:

$$\tilde{A}_{kl}(x) = \begin{cases} 0 & k < l \\ x & k = l \\ 1 & k > l \end{cases} \quad B_{kl} = \frac{1}{2}\begin{cases} \mathrm{i} & k < l \\ 0 & k = l \\ -\mathrm{i} & k > l \end{cases} \quad k, l \in [D].$$

The aim of this section is to compute $\det\left[\tilde{A}(x) + \epsilon B\right]$ up to first order in ϵ and to show that the first-order term has to vanish at $x = 0$ under some assumptions. On the way, we need to prove several lemmas which are of little interest in themselves. The first is the following sum formula which we will use several times:

Lemma C.1. *The following identity is true for $k \in \mathbb{N}_0$:*

$$\sum_{j=0}^{k} \left[-\frac{(x-1)^j}{x^{j+2}} \right] + \frac{1}{x} = \frac{(x-1)^{k+1}}{x^{k+2}}.$$

Proof. The statement follows by induction. □

We want to give the inverse of $\tilde{A}(x)$ provided $x \neq 0$.

Lemma C.2. *Assume that $x \neq 0$. Then, the inverse of $\tilde{A}(x)$ is given by*

$$C_{kl}(x) := \begin{cases} 0 & k < l \\ \frac{1}{x} & k = l \\ -\frac{(x-1)^{k-l-1}}{x^{k-l+1}} & k > l \end{cases} \quad k, l \in [D].$$

Proof. First, note that $\tilde{A}(x)$ is invertible for $x \neq 0$, since $\det\left[\tilde{A}(x)\right] = x^D$. We want to show that $C(x)\tilde{A}(x) = \mathbb{1}$. For now, let $C(x)\tilde{A}(x) =: F$. Note that F is lower triangular since $C(x)$ and $\tilde{A}(x)$ are. For $i \geq j$, we find

$$F_{ij} = \sum_{k=j}^{i} C_{ik}(x)\tilde{A}_{kj}(x).$$

For $i = j$, we have $F_{ii} = \frac{1}{x}x = 1$. For $i > j$, we obtain

$$\begin{aligned} F_{ij} &= -\frac{(x-1)^{i-j-1}}{x^{i-j+1}}x + \sum_{k=j+1}^{i-1}\left[-\frac{(x-1)^{i-k-1}}{x^{i-k+1}}\right] + \frac{1}{x} \\ &= -\frac{(x-1)^{i-j-1}}{x^{i-j}} + \sum_{k=0}^{i-j-2}\left[-\frac{(x-1)^k}{x^{k+2}}\right] + \frac{1}{x} \\ &= 0. \end{aligned}$$

The last equality follows by Lemma C.1. Hence, $F_{ij} = \delta_{ij}$. □

This can be used to compute the trace of $\tilde{A}^{-1}(x)B$.

Lemma C.3. *Let $x \neq 0$. Then,*

$$\mathrm{Tr}\left(\tilde{A}^{-1}(x)B\right) = \frac{\mathrm{i}}{2}\left[1 - \frac{D}{x} - \frac{(x-1)^D}{x^D}\right].$$

Proof. We first need to compute the diagonal entries of $\tilde{A}^{-1}(x)B$. Since $B_{ii} = 0\,\forall i \in [D]$ and $\tilde{A}^{-1}(x)$ is lower triangular, we have

$$\begin{aligned} \left[\tilde{A}^{-1}(x)B\right]_{jj} &= \sum_{k=1}^{j-1}\tilde{A}^{-1}_{jk}(x)B_{kj} = \frac{1}{2}\sum_{k=1}^{j-1}\left[-\frac{\mathrm{i}(x-1)^{j-k-1}}{x^{j-k+1}}\right] \\ &= \frac{1}{2}\sum_{k=0}^{j-2}\left[-\frac{\mathrm{i}(x-1)^k}{x^{k+2}}\right]. \end{aligned}$$

Taking the trace of this, we obtain

$$\begin{aligned} \mathrm{Tr}\left(\tilde{A}^{-1}(x)B\right) &= \frac{1}{2}\sum_{j=2}^{D}\sum_{k=0}^{j-2}\left[-\frac{\mathrm{i}(x-1)^k}{x^{k+2}}\right] \\ &= \frac{\mathrm{i}}{2}\sum_{j=0}^{D-2}\left[\frac{(x-1)^{j+1}}{x^{j+2}} - \frac{1}{x}\right] \\ &= \frac{\mathrm{i}}{2} - \frac{\mathrm{i}}{2}\frac{(x-1)^D}{x^D} - D\frac{\mathrm{i}}{2x}, \end{aligned}$$

where we have used Lemma C.1 in the second and third equality. □

Finally, we can use these computations to expand $\tilde{A}(x) + \epsilon B$ to first order in ϵ.

Lemma C.4. *We can expand the determinant of* $\tilde{A}(x) + \epsilon B$ *in terms of* ϵ *as*

$$\det\left[\tilde{A}(x) + \epsilon B\right] = x^D - \frac{\epsilon \mathrm{i}}{2}\left[Dx^{D-1} + (x-1)^D - x^D\right] + \mathcal{O}(|\epsilon|^2).$$

Proof. Let $x \neq 0$. Let $f : \mathbb{C} \to \mathbb{C}$, $f(\epsilon) = \det\left[\tilde{A}(x) + \epsilon B\right]$. By Taylor's theorem, we have

$$f(\epsilon) = f(0) + f'(0)\epsilon + \mathcal{O}(|\epsilon|^2).$$

By Jacobi's formula, it follows that

$$\frac{d}{dt}\det\left[\tilde{A}(x) + tB\right]\Big|_{t=0} = \mathrm{Tr}\left(\mathrm{adj}(\tilde{A}(x))B\right).$$

Using that $\tilde{A}(x)\mathrm{adj}(\tilde{A}(x)) = \det[\tilde{A}(x)]\mathbb{1}$ by the definition of the adjugate matrix, we infer

$$\det\left[\tilde{A}(x) + \epsilon B\right] = \det\left[\tilde{A}(x)\right] + \epsilon \det\left[\tilde{A}(x)\right]\mathrm{Tr}\left(\tilde{A}^{-1}(x)B\right) + \mathcal{O}(|\epsilon|^2).$$

By Lemma C.3 and $\det\left[\tilde{A}(x)\right] = x^D$, the statement follows for $x \neq 0$. The result extends to $x = 0$ by continuity. □

Lemma C.5. *Let* p, q, $r \in \mathbb{C}[x, y]$ *such that* $q(0, 0) = 0 = r(0, 0)$ *and* $p = q \cdot r$. *Let the expansion in* ϵ *of* $p(x, \epsilon)$ *be*

$$p(x, \epsilon) = \sum_{k=0}^{D} p_k(x)\epsilon^k$$

for D *the degree of* p *in* y *and* $p_k(x) \in \mathbb{C}[x]$ *for all* $k \in [D] \cup \{0\}$. *Then,* $x = 0$ *is a root of* $p_1(x)$.

Proof. By the above expansion, we can write

$$p_1(x) = \frac{d}{d\epsilon}p(x, \epsilon)\Big|_{\epsilon=0}.$$

Using the definition of p, we obtain

$$p_1(0) = \frac{d}{d\epsilon}q(0, \epsilon)\Big|_{\epsilon=0} r(0, 0) + q(0, 0)\frac{d}{d\epsilon}r(0, \epsilon)\Big|_{\epsilon=0},$$

which is zero since we assumed $q(0, 0) = 0 = r(0, 0)$. □

Appendix D. No Information Without Disturbance for Positive Maps

In the ordinary setting, the statement that there is no information without disturbance is proven for completely positive maps, because those are the physically relevant evolutions of the system. The statement then has a short proof using Choi matrices. In this section, we show that the statement still holds for merely positive maps. This is used, e.g., in Sect. 9.

Lemma D.1. *Let $\mathcal{T}_i : \mathcal{M}_D \to \mathcal{M}_D$, $i \in [s]$, be a collection of positive linear maps such that*

$$\sum_{i=1}^{s} \mathcal{T}_i = \mathrm{id}.$$

Then, $\mathcal{T}_i = c_i \mathrm{id}$ for some $c_i \geq 0$ for all $i \in [s]$ and $\sum_{i=1}^{s} c_i = 1$.

Proof. Let $|\psi\rangle\langle\psi| \in \mathcal{S}(\mathbb{C}^D)$. Then, $\mathcal{T}_i(|\psi\rangle\langle\psi|) = c_i(\psi,\psi)\,|\psi\rangle\langle\psi|$ for some number $c_i(\psi,\psi) \geq 0$ and for all $i \in [s]$. This follows from positivity of the maps, the fact that they sum to the identity and because the rank-one projections are extremal in the set of states. We have to show that the constant does not depend on the state. Consider $A := (x\,|\psi\rangle + y\,|\phi\rangle)(\bar{x}\,\langle\psi| + \bar{y}\,\langle\phi|)$, with $|\psi\rangle$, $|\phi\rangle$ orthonormal, x, $y \in \mathbb{C}$. Again, $\mathcal{T}_i(A) = c_i(A)A$. By linearity,

$$\mathcal{T}_i(A) = c_i(\psi,\psi)|x|^2\,|\psi\rangle\langle\psi| + c_i(\phi,\phi)|y|^2\,|\phi\rangle\langle\phi| + \mathcal{T}_i(x\,|\psi\rangle\langle\phi|\,\bar{y} + y\,|\phi\rangle\langle\psi|\,\bar{x}).$$

Let $\tilde{A} = a\,|\psi\rangle\langle\psi| + b\,|\phi\rangle\langle\phi| + c\,|\phi\rangle\langle\psi| + \bar{c}\,|\psi\rangle\langle\phi|$. This matrix is positive semidefinite for $a \geq 0$, $ab - |c|^2 \geq 0$, $b \geq 0$, $c \in \mathbb{C}$. Note that we can scale $a \to \lambda a$, $b \to \frac{1}{\lambda} b$ for $\lambda \in \mathbb{R} \setminus \{0\}$ while keeping c constant. With $\mathcal{T}_i(\tilde{A}) \leq \tilde{A}$ and $\mathcal{T}_i(\tilde{A}) \geq 0$, we can infer

$$\langle\theta|\,\mathcal{T}_i(x\,|\psi\rangle\langle\phi|\,\bar{y} + y\,|\phi\rangle\langle\psi|\,\bar{x})\,|\theta\rangle = 0 \quad \forall\,|\theta\rangle \in \{\,|\psi\rangle, |\phi\rangle\,\}$$

by scaling with an appropriate λ. Hence,

$$\mathcal{T}_i(x\,|\psi\rangle\langle\phi|\,\bar{y} + y\,|\phi\rangle\langle\psi|\,\bar{x}) = c_i(x\psi, y\phi)\,|\psi\rangle\langle\phi| + \bar{c}_i(x\psi, y\phi)\,|\phi\rangle\langle\psi|\,.$$

Thus, computing $\langle\theta_1|\,\mathcal{T}_i(A)\,|\theta_2\rangle$ for $|\theta_1\rangle$, $|\theta_2\rangle \in \{\,|\psi\rangle, |\phi\rangle\,\}$ yields that both $c_i(\psi,\psi) = c_i(\phi,\phi)$ and $c_i(x\psi, y\phi) = c_i(\psi,\psi)x\bar{y}$. Thus, the constants do not depend on $|\psi\rangle$ and $|\phi\rangle$. Choosing an orthonormal basis and the corresponding usual basis of Hermitian operators, this implies that $\mathcal{T}_i = c_i \mathrm{id}$ for $c_i \geq 0$ for all $i \in [s]$. □

Appendix E. Complexity of Block Diagonalization and InterpolationSDP

E.1. Block Diagonalization

In this section, we will analyze the complexity of determining the minimal compression dimension. This is needed in Sect. 8. We start with the block diagonalization part. Assume we are given linearly independent Hermitian operators $\{E_1, \ldots, E_k\} = \mathcal{O} \subset \mathcal{M}_D$ with entries in $\mathbb{Q}(\mathrm{i})$ (the complex numbers with rational real and imaginary part). We first need to determine the composition of the C*-algebra generated by $\mathcal{O}$ into irreducible components,

$$\mathrm{C}^*(\mathcal{O}) = U^* \left(\bigoplus_{i=1}^{s} \mathcal{M}_{D_i} \otimes \mathbb{1}_{m_i} \right) U, \tag{21}$$

where $U \in \mathcal{M}_D$ is unitary and $m_i \in \mathbb{N}$ for all $i \in [s]$. For this, we use the complex version of the algorithm proposed in [24]. This algorithm is formulated over the real numbers and can be adapted to the complex case. However, we will show that we can still use it if we only allow for algebraic numbers, which we will denote by $\mathbb{A}$. This is more realistic for practical computations. For the real algebraic numbers, we will write $\mathbb{A}_{\mathbb{R}}$. By Eq. (21), we can write

$$E_j = U^* \left(\bigoplus_{i=1}^{s} E_j^i \otimes \mathbb{1}_{m_i} \right) U \quad \forall j \in [k]$$

with $E_j^i \in \mathcal{M}_{D_i}$ for all $i \in [s]$, $j \in [k]$. We note that the entries of U, E_j^i can be chosen to be in $\mathbb{A}$, since they are the solutions to a system of polynomial equalities, which we can split into real and imaginary part. As $\mathbb{A}_{\mathbb{R}}$ is a real closed field, it follows by the Tarski transfer principle [23, 11.2.3] that this system of equations has a solution in $\mathbb{A}_{\mathbb{R}}$ if and only if it has a solution in $\mathbb{R}$. The latter is guaranteed by Eq. (21).

If the $E_1, \ldots, E_k$ do not linearly span $\mathrm{C}^*(\mathcal{O})$, we may find such a basis by a procedure similar to the one described in [24, comment after Proposition 5]. Note that in the complex case, we need to add elements of the form $\mathrm{i}(AB - BA)/2$ in each step as well. We further assume that we are given a finite set $\mathcal{B} \subset \mathbb{Q}$ with at least $(s/\epsilon) \max_{i \in [s]} D_i$ elements for some $\epsilon \in (0, 1)$. Choosing $r \in \mathcal{B}^k$ randomly from a uniform distribution, the element $E(r) = r_1 E_1 + \ldots r_k E_k$ is generic with probability at least $1 - \epsilon$. Generic means that elements in the different simple components in Eq. (21) have different eigenvalues. This can be guaranteed by avoiding the zero set of a polynomial which is the product of the resultants of the characteristic polynomials for the respective blocks. The lower bound follows from the Schwartz–Zippel lemma [29, Corollary 1] applied to that polynomial and a union bound. See [24, Proposition 3] for details. We can rescale r such that $r \in \mathbb{Q}$, $\|r\|_2 \leq 1$. Then, we can compute the characteristic polynomial of $E(r)$ and use the (probabilistic) factorization algorithm based on basis reduction in [34, Corollary 16.25] to factor it into irreducible components. Note that this gives us the eigenvalues of $E(r)$ with their respective algebraic multiplicities, since the algebraic numbers are defined by their minimal polynomials. Using Gaussian elimination, we can obtain the corresponding eigenvectors. Grouping the eigenvalues into sets as described in [24, Proposition 2], we have found the decomposition into irreducible elements. The second part of the algorithm, finding the irreducible factors, can be carried out exactly as described by [24]. The overall complexity is dominated by the factorization of the characteristic polynomial. Its maximal coefficient has modulus at most $(kDM)^D$, where $M = \max_{i \in k} \|E_i\|_\infty$. Therefore, the factorization needs an expected number of $\mathcal{O}(D^{10} \mathrm{polylog}(k) \mathrm{polylog}(D) \mathrm{polylog}(M))$ arithmetic operations. The algorithm succeeds with probability at least $1 - \epsilon$, because the element $E(r)$ needs to be generic. Thus, we have obtained $E_1, \ldots, E_k$ in block diagonal form as required for Algorithm 1.

E.2. Complexity of InterpolationSDP

In the rest of this section, we will comment on the complexity of solving the semidefinite program $\mathrm{InterpolationSDP}(E_1, \ldots, E_k)(j)$. We have to convert the unbounded optimization problem into a feasibility problem to be able to practically solve it. The new SDP is:

Given $E_i^j \in \mathcal{M}_{D_j}^{\text{herm}}$, $j \in [s]$, $i \in [k]$, *determine whether there are* $H_i \in \mathcal{M}_{D_1}$, $i \in [k+1]$ *such that*

$$\sum_{i=1}^{k} \begin{bmatrix} 0 & & & \\ & E_i^2 & & \\ & & \ddots & \\ & & & E_i^s \end{bmatrix} \otimes H_i \geq 0$$

$$-1 - \left(\sum_{i=1}^{k} \operatorname{Tr}\left((E_i^1)^T H_i \right) \right) \geq 0.$$

Proposition E.1. *If the* E_i^j *have entries in* $\mathbb{Q}(\mathrm{i})$ *for all* $j \in [s]$, $i \in [k]$, *then the feasibility of this SDP can be determined in* $\mathcal{O}(kD_1^6 D^4) + (D_1 D)^{\mathcal{O}(kD_1^2)}$ *operations.*

Proof. This follows from the results in [28]. Theorem 5.7 of that paper states that the given symmetric $n \times n$ matrices $Q_0, \ldots, Q_m$ with integer entries, the question whether there are real numbers $x_1, \ldots, x_m$ such that

$$Q_0 + x_1 Q_1 + \ldots + x_m Q_m \geq 0$$

can be decided using $\mathcal{O}(mn^4) + n^{\mathcal{O}(\min\{\, m, n^2 \,\})}$ operations. To use this theorem, we have to convert the SDP into standard form. This can be done using a basis of the Hermitian matrices and expressing the H_i as a real combination of basis elements. The two constraints can be combined into one writing them as a block matrix. Multiplying the equations by an appropriate positive integer, we can assume that they have integer coefficients. Finally, we can convert a complex SDP into a real SDP while increasing the dimension of the matrices by a factor of 2. Thus, we have $m = kD_1^2$ and $n = 2(D_1 D + 1)$ and the result follows by [28, Theorem 5.7]. □

In our application of the algorithm, we assumed that $k \leq D^2$. If we have $k = \mathcal{O}(1)$ and $D_1 = \mathcal{O}(1)$, which means that we are interested in just a few effect operators and we have an upper bound on the block dimension uniform in D, the SDP can be solved in a number of operations polynomial in D. If this is not the case, the performance of the algorithm can be significantly worse. The reason for this is that the separation between the two cases interpolation possible/impossible can become double exponentially small if we bound the operator norm of the H_i we allow.

Acknowledgements

AB would like to thank Michael Kech for helpful discussions concerning the use of semi-algebraic sets and the Stack Exchange user Simone Weil for pointing out the strategy to prove the existence of irreducible hyperbolic polynomials in any dimension. Further, AB would like to thank the referee for helpful comments and for proposing the alternative argument in the proof of Lemma

5.2. MMW thanks John Watrous for suggesting the problem and David Pérez García as well as John Watrous for interesting discussions.

References

[1] Aubrun, G., Lancien, C.: Zonoids and sparsification of quantum measurements. Positivity **20**(1), 1–23 (2016)

[2] Arveson, W.: Subalgebras of C^*-algebras II. Acta Math. **128**(1), 271–308 (1972)

[3] Bochnak, J., Coste, M., Roy, M.-F.: Real Algebraic Geometry. Volume 36 of Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer, New York (1998)

[4] Bix, R.: Conics and Cubics, 2nd edn. Springer, New York (2006)

[5] Blackadar, B.: Operator Algebras: Theory of C*-Algebras and Von Neumann Algebras. Volume 13 of Encyclopaedia of Mathematical Sciences. Springer, New York (2006)

[6] Böttcher, A., Spitkovsky, I.M.: A gentle guide to the basics of two projections theory. Linear Algebra Appl. **432**(6), 1412–1459 (2010)

[7] Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)

[8] Choi, M.-D.: Completely positive linear maps on complex matrices. Linear Algebra Appl. **10**(3), 285–290 (1975)

[9] Dixmier, J.: Les algèbres d'opérateurs dans l'espace Hilbertien. Gauthier-Villars, (1957)

[10] Farenick, D.R.: Algebras of Linear Transformations. Springer, New York (2001)

[11] George, A., Ikramov, K.D.: Common invariant subspaces of two matrices. Linear Algebra Appl. **287**(1), 171–179 (1999)

[12] Güler, O.: Hyperbolic polynomials and interior point methods for convex programming. Math. Oper. Res. **22**(2), 350–377 (1997)

[13] Halmos, P.R.: Two subspaces. Trans. Am. Math. Soc. **144**, 381–389 (1969)

[14] Heinosaari, T., Jivulescu, M.A., Reeb, D., Wolf, M.M.: Extending quantum operations. J. Math. Phys. **53**(10), 102208 (2012)

[15] Horn, R.A.: Matrix Analysis, 2nd edn. Cambridge University Press, Cambridge (2012)

[16] Heinosaari, T., Ziman, M.: The Mathematical Language of Quantum Theory. Cambridge University Press, Cambridge (2012)

[17] Kato, T.: Perturbation Theory for Linear Operators. Volume 132 of Die Grundlehren der mathematischen Wissenschaften. Springer, New York (1966)

[18] Keyl, M.: Fundamentals of quantum information theory. Phys. Rep. **369**(5), 431–548 (2002)

[19] Kech, M., Wolf, M.M.: Constrained quantum tomography of semi-algebraic sets with applications to low-rank matrix recovery. Inf. Inference **6**(2), 171–195 (2017)

[20] Lindblad, G.: A general no-cloning theorem. Lett. Math. Phys. **47**(2), 189–196 (1999)

[21] Lewis, A.S., Parrilo, P.A., Ramana, M.V.: The Lax conjecture is true. Proc. Am. Math. Soc. **133**, 2495–2499 (2005)

[22] Lomonosov, V., Rosenthal, P.: The simplest proof of Burnside's theorem on matrix algebras. Linear Algebra Appl. **383**, 45–47 (2004)

[23] Marshall, M.: Positive Polynomials and Sums of Squares. Volume 146 of Mathematical Surveys and Monographs. American Mathematical Society, Providence (2008)

[24] Murota, K., Kanno, Y., Kojima, M., Kojima, S.: A numerical algorithm for block-diagonal decomposition of matrix $*$-algebras with application to semidefinite programming. Jpn. J. Ind. Appl. Math. **27**(1), 125–160 (2010)

[25] Nielsen, M.A., Chuang, I.L.: Quantum computation and quantum information, 10th anniversary edition. Cambridge University Press, Cambridge (2010)

[26] Nuij, W.: A note on hyperbolic polynomials. Math. Scand. **23**, 69–72 (1968)

[27] Paulsen, V.: Completely Bounded Maps and Operator Algebras. Cambridge University Press, Cambridge (2002)

[28] Porkolab, L., Khachiyan, L.: On the complexity of semidefinite programs. J. Glob. Optim. **10**, 351–365 (1997)

[29] Schwartz, J.T.: Fast probabilistic algorithms for verification of polynomial identities. J. ACM **27**(4), 701–717 (1980)

[30] Schuhmacher, B.: Quantum coding. Phys. Rev. A **51**(4), 2738–2747 (1995)

[31] Stark, C.J., Harrow, A.W.: Compressibility of positive semidefinite factorizations and quantum models. IEEE Trans. Inf. Theory **62**(5), 2867–2880 (2016)

[32] Terhal, B.M., Horodecki, P.: A Schmidt number for density matrices. Phys. Rev. A **61**(4), 040301 (2000)

[33] Vinnikov, V.: Complete description of determinantal representations of smooth irreducible curves. Linear Algebra Appl. **125**, 103–140 (1989)

[34] von zur Gathen, J.: Modern Computer Algebra. Cambridge University Press, Cambridge (2013)

[35] Wehner, S., Christandl, M., Doherty, A.C.: Lower bound on the dimension of a quantum system given measured data. Phys. Rev. A **78**, 062112 (2008)

[36] Wedin, P.: On Angles Between Subspaces of a Finite Dimensional Inner Product Space, pp. 263–285. Springer, New York (1983)

[37] Winter, A.: "Extrinsic" and "intrinsic" data in quantum measurements: asymptotic convex decomposition of positive operator valued measures. Commun. Math. Phys. **244**(1), 157–185 (2004)

[38] Wolf, M.M.: Quantum channels and operations. Lecture notes (2012). http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf

Andreas Bluhm, Lukas Rauber and Michael M. Wolf
Zentrum Mathematik
Technische Universität München
Boltzmannstr. 3
85748 Garching
Germany
e-mail: bluhm@ma.tum.de

Lukas Rauber
e-mail: lukas.rauber@tum.de

Michael M. Wolf
e-mail: m.wolf@tum.de

Communicated by David Perez-Garcia.
Received: August 31, 2017.
Accepted: January 25, 2018.

A.2 Joint measurability of quantum effects and the matrix diamond

Joint measurability of quantum effects and the matrix diamond

Andreas Bluhm and Ion Nechita

Quantum incompatibility is one of the defining properties of quantum mechanics and is essential for tasks such as quantum cryptography. A famous example of incompatible observables are the momentum and the position of a particle. In this work, we prove a connection between the compatibility of binary quantum measurements and an inclusion problem of free spectrahedra.

We start our work by reviewing some results on the compatibility of quantum measurements in Section III. In particular, we discuss the relation between different noise models which can be used to quantify incompatibility in Proposition III.4. In Section IV, we collect some results on the inclusion of free spectrahedra which are needed for the rest of the work. The free spectrahedron which is particularly relevant for this work is the matrix diamond, which is a matricial relaxation of the ℓ_1-ball. The connection between the compatibility of binary measurements and the inclusion of free spectrahedra is made in Section V. Our first main result is Theorem V.3. It connects the compatibility of a collection of binary measurements to the inclusion of the matrix diamond into a free spectrahedron defined by the measurements and has three parts. The first part states that the given operators define valid measurements if and only if the inclusion holds at level 1. The second part asserts that the operators define compatible measurements if and only if the inclusion holds at the level of free spectrahedra. The third part finds that inclusion at level k holds if and only if all k-dimensional compressions of the given operators define compatible measurements. In this sense, the inclusion of the different levels of the free spectrahedra correspond to different degrees of compatibility. Our second main result is Theorem V.7 which proves that the balanced compatibility region of g binary measurements in dimension d corresponds to the d-dimensional inclusion set of the matrix diamond with g coordinates.

This correspondence allows us to use results on the inclusion of free spectrahedra to prove bounds on the balanced compatibility region. Section VII gives lower bounds, whereas Section VIII.B provides upper bounds. Theorem VII.7 states that the balanced compatibility region for binary measurements always contains the positive orthant of the Euclidean unit ball. If the dimension is exponential in the number of measurements, Theorem VIII.8 proves that the reverse inclusion is also true, thus fully characterizing the compatibility region in this case. The proof of Theorem VIII.8 also provides the most incompatible effects in this situation as pointed out in Remark VIII.10.

It is also possible to use quantum information techniques to obtain bounds on the inclusion set for the matrix diamond. Section VI reviews lower bounds derived from approximate cloning whereas Section VIII.A provides upper bounds we infer from the study of mutually unbiased bases (MUBs) and symmetric informationally complete positive operator-valued measures (SIC-POVMs).

We conclude in Section IX with a discussion of all the bounds obtained, both on the compatibility regions for different versions of noise and on the inclusion sets for the matrix diamond. Finally, we show which inclusion problem corresponds to the compatibility of a binary measurement and a measurement with three outcomes, but leave the general case for Article IV.

I was significantly involved in finding the ideas and carrying out the scientific work of all parts of this article. Furthermore, I was in charge of the writing of all parts with the exception of Sections VI and VIII.A.

Permission to include:

Andreas Bluhm and Ion Nechita
Joint measurability of quantum effects and the matrix diamond.
Journal of Mathematical Physics 59, 112202 (2018).

Permission to Reuse Content

REUSING AIP PUBLISHING CONTENT

Permission from AIP Publishing is required to:

- republish content (e.g., excerpts, figures, tables) if you are not the author
- modify, adapt, or redraw materials for another publication
- systematically reproduce content
- store or distribute content electronically
- copy content for promotional purposes

To request permission to reuse AIP Publishing content, use RightsLink® for the fastest response or contact AIP Publishing directly at rights@aip.org (mailto:rights@aip.org) and we will respond within one week:

For RightsLink, use Scitation to access the article you wish to license, and click on the Reprints and Permissions link under the TOOLS tab. (For assistance click the “Help” button in the top right corner of the RightsLink page.)

To send a permission request to rights@aip.org (mailto:rights@aip.org), please include the following:

- Citation information for the article containing the material you wish to reuse
- A description of the material you wish to reuse, including figure and/or table numbers
- The title, authors, name of the publisher, and expected publication date of the new work
- The format(s) the new work will appear in (e.g., print, electronic, CD-ROM)
- How the new work will be distributed and whether it will be offered for sale

Authors do **not** need permission from AIP Publishing to:

- quote from a publication (please include the material in quotation marks and provide the customary acknowledgment of the source)
- reuse any materials that are licensed under a Creative Commons CC BY license (please format your credit line: “Author names, Journal Titles, Vol.#, Article ID#, Year of Publication; licensed under a Creative Commons Attribution (CC BY) license.”)
- reuse your own AIP Publishing article in your thesis or dissertation (please format your credit line: “Reproduced from [FULL CITATION], with the permission of AIP Publishing”)
- make multiple copies of articles–although you must contact the Copyright Clearance Center (CCC) at www.copyright.com (http://www.copyright.com/) to do this

REUSING CONTENT PUBLISHED BY OTHERS

To request another publisher’s permission to reuse material in AIP Publishing articles, please use our

JOURNAL OF MATHEMATICAL PHYSICS **59**, 112202 (2018)

Joint measurability of quantum effects and the matrix diamond

Andreas Bluhm[1,a)] and Ion Nechita[2,b)]
[1]*Zentrum Mathematik, Technische Universität München, Boltzmannstrasse 3, 85748 Garching, Germany*
[2]*Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, Paris, France*

(Received 19 July 2018; accepted 1 November 2018; published online 28 November 2018)

In this work, we investigate the joint measurability of quantum effects and connect it to the study of free spectrahedra. Free spectrahedra typically arise as matricial relaxations of linear matrix inequalities. An example of a free spectrahedron is the matrix diamond, which is a matricial relaxation of the ℓ_1-ball. We find that joint measurability of binary positive operator valued measures is equivalent to the inclusion of the matrix diamond into the free spectrahedron defined by the effects under study. This connection allows us to use results about inclusion constants from free spectrahedra to quantify the degree of incompatibility of quantum measurements. In particular, we completely characterize the case in which the dimension is exponential in the number of measurements. Conversely, we use techniques from quantum information theory to obtain new results on spectrahedral inclusion for the matrix diamond. *Published by AIP Publishing.* https://doi.org/10.1063/1.5049125

I. INTRODUCTION

One of the defining properties of quantum mechanics is the existence of incompatible observables, i.e., measurements that cannot be performed simultaneously.[6,25] A classic example of this behavior is the observables of position and momentum. One of the central notions to capture this property of quantum mechanics is joint measurability. Observables are jointly measurable if they arise as marginals from a common observable. This has practical implications for quantum information tasks[8] as only incompatible observables can violate Bell inequalities.[16]

It is well-known that incompatible observables can be made compatible by adding a sufficient amount of noise.[11] Although many works study compatibility questions for concrete observables (see Ref. 22 for a topical review), there has also been interest in how much incompatibility there is in quantum mechanics and other generalized probabilistic theories.[11,17] In the present work, we continue this line of research by studying the degree of incompatibility in quantum mechanics in more detail. We will be interested in the compatibility regions for a fixed number of binary measurements in fixed dimension and for different types of noise.

For this, we will use tools from the study of free spectrahedra (see Ref. 26 for a general introduction). Concretely, we are interested in the problem of (free) spectrahedral inclusion.[27] Originally, the inclusion of free spectrahedra has been introduced as a relaxation to study the inclusion of ordinary spectrahedra.[3,28] In contrast to that, we will be interested in the inclusion constants for their own sake. Often, results on free spectrahedral inclusion work for large classes of spectrahedra, e.g., spectrahedra with symmetries.[14,28] Recently, results have been found which study maximal and minimal free spectrahedra for the p-norm unit balls.[38] It is especially the latter work which is most useful to us. We would also like to mention that another point of contact between quantum information theory and free analysis is the extension (or interpolation) problem for completely positive maps; see Refs. 1 and 20.

a)Electronic mail: andreas.bluhm@ma.tum.de
b)Electronic mail: nechita@irsamc.ups-tlse.fr

In this work, we establish the connection between free spectrahedral inclusion and joint measurability. The matricial relaxation of the ℓ_1-ball is known as the *matrix diamond* and plays a central role in our setting. We can then use the results on inclusion constants for this free spectrahedron to characterize the degree of incompatibility of quantum effects in different settings. Conversely, we translate techniques to prove upper and lower bounds on quantum incompatibility to study spectrahedral inclusion. Let us note that since the problems of joint measurability and quantum steering are closely related,[43] many of our results can be translated to the steering framework.

II. MAIN RESULTS

In this section, we will briefly outline the main findings of our work. Its main contribution is to connect the following two seemingly unrelated problems.

One is the problem of *joint measurability of binary quantum observables*. Given a g-tuple of quantum effects $E_1, \ldots, E_g$, we can ask the question of how much noise we have to add to the corresponding measurements to make them jointly measurable. Joint measurability means that there exists a joint positive operator valued measure (POVM) $\left\{R_{i_1,\ldots,i_g}\right\}$ from which the binary POVMs we are interested in arise as marginals. Noise can be added in different ways to a measurement. We will mainly consider the case in which we take convex combinations of a quantum measurement with a fair coin, i.e.,

$$E' := sE + (1-s)I/2,$$

for $s \in [0, 1]$. The set of g-tuples $s \in [0, 1]^g$ which make *any* g binary POVMs of dimension d compatible will be denoted as $\Gamma(g, d)$.

The other problem comes from the field of *free spectrahedra*. A free spectrahedron is a matricial relaxation of an ordinary spectrahedron. The free spectrahedron $\mathcal{D}_A$ is then the set of self-adjoint matrix g-tuples X of arbitrary dimension which fulfill a given linear matrix inequality

$$\sum_{i=1}^{g} A_i \otimes X_i \leq I.$$

If we only consider the scalar elements $\mathcal{D}_A(1)$ of this set, this is just the ordinary spectrahedron defined by the matrix tuple A. The inclusion problem for free spectrahedra is to find the scaling factors $s \in \mathbb{R}^g_+$ such that the implication

$$\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1) \Rightarrow s \cdot \mathcal{D}_A \subseteq \mathcal{D}_B \tag{1}$$

is true. We will be interested in the case in which $\mathcal{D}_A$ is the matrix diamond, i.e., the set of matrices X such that

$$\sum_{i=1}^{g} \epsilon_i X_i \leq I, \qquad \forall \epsilon \in \{-1, 1\}^g.$$

The set of all such s which make the implication in Eq. (1) true for any $B \in (\mathcal{M}_d^{sa})^g$ in this case will be written as $\Delta(g, d)$.

The main contribution of our work is to relate these two problems and use this connection to characterize $\Gamma(g, d)$. In Theorem V.3, we find the following:

Theorem II.1. *Let $E \in \left(\mathcal{M}_d^{sa}\right)^g$ and let $2E - I := (2E_1 - I_d, \ldots, 2E_g - I_d)$. We have*

1. *$\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_{2E-I}(1)$ if and only if $E_1, \ldots, E_g$ are quantum effects.*
2. *$\mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{2E-I}$ if and only if $E_1, \ldots, E_g$ are jointly measurable quantum effects.*
3. *$\mathcal{D}_{\diamond,g}(k) \subseteq \mathcal{D}_{2E-I}(k)$ for $k \in [d]$ if and only if for any isometry $V : \mathbb{C}^k \hookrightarrow \mathbb{C}^d$, the induced compressions $V^*E_1V, \ldots, V^*E_gV$ are jointly measurable quantum effects.*

This shows that there is a one-to-one correspondence between different levels of the spectrahedral inclusion problem and different degrees of compatibility. Furthermore, we show in Theorem V.7 that finding spectral inclusion constants corresponds to making POVMs compatible through adding noise.

Theorem II.2. *It holds that* $\Gamma(g, d) = \Delta(g, d)$.

This result allows us to use the results on spectrahedral inclusion in order to characterize the set $\Gamma(g, d)$. We find that the higher dimensional generalization of the positive quarter of the unit circle plays an important role in this,

$$\mathrm{QC}_g := \left\{ s \in \mathbb{R}^g_+ : \sum_{i=1}^{g} s_i^2 \leq 1 \right\}.$$

The adaptation of some results of Ref. 38 allows us to show in Theorem VII.7:

Theorem II.3. *Let* $g, d \in \mathbb{N}$. *Then, it holds that* $QC_g \subseteq \Gamma(g, d)$. *In other words, for any g-tuple* $E_1, \ldots, E_g$ *of quantum effects and any positive vector* $s \in \mathbb{R}^g_+$ *with* $\|s\|_2 \leq 1$, *the g-tuple of noisy effects*

$$E_i' = s_i E_i + (1 - s_i)\frac{I_d}{2}$$

is jointly measurable.

If the dimension of the effects under study is exponential in the number of measurements, Theorem VIII.8 provides us with a converse result. Again, this theorem is based on a result of Ref. 38.

Theorem II.4. *Let* $g \geq 2$, $d \geq 2^{\lceil (g-1)/2 \rceil}$. *Then,* $\Gamma(g, d) \subseteq QC_g$.

Thus, for $g \geq 2$, $d \geq 2^{\lceil (g-1)/2 \rceil}$, we infer that $\Gamma(g, d) = \mathrm{QC}_g$; this equality was known previously only in the case $g = 2$. However, this can no longer be the case for many measurements in low dimensions, as we point out in Sec. IX. For other types of noise added to quantum measurements, we can use similar results to give upper and lower bounds on the compatibility regions. The bounds we obtain with our techniques improve greatly on past results in the quantum information literature. As an example, the best lower bound in the symmetric case came from cloning and was of order $1/g$ for fixed g and large d; see Proposition VI.2. Our results yield a lower bound of $1/\sqrt{g}$, which turns out to be exact in the regime $g \ll d$; we refer the reader to Sec. IX for a detailed comparison of these bounds.

Conversely, we can use techniques from quantum information such as asymmetric cloning (Sec. VI) to give bounds on spectrahedral inclusion in different settings. In particular, we introduce in this work a generalization of the notion of inclusion constants from Ref. 28 in two directions: first by restricting both the size and the number of the matrices appearing in the spectrahedron and then by allowing asymmetric scalings of the spectrahedra; see Definition IV.1. Our contribution to the inclusion theory of free spectrahedra is going beyond the results from Ref. 38 by studying the asymmetric and size-dependent inclusion constants.

III. CONCEPTS FROM QUANTUM INFORMATION THEORY

In this section, we will start by reviewing some notions from quantum information theory related to measurements. Subsequently, we will define several versions of incompatibility of quantum measurements and show basic relations between them. For an introduction to the mathematical formalism of quantum mechanics, see Ref. 24 or Ref. 45, for example.

Before we move to the quantum formalism, let us introduce some basic notation. For brevity, we will write $[n] := \{1, \ldots, n\}$ for $n \in \mathbb{N}$ and $\mathbb{R}^g_+$ for $\{x \in \mathbb{R}^g : x_i \geq 0 \ \forall i \in [g]\}$, $g \in \mathbb{N}$. Additionally, $\lceil \cdot \rceil : \mathbb{R} \to \mathbb{Z}$ will be the ceiling function. Furthermore, for $n, m \in \mathbb{N}$, let $\mathcal{M}_{n,m}$ be the set of complex $n \times m$ matrices. If $m = n$, we will just write $\mathcal{M}_n$. We will write $\mathcal{M}_n^{\mathrm{sa}}$ for the self-adjoint matrices and $\mathcal{U}_d$ for the unitary $d \times d$ matrices. $I_n \in \mathcal{M}_n$ will be the identity matrix. We will often drop the index if the dimension is clear from the context. For $A \in (\mathcal{M}_d^{sa})^g$, let $\mathcal{OS}_A$ be the *operator system* defined by the g-tuple A, i.e.,

$$\mathcal{OS}_A := \text{span}\{I_d, A_i : i \in [g]\}.$$

Moreover, we will often write for such tuples $2A - I := (2A_1 - I_d, \ldots, 2A_g - I_d)$ and $V^*AV := (V^*A_1V, \ldots, V^*A_gV)$ for $V \in \mathcal{M}_{d,k}$, $k \in \mathbb{N}$.

A quantum mechanical system is described by its *state* $\rho \in \mathcal{S}(\mathcal{H})$, where $\mathcal{H}$ is the Hilbert space associated with the system and

$$\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H}) : \rho \geq 0, \text{tr}[\rho] = 1\}.$$

In the present work, all Hilbert spaces will be finite dimensional. To describe transformations between quantum systems, we will use the concept of completely positive maps. Let $\mathcal{T} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a linear map with $\mathcal{H}$ and $\mathcal{K}$ being two Hilbert spaces. This map is k-positive if for $k \in \mathbb{N}$, the map $\mathcal{T} \otimes \text{Id}_k : \mathcal{B}(\mathcal{H}) \otimes \mathcal{M}_k \to \mathcal{B}(\mathcal{K}) \otimes \mathcal{M}_k$ is a positive map. A map is called completely positive if it is k-positive for all $k \in \mathbb{N}$. If this map is additionally trace preserving, it is called a *quantum channel*.

Let Eff_d be the set of d-dimensional *quantum effects*, i.e.,

$$\text{Eff}_d := \left\{E \in \mathcal{M}_d^{\text{sa}} : 0 \leq E \leq I\right\}.$$

Effect operators are useful to describe quantum mechanical measurements. In quantum information theory, measurements correspond to *positive operator valued measures* (POVMs). A POVM is a set $\{E_i\}_{i\in\Sigma}$, $E_i \in \text{Eff}_d$ for all $i \in \Sigma$, such that

$$\sum_{i\in\Sigma} E_i = I.$$

Here, Σ is the set of measurement outcomes, which we will assume to be finite for simplicity and equal to $[m]$ for some $m \in \mathbb{N}$. For the case of binary POVMs ($m = 2$), we will identify the POVM $\{E, I - E\}$ with its effect operator $E \in \text{Eff}_d$.

If a collection of POVMs can be written as marginals of a common POVM with more outcomes, we will say that they are jointly measurable (see Ref. 22 for an introduction to the topic).

Definition III.1 (**Jointly measurable POVMs**). *We consider a collection of d-dimensional POVMs $\left\{E_j^{(i)}\right\}_{j\in[m_i]}$, where $m_i \in \mathbb{N}$ for all $i \in [g]$, $g \in \mathbb{N}$. These POVMs are* jointly measurable *or* compatible *if there is a d-dimensional POVM $\left\{R_{j_1,\ldots,j_g}\right\}$ with $j_i \in [m_i]$ such that for all $u \in [g]$ and $v \in [m_u]$,*

$$E_v^{(u)} = \sum_{\substack{j_i\in[m_i]\\ i\in[g]\setminus\{u\}}} R_{j_1,\ldots,j_{u-1},v,j_{u+1},\ldots,j_g}.$$

It is well-known[22] that not all quantum mechanical measurements are compatible. In concrete situations, the joint measurability of POVMs can be checked using a semidefinite program (SDP) (see, e.g., Ref. 47). Note that the SDP for g binary POVMs has 2^g variables, so when the number of effects g is large, it becomes computationally costly to decide compatibility. However, incompatible measurements can be made compatible by adding noise to the respective measurements. A trivial measurement is a POVM in which all effects are proportional to the identity. Adding noise to a measurement then means taking a convex combination of the original POVM and a trivial measurement. In order to quantify incompatibility of measurements, we can define several sets which differ in the type of noise we allow. We will restrict ourselves to binary POVMs in this work. For our first set, we allow different types of noise for every POVM,

$$\Gamma^{all}(g,d) := \left\{s \in [0,1]^g : \forall E_1, \ldots, E_g \in \text{Eff}_d, \exists a \in [0,1]^g \text{ s.t. } s_iE_i + (1-s_i)a_iI \text{ are compatible}\right\}.$$

Another possibility is to consider only balanced noise,

$$\Gamma(g,d) := \left\{s \in [0,1]^g : \ s_iE_i + \frac{1-s_i}{2}I \text{ are compatible } \forall E_1, \ldots, E_g \in \text{Eff}_d\right\}.$$

Sometimes, it is inconvenient that the map from the original measurements to the ones with added noise is non-linear in the effect operators. To remedy this, we define

$$\Gamma^{lin}(g,d) := \left\{ s \in [0,1]^g : s_i E_i + (1-s_i)\frac{\mathrm{tr}[E_i]}{d} I \text{ are compatible } \forall E_1, \ldots, E_g \in \mathrm{Eff}_d \right\}.$$

The restriction of this set to equal weights has appeared before in the context of quantum steering.[21,44]

Instead of restricting the type of noise allowed, we can also consider less general POVMs and restrict to those which are unbiased,

$$\Gamma^{0}(g,d) := \left\{ s \in [0,1]^g : s_i E_i + \frac{1-s_i}{2} I \text{ are compatible } \forall E_1, \ldots, E_g \in \mathrm{Eff}_d \text{ s.t. } \mathrm{tr}[E_i] = \frac{d}{2} \right\}.$$

Finally, let us introduce a set of parameters related to (asymmetric) cloning of quantum states,

$$\begin{aligned} \Gamma^{clone}(g,d) := \{ s \in [0,1]^g \ : \ \exists \mathcal{T} : \mathcal{M}_d^{\otimes g} \to \mathcal{M}_d \text{ unital and completely positive s.t.} \\ \forall X \in \mathcal{M}_d,\ \forall i \in [g],\ \mathcal{T}\left(I^{\otimes(i-1)} \otimes X \otimes I^{\otimes(n-i)}\right) = s_i X + (1-s_i)\frac{\mathrm{tr}[X]}{d} I \}. \end{aligned} \tag{2}$$

All these sets are convex sets, as the next proposition shows.

Proposition III.2. $\Gamma^{\#}(g, d)$ *is convex for* $d, g \in \mathbb{N}$ *and* $\# \in \{all, \emptyset, lin, 0, clone\}$.

Proof. We only prove the proposition for $\Gamma(g, d)$ here because the proofs for the other sets are very similar. Let $s, t \in \Gamma(g, d)$ and $\lambda \in [0, 1]$. Let further $E_1, \ldots, E_g \in \mathrm{Eff}_d$. By the choice of s and t, we know that the $s_i E_i + (1 - s_i)I/2$ and the $t_i E_i + (1 - t_i)I/2$ are each jointly measurable and give rise to joint POVMs $R_{i_1,\ldots,i_g}$ and $R'_{i_1,\ldots,i_g}$, respectively. Then,

$$\lambda R_{i_1,\ldots,i_g} + (1-\lambda) R'_{i_1,\ldots,i_g}$$

is again a POVM and it can easily be verified that

$$\begin{aligned} &\sum_{\substack{i_j \in [2] \\ j \in [g] \setminus \{u\}}} \lambda R_{i_1,\ldots,i_{u-1},1,i_{u+1},\ldots,i_g} + (1-\lambda) R'_{i_1,\ldots,i_{u-1},1,i_{u+1},\ldots,i_g} \\ &= [\lambda s_u + (1-\lambda) t_u] E_u + [1 - (\lambda s_u + (1-\lambda) t_u)] I/2. \end{aligned}$$

As the effects were arbitrary, this proves the assertion for $\Gamma(g, d)$. □

Remark III.3. Using convexity, it can easily be seen that $(1/g, \ldots, 1/g) \in \Gamma^{\#}(g, d)$, *where* $\# \in \{all, \emptyset, lin, 0, clone\}$. *It can be seen that the standard basis vector* e_i *is in each of the sets for* $i \in [g]$. *The above statement then follows by convexity. See also Ref. 22 for an intuitive argument.*

In the next proposition, we collect some relations between the different sets we have defined (Fig. 1).

Proposition III.4. Let $g, d \in \mathbb{N}$. *Then the following inclusions are true:*

1. $\Gamma(g, d) \subseteq \Gamma^{all}(g, d)$;

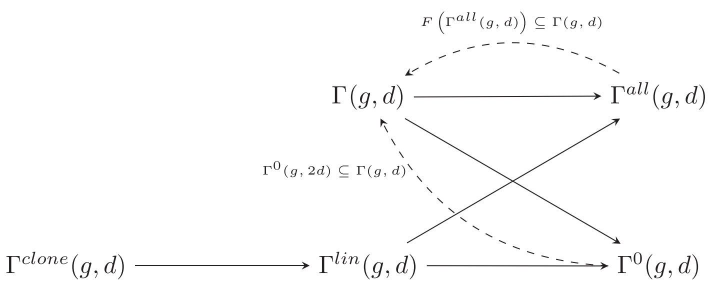

FIG. 1. The different inclusions for the sets $\Gamma^{\#}$, $\# \in \{\emptyset, lin, all, 0, clone\}$ proven in Proposition III.4. Full arrows represent inclusions of sets, while dashed arrows represent special conditions.

2. $\Gamma^{lin}(g, d) \subseteq \Gamma^{all}(g, d)$;
3. $\Gamma(g, d) \subseteq \Gamma^0(g, d)$;
4. $\Gamma^{lin}(g, d) \subseteq \Gamma^0(g, d)$;
5. $\Gamma^{clone}(g, d) \subseteq \Gamma^{lin}(g, d)$;
6. $\Gamma^0(g, 2d) \subseteq \Gamma(g, d)$;
7. $F\left(\Gamma^{all}(g, d)\right) \subseteq \Gamma(g, d)$, *where*

$$F : [0, 1]^g \to [0, 1]^g, \qquad F(s_1, \ldots, s_g) = \left(\frac{s_1}{2 - s_1}, \ldots, \frac{s_g}{2 - s_g}\right).$$

Proof. The first two assertions are true since we restrict the trivial measurements we are mixing with in both cases. The third assertion follows in the same way, but this time compatibility has to hold for less states. The fourth assertion follows since $\mathrm{tr}[E_i]/d = 1/2$ for the effects considered for $\Gamma^0(g, d)$. For the fifth claim, let $s \in \Gamma^{clone}(g, d)$ be an arbitrary scaling g-tuple and consider quantum effects $E_1, \ldots, E_g \in \mathrm{Eff}_d$. Define, for every bit-string b of length g

$$F_b := \mathcal{T}(E_1^{(b_1)} \otimes E_2^{(b_2)} \otimes \cdots \otimes E_g^{(b_g)}),$$

where we set $E_i^{(1)} = E_i$ and $E_i^{(0)} = I - E_i$ and $\mathcal{T}$ is a map as in (2). Since the map $\mathcal{T}$ is (completely) positive, we have that all the operators F_b are positive semidefinite. Moreover, the marginals can be computed as follows:

$$\sum_{b \in \{0,1\}^g,\, b_i = 1} F_b = \mathcal{T}(I^{\otimes i-1} \otimes E_i \otimes I^{\otimes n-i}) = s_i E_i + (1 - s_i)\frac{\mathrm{tr}[E_i]}{d} I =: E_i',$$

which shows that the mixed effects E_i' are compatible, proving the claim.

For the sixth assertion, let $s \in \Gamma^0(g, 2d)$. Then, for a g-tuple of arbitrary $d \times d$ quantum effects E_i, the quantum effects (of size $2d$)

$$s_i[E_i \oplus (I_d - E_i)] + (1 - s_i)\frac{d}{2d} I_{2d}$$

are unbiased ($\mathrm{tr}[E_i \oplus (I_d - E_i)] = d$) and thus compatible. Truncating the above effects to their upper-left corner proves the claim.

Let us now prove the seventh and final claim. It is enough to show that, for any effect $E \in \mathrm{Eff}_d$ and any mixture $E' = sE + (1 - s)aI$ with some trivial effect aI ($a \in [0, 1]$), there is a further mixture $E'' = xE' + (1 - x)bI = yE + (1 - y)I/2$. Working out the relations between the parameters s, x, y, a, b, we find the following two equations:

$$y = xs \quad \text{and} \quad b = \frac{1 - xs - 2xa(1 - s)}{2(1 - x)}.$$

Asking that, for all values of $a \in [0, 1]$, b is also between 0 and 1, we obtain the desired inequality $y \le s/(2 - s)$. Let $(E_1', \ldots, E_g')$ be the compatible effects corresponding to s. Then $E_1', \ldots, E_{j-1}', b_j I, E_{j+1}', \ldots, E_g'$ are compatible as well since we obtain a joint POVM for the effects without E_j' by summing over the jth index and $b_j I$ is a trivial measurement and as such compatible with all effects. Then, we obtain the element $E'' = (x_1 E_1' + (1 - x_1)b_1 I, \ldots, x_g E_g' + (1 - x_g)b_g I)$ from $(E_1', \ldots, E_g')$ by successively taking convex combinations with elements of the form $(E_1', \ldots, E_{j-1}', b_j I, E_{j+1}', \ldots, E_g')$. As convex combinations of compatible tuples stay compatible (see proof of Proposition III.2), we infer that E'' is compatible and the assertion follows. □

Remark III.5. It would be interesting to see if, in general, $\Gamma(g, d') \subseteq \Gamma^{lin}(g, d)$ *for some parameter* d' *which depends on* d.

Now we show that these sets become smaller when we increase the dimension of the effects considered.

Proposition III.6. Let $g, d \in \mathbb{N}$ and $\# \in \{all, \emptyset, lin, 0, clone\}$. Then

$$\Gamma^{\#}(g, d+1) \subseteq \Gamma^{\#}(g, d).$$

Proof. Let us first show the inclusion for Γ^{all}; the proofs for Γ and Γ^0 are almost identical. Let $E_i \in \mathrm{Eff}_d$ for all $i \in [g]$. We can embed these effects into Eff_{d+1} by choosing $E_i' = E_i \oplus 0$. Let $s \in \Gamma^{all}(g, d+1)$. Then there exists an $a \in [0,1]^g$ such that the effects

$$s_i E_i' + (1 - s_i) a_i I_{d+1}, \qquad i \in [g],$$

are compatible. Let $V : \mathbb{C}^d \hookrightarrow \mathbb{C}^{d+1}$ be an isometry such that $V V^*$ is the projection onto the first d entries. It is easy to check that for a POVM $\{R_i\}_{i \in [g]}$ with $R_i \in \mathrm{Eff}_{d+1}$ for all $i \in [g]$, the set $\{V^* R_i V\}_{i \in [g]}$ is again a POVM with elements in Eff_d. Furthermore,

$$V^* \left[s_i E_i' + (1 - s_i) a_i I_{d+1} \right] V = s_i E_i + (1 - s_i) a_i I_d,$$

which implies that the $s_i E_i + (1 - s_i) a_i I_d$ are compatible and therefore $s \in \Gamma^{all}(g, d)$.

To prove the claim for Γ^{lin}, we use the same idea as before, but with the following linear embedding of $d \times d$ matrices into $(d+1) \times (d+1)$ matrices:

$$\Psi : \mathcal{M}_d \to \mathcal{M}_{d+1}, \qquad X \mapsto X \oplus \frac{\mathrm{tr}[X]}{d}.$$

As $\mathrm{tr}[\Psi(X)] = (d+1)/d \, \mathrm{tr}[X]$, the claim follows.

Finally, since Ψ can easily be seen to be completely positive and unital, we can use it together with the embedding V from above to define the cloning map for dimension d through the one for dimension $d+1$. With $\mathcal{T}_{d+1}$ and $\mathcal{T}_d$ being the maps appearing in (2) for $\Gamma^{clone}(g, d+1)$ and $\Gamma^{clone}(g, d)$, respectively, we obtain

$$\mathcal{T}_d(X) := V^* \mathcal{T}_{d+1}(\Psi^{\otimes g}(X)) V.$$

It can then be verified that the map indeed has the desired properties. □

Remark III.7. We would like to point out that the sets $\Gamma(g, d)$ give rise to compatibility criteria, *i.e.*, sufficient conditions for compatibility, *as follows. Let $s \in \Gamma(g, d)$ be such that $s_i > 0$ for all $i \in [g]$. Then, the following implication holds:*

$$\forall i \in [g], \; \frac{1}{s_i} E_i - \frac{1 - s_i}{2 s_i} I_d \in \mathrm{Eff}_d \quad \Longrightarrow \quad (E_1, \ldots, E_g) \text{ compatible.}$$

Indeed, using $s \in \Gamma(g, d)$ and the hypothesis, it follows that the effects

$$s_i \left[\frac{1}{s_i} E_i - \frac{1 - s_i}{2 s_i} I_d \right] + (1 - s_i) \frac{I_d}{2} = E_i$$

are compatible. These criteria become useful if the corresponding SDP is intractable.

IV. FREE SPECTRAHEDRA

In this section, we will review some concepts from the study of free spectrahedra which we will need in the rest of the paper. We will start with the definition of free spectrahedra and their inclusion. Then, we will review the link between spectrahedral inclusion and positivity properties of certain maps. All the theory needed in this work can be found in Refs. 14, 27, and 28.

Let $A \in \left(\mathcal{M}_d^{sa}\right)^g$. The *free spectrahedron at level n* corresponding to this g-tuple of matrices is the set

$$\mathcal{D}_A(n) := \left\{ X \in (\mathcal{M}_n^{sa})^g : \sum_{i=1}^{g} A_i \otimes X_i \leq I_{nd} \right\}.$$

For $n = 1$, this is a usual spectrahedron defined by a *linear matrix inequality*. The *free spectrahedron* is then defined as the (disjoint) union of all these sets, i.e.,

$$\mathcal{D}_A := \bigsqcup_{n \in \mathbb{N}} \mathcal{D}_A(n).$$

A free spectrahedron which we will very often encounter is the *matrix diamond of size g*. It is defined as

$$\mathcal{D}_{\diamond,g}(n) = \left\{ X \in (\mathcal{M}_n^{sa})^g : \sum_{i=1}^{g} \epsilon_i X_i \leq I_n \ \forall \epsilon \in \{-1,+1\}^g \right\}.$$

To see that this is a free spectrahedron, we can take the direct sum of all these constraints. The matrices defining this free spectrahedron are thus diagonal. At level 1, the matrix diamond is the unit ball of the ℓ_1-norm. Therefore, it is obviously bounded. For free spectrahedra, the inclusion $\mathcal{D}_A \subseteq \mathcal{D}_B$ means $\mathcal{D}_A(n) \subseteq \mathcal{D}_B(n)$ for all $n \in \mathbb{N}$. Inclusion at the level of spectrahedra ($n = 1$) does not guarantee inclusion of the free spectrahedra. That is, the implication

$$\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1) \Rightarrow \mathcal{D}_A \subseteq \mathcal{D}_B$$

does not hold in general. However, scaling the set $\mathcal{D}_A$ down, the implication becomes eventually true.

Definition IV.1. Let $\mathcal{D}_A$ be the free spectrahedron defined by $A \in \left(\mathcal{M}_d^{sa}\right)^g$. The inclusion set $\Delta_{\mathcal{D}_A}(k)$ is defined as

$$\Delta_{\mathcal{D}_A}(k) := \left\{ s \in \mathbb{R}_+^g : \forall B \in \left(\mathcal{M}_k^{sa}\right)^g \ \mathcal{D}_A(1) \subseteq \mathcal{D}_B(1) \Rightarrow s \cdot \mathcal{D}_A \subseteq \mathcal{D}_B \right\}.$$

For $\mathcal{D}_A = \mathcal{D}_{\diamond,g}$, we will write $\Delta(g, k)$ for brevity. Here, the set

$$s \cdot \mathcal{D}_A := \left\{ (s_1 X_1, \ldots, s_g X_g) : X \in \mathcal{D}_A \right\}$$

is the (asymetrically) scaled *free spectrahedron.*

Note that the definition above generalizes the one from Ref. 28 by restricting the size of the matrices defining the containing spectrahedra and by allowing non-symmetric scaling; one recovers the definition of inclusion constants from Ref. 28 by considering the largest constant $s \geq 0$ such that

$$\underbrace{(s, \ldots, s)}_{g \text{ times}} \in \Delta_{\mathcal{D}_A}(k), \qquad \forall k \geq 1.$$

As in the case of POVMs, we are also interested in the inclusion constant set where we restrict to inclusion into free spectrahedra defined by traceless matrices.

Definition IV.2. Let $\mathcal{D}_A$ be a spectrahedron defined by $A \in \left(\mathcal{M}_d^{sa}\right)^g$. The traceless-restricted *inclusion set $\Delta^0_{\mathcal{D}_A}(d)$ of $\mathcal{D}_A$ is defined as*

$$\Delta^0_{\mathcal{D}_A}(k) := \{ s \in \mathbb{R}_+^g : \forall C \in (\mathcal{M}_k^{sa})^g \text{ s.t. } \forall i \in [g],\ \mathrm{tr}[C_i] = 0, \\ \mathcal{D}_A(1) \subseteq \mathcal{D}_C(1) \implies s \cdot \mathcal{D}_A \subseteq \mathcal{D}_C \}.$$

For $\mathcal{D}_A = \mathcal{D}_{\diamond,g}$, we will again write $\Delta^0(g, k)$ for brevity.

The next proposition shows that both inclusion sets we have defined are convex.

Proposition IV.3. Let $A \in \left(\mathcal{M}_d^{sa}\right)^g$. Both $\Delta_{\mathcal{D}_A}(k)$ and $\Delta^0_{\mathcal{D}_A}(k)$ are convex.

Proof. Let $B \in \left(\mathcal{M}_k^{sa}\right)^g$ and $X \in \mathcal{D}_A$. Furthermore, let $s, t \in \Delta_{\mathcal{D}_A}(k)$ and $\lambda \in [0, 1]$. The assumptions on s, t yield

$$\sum_{i=1}^{g} B_i \otimes (\lambda s_i + (1-\lambda)t_i)X = \lambda \sum_{i=1}^{g} B_i \otimes s_i X + (1-\lambda) \sum_{j=1}^{g} B_j \otimes t_j X \le I.$$

This proves the first assertion because B was arbitrary. The second assertion follows in a very similar manner. □

The inclusion of spectrahedra can be related to positivity properties of a certain map. Let $A \in (\mathcal{M}_D^{sa})^g$ and $B \in (\mathcal{M}_d^{sa})^g$ define the free spectrahedra $\mathcal{D}_A$ and $\mathcal{D}_B$, respectively. Let $\Phi : \mathcal{OS}_A \to \mathcal{M}_d$ be the unital map defined as

$$\Phi : A_i \mapsto B_i, \qquad \forall i \in [g].$$

Then, we can find a one-to-one relation between properties of Φ and the inclusion of the free spectrahedra at different levels. This has been proven in Ref. 27 (Theorem 3.5) for real spectrahedra and we include a proof in the complex case for convenience.

Lemma IV.4. Let $A \in (\mathcal{M}_D^{sa})^g$ and $B \in (\mathcal{M}_d^{sa})^g$. Furthermore, let $\mathcal{D}_A(1)$ be bounded. Then $\mathcal{D}_A(n) \subseteq \mathcal{D}_B(n)$ holds if and only if Φ as given above is n-positive. In particular, $\mathcal{D}_A \subseteq \mathcal{D}_B$ if and only if Φ is completely positive.

Proof. The "only if" direction is true by the unitality and n-positivity of Φ. For the "if" direction, let $Y \in \mathcal{M}_n^{sa}(\mathcal{OS}_A)$. Without loss of generality, we can assume $I_D, A_1, \ldots, A_g$ to be linearly independent. Then

$$Y = I_D \otimes X_0 - \sum_{i=1}^{g} A_i \otimes X_i$$

for $(X_0, \ldots, X_g) \in \mathcal{M}_n^g$. We claim that $X_0, \ldots, X_g$ are self-adjoint. Then $(I_D \otimes e_i^*)(Y - Y^*)(I_D \otimes e_j) = 0$ for all $i, j \in [n]$ and an orthonormal basis $\{e_i\}_{i=1}^n$ of $\mathbb{C}^n$ if and only if $\langle e_i, (X_l - X_l^*)e_j\rangle$ for all $i, j \in [n]$ and for all $l \in [g] \cup \{0\}$. This proves the claim. If $Y \geq 0$, it holds that $X_0 \geq 0$. Let us assume that this is not the case. Then there exists an $x \in \mathbb{C}^n$ such that $\langle x, X_0 x\rangle < 0$. Positivity of Y yields

$$-\sum_{i=1}^{g} \langle x, X_i x\rangle A_i > 0.$$

Therefore, $\lambda(\langle x, X_1 x\rangle, \ldots, \langle x, X_g x\rangle) \in \mathcal{D}_A(1)$ for all $\lambda \geq 0$. This contradicts the assumption that $\mathcal{D}_A(1)$ is bounded. Let us now assume that $Y \geq 0$ and that $X_0 > 0$. Then $(\Phi \otimes Id_n)Y \geq 0$ because $X_0^{-1/2} X X_0^{-1/2} \in \mathcal{D}_A(n) \subseteq \mathcal{D}_B(n)$. For $Y \geq 0$ and $X_0 \geq 0$, positivity of $(\Phi \otimes Id_n)Y$ follows from exchanging X_0 by $X_0 + \epsilon I_n$, $\epsilon > 0$, and letting ϵ go to zero. □

Remark IV.5. The complete positivity of Φ can be checked using an SDP[20,27]*. Therefore, the inclusion problem at the level of free spectrahedra is efficiently solvable. This is not necessarily the case for the usual spectrahedra at level 1 because checking the positivity of a linear map is in general a hard problem (the set of positive maps between matrix algebras is dual to the set of separable states, and deciding weak membership into the latter set is known to be NP-hard*[18]*). Seeing the free spectrahedral inclusion problem as a* relaxation *of the corresponding problem for (level 1) spectrahedra is a very useful idea in optimization; see Refs. 3 and 27.*

Using the previous lemma, we obtain a useful corollary if we assume that $\mathcal{D}_A(1)$ is bounded, which is enough for us.

Corollary IV.6. Let $A \in (\mathcal{M}_D^{sa})^g$ and $B \in (\mathcal{M}_d^{sa})^g$. Moreover, let $\mathcal{D}_A(1)$ be bounded. Then $\mathcal{D}_A(d) \subseteq \mathcal{D}_B(d)$ if and only if $\mathcal{D}_A \subseteq \mathcal{D}_B$.

Proof. From Lemma IV.4, $\mathcal{D}_A(d) \subseteq \mathcal{D}_B(d)$ is equivalent to Φ being d-positive. Since Φ maps to $\mathcal{M}_d$, this is equivalent to the complete positivity of the map [see Ref. 39 (Theorem 6.1)]. The claim then follows by another application of Lemma IV.4. □

Remark IV.7. The result of Corollary IV.6 without the boundedness assumption has appeared before in Ref. 28 (Lemma 2.3) for real spectrahedra with a longer proof. Their proof carries over to the complex setting. Therefore, the boundedness assumption is not necessary, but it shortens the proof considerably.

V. SPECTRAHEDRAL INCLUSION AND JOINT MEASURABILITY

In this section, we establish the link between joint measurability of effects and the inclusion of free spectrahedra. The main result of this work is Theorem V.3 which we prove at the end of this section. It connects the inclusion of the matrix diamond into a spectrahedron with the joint measurability of the quantum effects defining this spectrahedron. Before we can prove the theorem, we will need two lemmas concerning compressed versions of a (free) spectrahedron.

Lemma V.1. Let $k \in \mathbb{N}$, $1 \leq k \leq d$, and let $V : \mathbb{C}^k \hookrightarrow \mathbb{C}^d$ be an isometry. For $A \in \left(\mathcal{M}_d^{sa}\right)^g$, it holds that $\mathcal{D}_A \subseteq \mathcal{D}_{V^*AV}$.

Proof. Let $X \in \mathcal{D}_A(n)$. Then from the definition,

$$\sum_{i=1}^{g} A_i \otimes X_i \leq I_{dn}.$$

Multiplying the equation by $V \otimes I_n$ from the right and by its adjoint from the left, it follows that

$$\sum_{i=1}^{g} V^* A_i V \otimes X_i \leq I_{kn}.$$

Here, we have used that V is an isometry and that the map $Y \mapsto W^* Y W$ for matrices Y, W of appropriate dimensions is completely positive. □

Lemma V.2. Let $k \in \mathbb{N}$, $1 \leq k \leq d$, and let $V : \mathbb{C}^k \hookrightarrow \mathbb{C}^d$ be an isometry. For $A \in \left(\mathcal{M}_d^{sa}\right)^g$, it holds that

$$\mathcal{D}_A(k) = \bigcap_{V:\mathbb{C}^k \hookrightarrow \mathbb{C}^d \text{ isometry}} \mathcal{D}_{V^*AV}(k). \tag{3}$$

Proof. Let us denote the right-hand side of Eq. (3) by $\mathcal{C}$. From Lemma V.1, it follows that $\mathcal{D}_A(k) \subseteq \mathcal{C}$. For the reverse inclusion, let $X \in \mathcal{C}$. This implies especially $X \in \left(\mathcal{M}_k^{sa}\right)^g$. We write

$$Y := \sum_{i=1}^{g} A_i \otimes X_i - I_{dk}.$$

To prove the assertion, we need to show that $Y \geq 0$. Let $y \in \mathbb{C}^d \otimes \mathbb{C}^k$ be a unit vector with Schmidt decomposition

$$y = \sum_{i=1}^{k} \sqrt{\lambda_i} e_i \otimes f_i,$$

where $\{e_i\}_{i=1}^k$ and $\left\{f_j\right\}_{j=1}^k$ are orthonormal families in $\mathbb{C}^d$ and $\mathbb{C}^k$, respectively. Furthermore, $\lambda_i \geq 0$ and such that $\sum_{i=1}^k \lambda_i = 1$. We then have

$$\langle y, Yy \rangle = \sum_{i,j=1}^{k} \sqrt{\lambda_i \lambda_j} \langle (e_i \otimes f_i), Y(e_j \otimes f_j) \rangle.$$

Let $\Omega = \sum_{i=1}^k g_i \otimes g_i$ be an unnormalized maximally entangled state with an orthonormal basis $\{g_i\}_{i=1}^k$ of $\mathbb{C}^k$. Moreover, let $V : \mathbb{C}^k \to \mathbb{C}^d$ and $Q : \mathbb{C}^k \to \mathbb{C}^k$ be defined as

$$V = \sum_{i=1}^k e_i g_i^*, \qquad Q = \sum_{j=1}^k \sqrt{\lambda_j} f_j g_j^*.$$

Then, $V : \mathbb{C}^k \hookrightarrow \mathbb{C}^d$ is an isometry. Therefore, $(V^* \otimes I_k)Y\,(V \otimes I_k) \geq 0$ by assumption, as $X \in \mathcal{C}$. Hence,

$$\langle y, Yy\rangle = \langle \Omega, (I_k \otimes Q^*)(V^* \otimes I_k)Y(V \otimes I_k)(I_k \otimes Q)\Omega\rangle \geq 0.$$

Thus, Y is positive semidefinite because y was arbitrary. □

Theorem V.3. *Let $E \in \left(\mathcal{M}_d^{sa}\right)^g$ and let $2E - I := (2E_1 - I_d, \ldots, 2E_g - I_d)$. We have*

1. *$\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_{2E-I}(1)$ if and only if $E_1, \ldots, E_g$ are quantum effects.*
2. *$\mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{2E-I}$ if and only if $E_1, \ldots, E_g$ are jointly measurable quantum effects.*
3. *$\mathcal{D}_{\diamond,g}(k) \subseteq \mathcal{D}_{2E-I}(k)$ for $k \in [d]$ if and only if for any isometry $V : \mathbb{C}^k \hookrightarrow \mathbb{C}^d$, the induced compressions $V^*E_1V, \ldots, V^*E_gV$ are jointly measurable quantum effects.*

Proof. Let us start with the first point. Since $\mathcal{D}_{\diamond,g}(1)$ is a convex polytope, we need to check inclusion only at extreme points. That means that the first assertion holds if and only if $\pm e_i \in \mathcal{D}_{2E-I}$ for all $i \in [g]$, where $\{e_i\}_{i=1}^g$ is the standard basis in $\mathbb{R}^g$. We have

$$e_i \in \mathcal{D}_E(1) \iff 2E_i - I \leq I \iff E_i \leq I,$$
$$-e_i \in \mathcal{D}_E(1) \iff -(2E_i - I) \leq I \iff E_i \geq 0,$$

proving the first claim.

We now characterize the free spectrahedral inclusion from the second point. In the following, we will identify diagonal matrices with vectors and the subalgebra of diagonal $2^g \times 2^g$-matrices with $\mathbb{C}^{2^g}$. The operator system associated with $\mathcal{D}_{\diamond,g}$ is

$$\mathbb{C}^{2^g} \supseteq \mathcal{OS}_{\diamond,g} := \operatorname{span}\{v_0, v_1, \ldots, v_g\},$$

where, indexing the 2^g coordinates by sign vectors $\varepsilon \in \{\pm 1\}^g$,

$$v_0(\varepsilon) = 1,$$
$$v_i(\varepsilon) = \varepsilon(i), \qquad \forall i \in [g].$$

Here, $\varepsilon(i)$ is the ith entry of the vector ε. The dimension of this operator system is $g + 1$. We define a map $\Phi : \mathcal{OS}_{\diamond,g} \to \mathcal{M}_d$ by

$$v_0 \mapsto I,$$
$$v_i \mapsto 2E_i - I \qquad , \forall i \in [g].$$

The spectrahedral inclusion $\mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{2E-I}$ holds if and only if the map Φ is completely positive (Lemma IV.4). If this is the case, Arveson's extension theorem [see Ref. 39 (Theorem 6.2) for a finite-dimensional version] guarantees the existence of a completely positive extension $\tilde{\Phi}$ of this map to the whole algebra $\mathbb{C}^{2^g}$ because $\mathcal{OS}_{\diamond,g}$ is an operator system. As $\mathbb{C}^{2^g}$ is a commutative matrix algebra, it is enough to show that the extension is positive [see Ref. 39 (Theorem 3.11)]. To find such an extension $\tilde{\Phi} : \mathbb{C}^{2^g} \to \mathcal{M}_d$, we consider the basis $(g_\eta)_{\eta \in \{\pm 1\}^g}$ of the vector space $\mathbb{C}^{2^g}$ which is defined as follows:

$$g_\eta(\varepsilon) = \mathbf{1}_{\varepsilon=\eta} \geq 0.$$

Here, $\mathbf{1}_{\varepsilon=\eta} = 1$ if $\varepsilon = \eta$ and zero otherwise. Let us write $G_\eta := \tilde{\Phi}(g_\eta)$; since the g_η are positive, the (complete) positivity of $\tilde{\Phi}$ is equivalent to $G_\eta \geq 0$, for all η.

We have, for all $\varepsilon \in \{\pm 1\}^g$,

$$1 = v_0(\varepsilon) = \sum_\eta \mathbf{1}_{\varepsilon=\eta} = \sum_\eta g_\eta(\epsilon)$$

and thus we can rewrite

$$\tilde{\Phi}(v_0) = I \iff \sum_\eta G_\eta = I.$$

We also have

$$v_i(\varepsilon) = \varepsilon(i) = 2\mathbf{1}_{\varepsilon(i)=+1} - 1 = 2 \sum_{\eta\,:\,\eta(i)=+1} \mathbf{1}_{\varepsilon=\eta} - 1 = 2 \sum_{\eta\,:\,\eta(i)=+1} g_\eta(\epsilon) - \sum_\eta g_\eta(\epsilon)$$

and thus we have, for all $i \in [g]$,

$$\tilde{\Phi}(v_i) = 2E_i - I \iff \sum_{\eta\,:\,\eta(i)=+1} G_\eta = E_i.$$

Collecting all these facts, we have shown that the map Φ extends to a (completely) positive map on the whole $\mathbb{C}^{2^g}$ if and only if there exist operators $(G_\eta)_{\eta\in\{\pm1\}^g}$ such that

$$\begin{aligned} \forall \eta \in \{\pm 1\}^g, \quad & G_\eta \geq 0, \\ & \sum_\eta G_\eta = I, \\ \forall i \in [g], \quad & \sum_{\eta\,:\,\eta(i)=+1} G_\eta = E_i, \end{aligned}$$

but these are precisely the conditions for the joint measurability of the effects $E_1, \ldots, E_g$ and we are done with the second point. For the third assertion, it follows from Lemma V.2 that $\mathcal{D}_{\diamond,g}(k) \subseteq \mathcal{D}_{2E-I}(k)$ holds if and only if $\mathcal{D}_{\diamond,g}(k) \subseteq \mathcal{D}_{2V^*EV-I}(k)$ for any isometry $V : \mathbb{C}^k \hookrightarrow \mathbb{C}^d$. Furthermore, Corollary IV.6 asserts that this is equivalent to $\mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{2V^*EV-I}$ for all isometries V as above. The claim then follows from the second assertion of this theorem. □

Remark V.4. The fact that the second point of the theorem above implies the third point, read on the quantum effects side of the equivalence, is a well-known fact[36]*: compressions of jointly measurable effects are jointly measurable.*

Remark V.5. If $E \in (\mathcal{M}_d^{sa})^g$ is a g-tuple of pairwise commuting matrices, then $\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_{2E-I}(1)$ implies $\mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{2E-I}$. This is true because the effects on the right-hand side generate a commutative matrix subalgebra. Inclusion at level one then implies that the corresponding map Φ is positive. As its range is contained in a commutative matrix algebra, this also implies that Φ is completely positive [see Ref. 39 *(Theorem 3.9)*]*, which then yields the inclusion at the level of free spectrahedra. This recovers the well-known result from quantum information theory that pairwise commuting effects are jointly measurable.*

Remark V.6. We can recover another result from quantum information theory, namely, that effects of the form aI, $a \in [0, 1]$, are trivially compatible with any effect. This corresponds to the fact that

$$\mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{2E-I} \iff \mathcal{D}_{\diamond,g+1} \subseteq \mathcal{D}_{(2E-I,\alpha I)} \tag{4}$$

for any $E \in (\mathcal{M}_d^{sa})^g$. Here, we write $\alpha = 2a - 1$, i.e., $\alpha \in [-1, 1]$. This can best be seen at the level of maps. It is easy to see that the v_i defining $\mathcal{D}_{\diamond,g}$ (c.f. proof of Theorem V.3) can be written as

$$v_i = \prod_{j=1}^{i-1} I_2 \otimes \mathrm{diag}[+1, -1] \otimes \prod_{j=i+1}^{g} I_2.$$

Let Φ_g be the map corresponding to the left-hand side and Φ_{g+1} be the one corresponding to the right-hand side of Eq. (4)*. For the "only if"-direction, we can simply define $\Phi_g(A) = \Phi_{g+1}(A \otimes I_2)$, where $A \in \mathbb{C}^{2^g}$. For the "if"-direction, we define the linear map $\Psi : \mathbb{C}^2 \to \mathbb{C}$ as*

$$\Psi:(1,0)\mapsto\frac{\alpha+1}{2},\qquad \Psi:(0,1)\mapsto\frac{1-\alpha}{2}.$$

This map is unital, positive, and therefore also completely positive. We can then set $\Phi_{g+1}=\Phi_g\otimes\Psi$. *It can then be checked using the above expression for the* v_i *that this map has indeed the desired properties.*

Theorem V.7. *It holds that* $\Gamma(g,d)=\Delta(g,d)$ *and that* $\Gamma^0(g,d)=\Delta^0(g,d)$.

Proof. Let $s\in\mathbb{R}^g$. Then $s\in\Gamma(g,d)$ if and only if $s_1E_1+(1-s_1)I/2,\ldots,s_gE_g+(1-s_g)I/2$ are jointly measurable for all effects $E_1,\ldots E_g\in\mathrm{Eff}_d$. It can easily be seen that

$$s\cdot\mathcal{D}_{\diamond,g}\subseteq\mathcal{D}_{2E-I}\iff\mathcal{D}_{\diamond,g}\subseteq\mathcal{D}_{s\cdot(2E-I)}\iff\mathcal{D}_{\diamond,g}\subseteq\mathcal{D}_{2E'-I},$$

where $E_i'=s_iE_i+(1-s_i)I/2$ for all $i\in[g]$. Therefore, Theorem V.3 implies that $s\in\Gamma(g,d)$ is equivalent to the implication $\mathcal{D}_{\diamond,g}(1)\subseteq\mathcal{D}_{2E-I}(1)\Rightarrow s\cdot\mathcal{D}_{\diamond,g}\subseteq\mathcal{D}_{2E-I}$ for all $E\in(\mathcal{M}_d^{sa})^g$. This is equivalent to $s\in\Delta(g,d)$ because $X\mapsto 2X-I$ is a bijection on $\mathcal{M}_d^{sa}$. The second assertion follows since $X\mapsto 2X-I$ is a bijection between $\left\{X\in\mathcal{M}_d^{sa}:\mathrm{tr}[X]=d/2\right\}$ and $\left\{A\in\mathcal{M}_d^{sa}:\mathrm{tr}[A]=0\right\}$. □

VI. LOWER BOUNDS FROM CLONING

In this section, we provide, using known facts about the set Γ^{clone}, lower bounds for Γ^{lin} and Γ. We start by recalling the main results from the theory of *symmetric* cloning.

Theorem VI.1 [**Ref. 33 (Theorem 7.2.1)**]. *For a quantum channel* $\mathcal{T}:\mathcal{M}_d^{\otimes N}\to\mathcal{M}_d^{\otimes g}$, *consider the quantities*

$$\mathcal{F}_{c,1}(\mathcal{T})=\inf_{j\in[g]}\inf_{\sigma\text{ pure}}\mathrm{tr}[\sigma^{(j)}\mathcal{T}(\sigma^{\otimes N})],$$

where $\sigma^{(j)}=I^{\otimes(j-1)}\otimes\sigma\otimes I^{\otimes(g-j)}\in\mathcal{M}_d^{\otimes g}$ *and*

$$\mathcal{F}_{c,all}=\inf_{\sigma\text{ pure}}\mathrm{tr}[\sigma^{\otimes g}\mathcal{T}(\sigma^{\otimes N})].$$

These quantities are both maximized by the optimal quantum cloner

$$\mathcal{T}(\rho)=\frac{d[N]}{d[g]}S_g(\rho\otimes I)S_g.$$

Here, $d[g]=\binom{d+g-1}{g}$ *is the dimension of the symmetric subspace* $\vee^g\mathbb{C}^d\subseteq(\mathbb{C}^d)^{\otimes g}$ *and* S_g *is the corresponding orthogonal projection.*

From Ref. 46 [Eq. (3.7)], we know further that

$$\tilde{\rho}:=\mathrm{tr}_{i^c}[\mathcal{T}(\rho)]=\gamma\rho+(1-\gamma)I/d,$$

where $\gamma=(g+d)/(g(1+d))$ (for $N=1$). Here, tr_{i^c} means tracing out all systems but the ith one. Going to the dual picture, we can compute that for $E\in\mathcal{M}_d$,

$$\mathrm{tr}[\tilde{\rho}E]=\mathrm{tr}\left[\rho\left(\gamma E+(1-\gamma)\frac{\mathrm{tr}E}{d}I\right)\right].$$

We can therefore identify $E'=\gamma E+(1-\gamma)I_E$, where I_E is the trivial effect $\mathrm{tr}\,\{E\}/dI$ depending on E. Therefore, γ is a lower bound on the joint measurability of a family of effects, with the POVM

$$\mathcal{T}^*\left(\left\{E_1^{(i)}\right\}_i\otimes\cdots\otimes\left\{E_g^{(j)}\right\}_j\right)$$

being the joint observable.

Inserting the expression for γ from the symmetric cloning bounds and using Proposition III.4, we obtain the following result; note that below, the second quantity is always larger than the third one.

Proposition VI.2. For all $g, d \geq 2$,

$$\frac{g+d}{g(1+d)}\underbrace{(1,1,\ldots,1)}_{g\text{ times}} \in \Gamma^{clone}(g,d) \subseteq \Gamma^{lin}(g,d) \subseteq \Gamma^{all}(g,d),$$

$$\frac{g+2d}{g(1+2d)}\underbrace{(1,1,\ldots,1)}_{g\text{ times}} \in \Gamma^{clone}(g,2d) \subseteq \Gamma^{0}(g,2d) \subseteq \Gamma(g,d) \subseteq \Gamma^{all}(g,d),$$

$$\frac{g+d}{g+d(2g-1)}\underbrace{(1,1,\ldots,1)}_{g\text{ times}} \in F(\Gamma^{clone}(g,d)) \subseteq F(\Gamma^{all}(g,d)) \subseteq \Gamma(g,d).$$

In the general, non-symmetric case, the exact form of the set $\Gamma^{clone}(g,d)$ has been computed, by different methods, in Refs. 32 and 42; the following restatement of the optimal cloning probabilities is taken from the former reference.

Theorem VI.3 [**Ref.** 32 **(Theorem 1, Sec. 2.3)**]. *For any $g, d \geq 2$,*

$$\Gamma^{clone}(g,d) = \left\{ s \in [0,1]^g \,:\, (g+d-1)\left[g - d^2 + d + (d^2-1)\sum_{i=1}^{g} s_i\right] \leq \left(\sum_{i=1}^{g}\sqrt{s_i(d^2-1)+1}\right)^2 \right\}. \tag{5}$$

Using the variables $t_i := s_i(d^2-1)+1 \in [1, d^2]$, we have the simpler expression

$$\Gamma^{clone}(g,d) = \left\{ s \in [0,1]^g \,:\, \|t\|_1 - \frac{\|t\|_{1/2}}{g+d-1} \leq d(d-1) \right\},$$

where $\|\cdot\|_p$ denotes the ℓ_p-quantity on $\mathbb{R}^g$: $\|t\|_p = (\sum_{i=1}^{g} |t_i|^p)^{1/p}$.

Proof. The formula is exactly Eq. (5) from Ref. 32, after the change of variables $F_i = s_i + (1 - s_i)/d$, for all $i \in [g]$. □

Remark VI.4. Note that the symmetric cloning optimal probability is recovered by setting $s_1 = s_2 = \cdots = s_g$ in the result above, yielding the maximal value

$$s_{\max} = \frac{g+d}{g(d+1)}.$$

Remark VI.5. In the regime $d \to \infty$, the left-hand side of (5) behaves like $d^3(\|s\|_1 - 1)$, whereas the right-hand side behaves like $d^2\|s\|_{1/2}$. Hence, asymptotically, the achievable cloning probabilities should satisfy $\sum_i s_i \leq 1$; the set of such values is the probability simplex, i.e., the convex hull of the points $\{e_i\}_{i=1}^{g}$, where e_i is the basis vector having a 1 in position i and zeros elsewhere.

We discuss next the special cases of pairs and triplets, i.e., $g = 2, 3$. The case most studied for (asymmetric) cloning is the $g = 2$ case [see, e.g., Ref. 13 or the more recent Ref. 19 (Theorem 3)]. We plot in the left panel of Fig. 2 the sets $\Gamma^{clone}(2, d)$ for various values of d, as subsets of $\Gamma(2, d) = QC_2$.

Proposition VI.6. For all $d \geq 2$, we have

$$\Gamma^{clone}(2,d) = \{(s,t) \in [0,1]^2 \,:\, s + t - \frac{2}{d}\sqrt{(1-s)(1-t)} \leq 1\}.$$

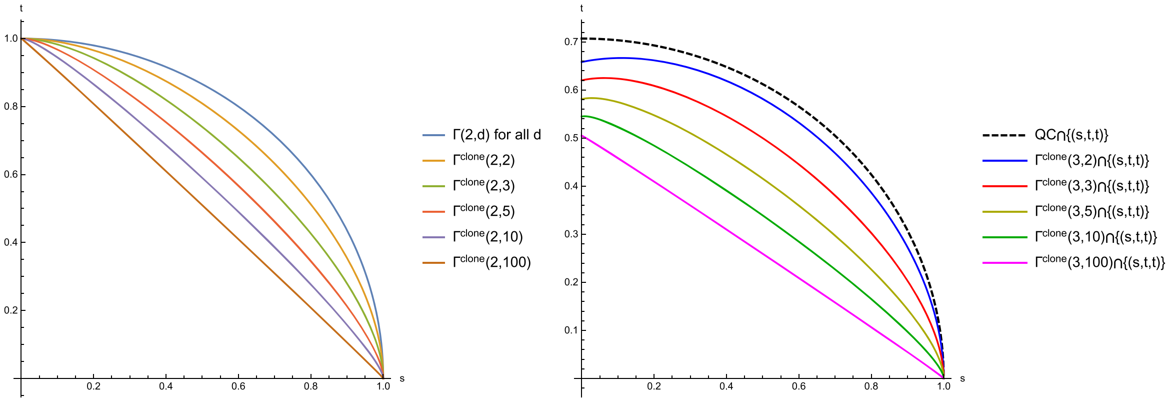

FIG. 2. Left: the sets $\Gamma^{clone}(2, d)$ for $d = 2, 3, 5, 10, 100$ as subsets of $\Gamma(2, d)$, which is a quarter-circle for all $d \geq 2$ (see Corollary VIII.9). Right: the cuts $\Gamma^{clone}(3, d) \cap \{(s, t, t)\}$ for $d = 2, 3, 5, 10, 100$.

Proof. To see that the condition above is equivalent to Eq. (5) for $g = 2$, one can solve both for t and show that the answer is the following:

$$t \leq 1 - s - \frac{2(1-s)}{d^2} + \frac{2\sqrt{1 - s + (d^2-1)s(1-s)}}{d^2}.$$

□

The case $g = 3$ is also worth mentioning since one can obtain manageable expressions for the set $\Gamma^{clone}(3, d)$. In the right panel of Fig. 2, we plot the slice $\Gamma^{clone}(3, d) \cap \{(s, t, t)\}$ for various values of d (this corresponds to asking that the "quality" of the second and third clones is identical) against the Euclidean ball (see Sec. VII B for the relevance of the quarter-circle).

Proposition VI.7. For all $d \geq 2$, we have, either in explicit or in the parametric form[31]

$$\Gamma^{clone}(3, d) = \Big\{(s, t, u) \in [0, 1]^3 \,:\, (d+2)\big[3 - d^2 + d + (d^2-1)(s+t+u)\big] \leq \left(\sqrt{(d^2-1)s+1} + \sqrt{(d^2-1)t+1} + \sqrt{(d^2-1)u+1}\right)^2\Big\}$$

$$= \Big\{\left(1 - b^2 - c^2 - \frac{2bc}{d+1}, 1 - c^2 - a^2 - \frac{2ca}{d+1}, 1 - a^2 - b^2 - \frac{2ab}{d+1}\right) : a^2 + b^2 + c^2 + 2(ab+bc+ca)/d \leq 1,\ a, b, c \geq 0\Big\}.$$

VII. LOWER BOUNDS FROM FREE SPECTRAHEDRA

A. Dimension dependent and symmetric lower bounds

This part basically reproduces the Proof of Theorem 1.4 in Ref. 28 and shows that making minor changes, the proof also works in the case where the spectrahedra are given in terms of complex instead of real matrices. Note that in this case, we obtain an inclusion constant of $2d$ instead of merely d. Let us first recall a lemma from Ref. 28, which was proved there for real matrices but carries over without change.

Lemma VII.1 (**Lemma 8.2 from Ref. 28**). *Suppose $T = (T_{j,l})$ is a $k \times k$ block matrix with blocks of equal size. If $\left\|T_{j,l}\right\|_\infty \leq 1$ for every $j, l \in [k]$, then $\|T\|_\infty \leq k$.*

Proposition VII.2. Let $A \in \left(\mathcal{M}_D^{sa}\right)^g$ and $B \in \left(\mathcal{M}_d^{sa}\right)^g$. Suppose further that $-\mathcal{D}_A \subseteq \mathcal{D}_A$ and that $\mathcal{D}_A(1)$ is bounded. If $\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1)$, then $\mathcal{D}_A \subseteq 2d\mathcal{D}_B$.

Proof. Fix some level n and consider $\{e_l\}_{l=1}^n$, the standard orthonormal basis for $\mathbb{C}^n$. Fix $1 \leq s \neq t \leq n$ and set $p_{s,t}^{\pm} = 1/\sqrt{2}(e_s \pm e_t) \in \mathbb{C}^n$, $\phi_{s,t}^{\pm} = 1/\sqrt{2}(e_s \pm \mathrm{i} e_t) \in \mathbb{C}^n$. Furthermore, let

$$P_{s,t}^{\pm} = I_D \otimes p_{s,t}^{\pm}, \qquad \Phi_{s,t}^{\pm} = I_D \otimes \phi_{s,t}^{\pm}.$$

Then $\left(P_{s,t}^{\pm}\right)^* P_{s,t}^{\pm} = I_D = \left(\Phi_{s,t}^{\pm}\right)^* \Phi_{s,t}^{\pm}$. Let $M \in \mathcal{M}_n^{sa}$, $C \in \mathcal{M}_D$. Then

$$\left(P_{s,t}^{+}\right)^*(C \otimes M)P_{s,t}^{+} - \left(P_{s,t}^{-}\right)^*(C \otimes M)P_{s,t}^{-} = 2C \otimes \mathrm{Re}(M)_{s,t},$$
$$\left(\Phi_{s,t}^{+}\right)^*(C \otimes M)\Phi_{s,t}^{+} - \left(\Phi_{s,t}^{-}\right)^*(C \otimes M)\Phi_{s,t}^{-} = -2C \otimes \mathrm{Im}(M)_{s,t}.$$

Let $X \in \mathcal{D}_A(n)$ and let $Z = \sum_j A_j \otimes X_j$. By hypothesis, $-X \in \mathcal{D}_A(n)$ as well; thus, $\pm Z \leq I_{Dn}$. By the above calculations, also $\pm\left(P_{s,t}^{\pm}\right)^* Z P_{s,t}^{\pm} \leq I_D$. Therefore, $\pm\left(p_{s,t}^{\pm}\right)^* X p_{s,t}^{\pm} \in \mathcal{D}_A(1)$ for all $s \neq t$. Convexity of $\mathcal{D}_A(1)$ together with the above implies $\pm\mathrm{Re}(X)_{s,t} = \pm(\mathrm{Re}(X_1)_{s,t}, \ldots, \mathrm{Re}(X_g)_{s,t}) \in \mathcal{D}_A(1)$. The same holds true for $s = t$ if one chooses $p_t = e_t$ and makes the necessary adjustments in the above argument. Considering $\phi_{s,t}^{\pm}$, we find that $\pm\mathrm{Im}(X)_{s,t} \in \mathcal{D}_A(1)$ for all $s, t \in [n]$ as well. Now set

$$T_{s,t} = \sum_j B_j \otimes (X_j)_{s,t}.$$

It holds that

$$\|T_{s,t}\|_\infty \leq \left\|\sum_j B_j \otimes \mathrm{Re}(X_j)_{s,t}\right\|_\infty + \left\|\sum_k B_k \otimes \mathrm{Im}(X_k)_{s,t}\right\|_\infty.$$

Moreover, we know that

$$-I_n \leq \sum_j B_j \otimes \mathrm{Re}(X_j)_{s,t} \leq I_n, \qquad -I_n \leq \sum_k B_k \otimes \mathrm{Im}(X_k)_{s,t} \leq I_n$$

as the real and imaginary parts of the entries of X have been found to be in $\mathcal{D}_A(1)$ and therefore also in $\mathcal{D}_B(1)$ by hypothesis. Combining the two findings, it follows that $\|T_{s,t}\|_\infty \leq 2$. An application of Lemma VII.1 to $T/2$ allows us to conclude that $\|T\|_\infty \leq 2n$. Thus,

$$-I_{dn} \leq \frac{1}{2n}\sum_j B_j \otimes X_j \leq I_{dn},$$

which implies $\pm 1/(2n)X \in \mathcal{D}_B(n)$. At level $n = d$, this implies $\mathcal{D}_A(d) \subseteq 2d\mathcal{D}_B(d)$. Since $B \in (\mathcal{M}_d^{sa})^g$, an application of Corollary IV.6 proves the assertion. □

Remark VII.3. The assumption that $\mathcal{D}_A(1)$ is bounded is not necessary and does not appear in Ref. 28; see Remark IV.7.

Exploiting the link between inclusion of free spectrahedra and joint measurability, the previous proposition corresponds to the following:

Corollary VII.4. Let $s = (1/(2d), \ldots, 1/(2d))$. Then $s \in \Gamma(g, d)$.

Proof. The matrix diamond is symmetric, i.e., it holds that $-\mathcal{D}_{\diamond,g} = \mathcal{D}_{\diamond,g}$. Furthermore, $\mathcal{D}_{\diamond,g}(1)$ is bounded. From Proposition VII.2, it follows that $\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_A(1)$ implies $\mathcal{D}_{\diamond,g} \subseteq 2d\mathcal{D}_A$ for any $A \in \left(\mathcal{M}_d^{sa}\right)$. Thus, $s \in \Delta(g, d)$. The claim then follows from Theorem V.7. □

Remark VII.5. In Ref. 28, Proposition VII.2 was proven for spectrahedra defined by real matrices and with d instead of $2d$. We point out that this result cannot hold in the complex case. Consider $d = 2$, $g = 3$, $A_i = \sigma_i$, and $B_i = \sigma_i^\top$ for $i \in [3]$. Here, σ_i are the usual Pauli matrices. In this case, the operator system spanned by the A_i is the whole matrix algebra. Let $s_0 = (s, s, s)$ with $s \in [0, 1]$. For $s_0\mathcal{D}_A \subseteq \mathcal{D}_B$ to hold, the map

$$\Phi'(A) = sA^\top + (1-s)\mathrm{tr}[A]/2I$$

for $A \in \mathcal{M}_2$ must be completely positive. This follows from Lemma IV.4 [see also Ref. 28 (Sec. 1.4)]. A short calculation shows that the minimal eigenvalue of the Choi matrix [see Ref. 45 *(Sec. 2.2.2)*] *of this map is $-s/2 + (1-s)/4$. For the Choi matrix to be positive, we thus require $s \le 1/3$, which is fulfilled by $s = 1/2d$ but not by $s = 1/d$. However, this calculation does not imply that $2d$ is optimal in Proposition VII.2, leaving this question open.*

B. Dimension independent lower bounds

We restate in this section one implication of Ref. 38 (Theorem 6.6), which, interpreted in terms of inclusion constants, yields Theorem VII.7. For the convenience of the reader and for the sake of being self-contained, we reproduce the proof with several simplifications and written in a language more familiar to quantum information specialists.

Let $\mathrm{MatBall}_g$ be the *matrix ball* [see Ref. 40 (Chap. 7), Ref. 28 (Sec. 14), and Ref. 14 (Sec. 9) for the different operator structures one can put on the ℓ_2-ball]

$$\mathrm{MatBall}_g := \{(X_1, \ldots, X_g) \in (\mathcal{M}_d^{sa})^g : \sum_{i=1}^g X_i^2 \le I\}.$$

We recall the following result from Ref. 38 (Lemma 6.5).

Lemma VII.6. For all $g \ge 1$, $\mathcal{D}_{\diamond,g} \subseteq \mathrm{MatBall}_g$.

Proof. Let $(X_1, \ldots, X_g) \in \mathcal{D}_{\diamond,g}$. By definition, we have that, for all $\varepsilon \in \{\pm 1\}^g$,

$$-I \le \sum_{i=1}^g \varepsilon_i X_i \le I.$$

Squaring the relation above, we get

$$\sum_i X_i^2 + \sum_{i \ne j} \varepsilon_i \varepsilon_j X_i X_j \le I.$$

Averaging the above inequality for all values of ε, we are left with $\sum_i X_i^2 \le I$, which is the claim we aimed for. □

Define the "quarter-circle"

$$\mathrm{QC}_g := \left\{ s \in \mathbb{R}_+^g : s_1^2 + \cdots + s_g^2 \le 1 \right\} \tag{6}$$

to be part of the unit disk contained in the positive orthant.

Theorem VII.7 [**Ref.** 38 **(Theorem 6.6)**]. *Let $A \in (\mathcal{M}_d^{sa})^g$. For any vector $s \in \mathbb{R}_+^g$ such that $\sum_i s_i^2 \le 1$ and any spectrahedron $\mathcal{D}_A$, whenever $\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_A(1)$, we have $s \cdot \mathrm{MatBall}_g \subseteq \mathcal{D}_A$. In particular, $s \cdot \mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_A$. In terms of inclusion constants, we have*

$$\forall g, d, \qquad \mathrm{QC}_g \subseteq \Delta(g, d).$$

Proof. Using Lemma VII.6, under the hypotheses, we only need to show the inclusion $s\mathrm{MatBall}_g \subseteq \mathcal{D}_A$. To this end, consider a g-tuple of $n \times n$ self-adjoint matrices $(X_1, \ldots, X_g)$ such that $\sum_{i=1}^g X_i^2 \le I$. We claim that this inequality implies that, for all s as in the statement,

$$\sum_{i=1}^g s_i |X_i| \le I.$$

Indeed, this follows from the general matrix inequality

$$\left\| \sum_i B_i C_i \right\|_\infty \leq \left\| \sum_i B_i B_i^* \right\|_\infty^{1/2} \left\| \sum_i C_i^* C_i \right\|_\infty^{1/2}.$$

The above inequality can be seen to hold by writing the B_i in the first row of a larger matrix and the C_i in the first column of another such matrix. Writing now $X_i = Y_i - Z_i$ with positive semidefinite operators Y_i, Z_i in such a way that $|X_i| = Y_i + Z_i$, we also have

$$\sum_{i=1}^g s_i Y_i + \sum_{i=1}^g s_i Z_i \leq I.$$

We interpret the last inequality as $\{s_i Y_i\}_{i=1}^g \sqcup \{s_i Z_i\}_{i=1}^g$ being a partial POVM and we apply the Naimark dilation theorem [see Ref. 35 (Sec. 2.2.8) or Ref. 45 (Theorem 2.42)]. Hence, there exists an isometry $V : \mathbb{C}^n \to \mathbb{C}^n \otimes \mathbb{C}^{2g+1}$ and $2g$ mutually orthogonal projections $P_i, Q_i \in \mathcal{M}_{n(2g+1)}$ such that $s_i Y_i = V^* P_i V$ and $s_i Z_i = V^* Q_i V$. We thus have $s_i X_i = V^*(P_i - Q_i)V$, and the operators $R_i := P_i - Q_i$ are commuting, normal, and with joint spectrum in $\mathcal{D}_{\diamond,g}(1)$. Thus, with E the joint spectral measure of R,

$$\begin{aligned}\sum_{i=1}^g A_i \otimes s_i X_i &= \sum_{i=1}^g A_i \otimes V^* R_i V \\ &= \int_{\mathcal{D}_A(1)} \left(\sum_{i=1}^g A_i y_i \right) \otimes V^* dE(y) V \\ &\leq I \otimes I.\end{aligned}$$

This shows that $s\mathrm{MatBall}_g(k) \subseteq \mathcal{D}_A(k)$ and the assertion follows as k was arbitrary. □

Remark VII.8. Corollary 7.17 of Ref. 14 shows that $(1/g, \ldots, 1/g) \in \Delta_{\mathcal{D}_A}$ *for all* $A \in (\mathcal{M}_d^{sa})^g$ *such that* $\mathcal{D}_A(1)$ *is invariant under projection onto some orthonormal basis. This result thus holds, in particular, for the matrix diamond, but also for more general spectrahedra. It corresponds to the observation of Remark III.3 that* $(1/g, \ldots, 1/g) \in \Gamma(g, d)$. *In the concrete situation that* $\mathcal{D}_A = \mathcal{D}_{\diamond,g}$, *the statement of Theorem VII.7 is much stronger, as one might expect.*

VIII. UPPER BOUNDS

We present in this section two upper bounds (i.e., containing sets) for the Γ and Γ^0 sets, one coming from quantum information theory[48] and another one coming from matrix convex set theory.[38] These two upper bounds are interesting in two different regimes: the first one applies when the number of POVMs is larger than the dimension of the quantum system, while the second one applies in the complementary regime, where the dimension is large with respect to the number of POVMs. Another important difference between the two results below is that the first one (Theorem VIII.2) deals with the set Γ^{lin}, while the second one (Theorem VIII.8) deals with the set Γ^0.

A. Zhu's necessary condition for joint measurability

We start by recalling Zhu's incompatibility criterion from Ref. 48; see also Ref. 49 for the mathematical details. To do so, define for a non-zero operator $A \in \mathcal{M}_d$,

$$\mathcal{G}(A) := \frac{|A\rangle\langle A|}{\mathrm{tr}[A]} \in \mathcal{M}_{d^2}^{sa},$$

where $|A\rangle \in \mathbb{C}^{d^2}$ is the vectorization of the matrix A. In the same vein, if $A^\circ := A - \mathrm{tr}[A]/dI$ denotes the traceless version of A, let

$$\overline{\mathcal{G}}(A) := \frac{|A^\circ\rangle\langle A^\circ|}{\mathrm{tr}[A]} \in \mathcal{M}_{d^2}^{sa}.$$

We also extend additively the definitions above to POVMs

$$\mathcal{G}^{\#}(\{E_i\}) := \sum_i \mathcal{G}^{\#}(E_i),$$

where $\mathcal{G}^{\#}$ denotes either $\mathcal{G}$ or $\overline{\mathcal{G}}$. Using the remarkable fact that the functions $\mathcal{G}^{\#}$ are subadditive, Zhu has showed the following result in Ref. 48 [Eqs. (10) and (11)].

Proposition VIII.1. If a set of g POVMs $\{E^{(1)}\}, \ldots, \{E^{(g)}\}$ on $\mathcal{M}_d$ are compatible, then

$$\min\{\mathrm{tr}[H] : H \geq \mathcal{G}(\{E^{(i)}\}),\ \forall i \in [g]\} = 1 + \min\{\mathrm{tr}[H] : H \geq \overline{\mathcal{G}}(\{E^{(i)}\}),\ \forall i \in [g]\} \leq d.$$

It turns out that the semidefinite program appearing in the result above is particularly easy in the case where the $d^2 \times d^2$ matrices $\overline{\mathcal{G}}(\{E^{(1)}\}), \ldots, \overline{\mathcal{G}}(\{E^{(g)}\})$ have orthogonal supports; if that is the case, then the optimal H is the sum of the matrices $\overline{\mathcal{G}}(\{E^{(i)}\})$ and the condition above reads

$$\sum_{i=1}^{g} \mathrm{tr}[\overline{\mathcal{G}}(\{E^{(i)}\})] \leq d - 1.$$

In order to exploit this phenomenon, let $G_{max}(d)$ be the maximal integer g such that there exist E_1, ..., E_g non-trivial orthogonal projections in $\mathcal{M}_d$ with the property $d\ \mathrm{tr}[E_iE_j] = (\mathrm{tr}[E_i])(\mathrm{tr}[E_j])$ for all $1 \leq i < j \leq g$.

Theorem VIII.2. *For all dimensions d and all $1 \leq g \leq G_{max}(d)$, we have*

$$\Gamma^{lin}(g, d) \subseteq \sqrt{d-1}\,\mathrm{QC}_g.$$

Proof. Let g, d be as in the statement and consider the 2-outcome POVMs $\{E_i, I_d - E_i\}$, $i \in [g]$, where E_i are such that $d\ \mathrm{tr}[E_iE_j] = (\mathrm{tr}[E_i])(\mathrm{tr}[E_j])$ for all $1 \leq i < j \leq g$. Since the effects E_i are non-trivial orthogonal projections, the previous condition is equivalent to $\mathrm{tr}[E_i^\circ E_j^\circ] = 0$, for all $i \neq j$. Fix $s \in \Gamma^{lin}(g, d)$; from the definition of the set $\Gamma^{lin}(g, d)$, it follows that the effects $s_iE_i + (1 - s_i)d^{-1}\mathrm{tr}[E_i]I$ are compatible and, thus, by Proposition VIII.1,

$$\sum_{i=1}^{g} \mathrm{tr}[\overline{\mathcal{G}}(\{s_iE_i + (1-s_i)d^{-1}\mathrm{tr}[E_i]I, s_i(I_d - E_i) + (1-s_i)d^{-1}(d - \mathrm{tr}[E_i])I\})] \leq d - 1. \tag{7}$$

Let us compute, for fixed i, the general term in the sum above. Start by computing

$$\overline{\mathcal{G}}(\{E_i, I_d - E_i\}) = \frac{|E_i^\circ\rangle\langle E_i^\circ|}{\mathrm{tr}[E_i]} + \frac{|(I_d - E_i)^\circ\rangle\langle (I_d - E_i)^\circ|}{d - \mathrm{tr}[E_i]} = \frac{d|E_i^\circ\rangle\langle E_i^\circ|}{\mathrm{tr}[E_i](d - \mathrm{tr}[E_i])}.$$

Using the fact that E_i is a (non-trivial) projection, we get

$$\mathrm{tr}[\overline{\mathcal{G}}(\{E_i, I_d - E_i\})] = \frac{d\ \mathrm{tr}[(E_i^\circ)^2]}{\mathrm{tr}[E_i](d - \mathrm{tr}[E_i])} = \frac{d[\mathrm{tr}[E_i^2] - \mathrm{tr}[E_i]^2/d]}{\mathrm{tr}[E_i](d - \mathrm{tr}[E_i])} = 1.$$

For the noisy version, mixing with the identity does not change the trace, hence

$$\begin{aligned}\overline{\mathcal{G}}(\{s_iE_i + (1-s_i)d^{-1}\mathrm{tr}[E_i]I, s_i(I_d - E_i) + (1-s_i)d^{-1}(d - \mathrm{tr}[E_i])I\}) \\ = \frac{ds_i^2|E_i^\circ\rangle\langle E_i^\circ|}{(\mathrm{tr}[E_i])(d - \mathrm{tr}[E_i])} = s_i^2\overline{\mathcal{G}}(\{E_i, I_d - E_i\}),\end{aligned}$$

and, thus, taking the trace,

$$\mathrm{tr}[\overline{\mathcal{G}}(\{s_iE_i + (1-s_i)d^{-1}\mathrm{tr}[E_i]I, s_i(I_d - E_i) + (1-s_i)d^{-1}(d - \mathrm{tr}[E_i])I\})] = s_i^2.$$

Zhu's condition (7) thus implies

$$\sum_{i=1}^{g} s_i^2 \le d - 1,$$

proving the claim. □

Remark VIII.3. An analysis of the proof above shows that the same result holds for the set Γ^0 *instead of* Γ^{lin}*, but with an extra restriction on the operators* E_i*: we must ask that* $\mathrm{tr}[E_i] = d/2$ *for all* $i \in [g]$*. We leave the existence of large tuples of such operators as an open problem.*

Let us now discuss the function $G_{max}(d)$. First, note that in order for the upper bound in the result above to be non-trivial, we must have $G_{max}(d) \ge d$. Below, we give two lower bounds on the function G_{max}, conditional on the existence of *mutually unbiased bases* (MUBs) and *symmetric informationally complete POVMs* (SIC-POVMs).

Recall that k orthonormal bases $\{x_j^{(i)}\}_{j=1}^d$, $i \in [k]$, are called *mutually unbiased* if and only if for all $i_1 \ne i_2$ and all j_1, j_2, $|\langle x_{j_1}^{(i_1)}, x_{j_2}^{(i_2)}\rangle| = 1/\sqrt{d}$. The maximal number of MUBs in $\mathbb{C}^d$ is $d + 1$ and this bound is attained if d is a prime power; very few other existence results are known; see Ref. 15 for a review. If there exist k MUBs in dimension d, then $G_{max}(d) \ge k$. This follows by setting $E_i = \sum_{j \in J_i} |x_j^{(i)}\rangle\langle x_j^{(i)}|$, where $\{x_j^{(i)}\}_{j=1}^d$ is the ith MUB and J_i is some non-trivial subset of $[d]$. We record next the following consequence of Theorem VIII.2.

Corollary VIII.4. If $d = p^n$ *is a prime power, then, for all* $g \le d + 1$,

$$\Gamma^{lin}(g, d) \subseteq \sqrt{d-1}\,\mathrm{QC}_g.$$

Recall that a unit rank POVM $\{d^{-1}|x_i\rangle\langle x_i|\}_{i=1}^{d^2}$ is called *symmetric and informationally complete* if and only if for all $i \ne j$, $|\langle x_i, x_j\rangle|^2 = 1/(d + 1)$. Whether a SIC-POVM exists in dimension d is a challenging question and an ongoing research subject. Analytic examples of SIC-POVMs have been constructed for $d = 1, \ldots, 21, 24, 28, 30, 31, 35, 37, 39, 43, 48, 124, 323$ and numerical constructions exist for much larger values of d; see Ref. 41 for a review and Ref. 2 for some recent progress. The use of SIC-POVMs yields the following corollary of Theorem VIII.2.

Corollary VIII.5. If there exists a SIC-POVM $\{x_i\}_{i=1}^{d^2}$ *in dimension d, then* $G_{max}(d + 1) \ge d^2$ *and thus, for all* $g \le d^2$*, we have*

$$\Gamma^{lin}(g, d + 1) \subseteq \sqrt{d}\,\mathrm{QC}_g.$$

Proof. Let $y_i = (\sqrt{t}, \sqrt{1-t}x_i)$ for $i \in [d^2]$, where $t \in [0, 1]$ is a parameter. Using the fact that the x_i are a SIC-POVM, we have that $|\langle y_i, y_j\rangle|^2 = t + (1 - t)/(d + 1)$, so for $t = 1/d^2$, $|\langle y_i, y_j\rangle|^2 = 1/d$. Setting $E_i = |y_i\rangle\langle y_i|$ then proves the claim. □

B. Pairwise anti-commuting unitary operators and spectrahedral inclusion constants

We now present a different type of upper bound, this time on the set $\Delta^0 = \Gamma^0$. What follows is based on Ref. 38 (Theorem 6.6). We adapt the proof there to our setting by taking into account the system dimension d. The main ingredient of the construction in Ref. 38 is the following Hurwitz-Radon-like result.

Lemma VIII.6 [**Ref.** 34 **or Ref.** 30 **(Theorem 1)**]. *For* $d = 2^k$, $k \in \mathbb{N}_0$*, there exist* $2k + 1$ *anti-commuting, self-adjoint, unitary matrices* $F_1, \ldots, F_{2k+1} \in \mathcal{U}_d$*. Moreover,* 2^k *is the smallest dimension where such a* $(2k + 1)$*-tuple exists.*

A $(2k+1)$-tuple as above is sometimes called a *spin system* in operator theory; see, e.g., Ref. 40. One can easily construct such matrices recursively as follows. For $k = 0$, simply take $F_1^{(0)} := [1]$. For $k \geq 1$, define

$$F_i^{(k+1)} = \sigma_X \otimes F_i^{(k)} \; \forall i \in [2k+1] \quad \text{and} \quad F_{2k+2}^{(k+1)} = \sigma_Y \otimes I_{2^k}, \quad F_{2k+3}^{(k+1)} = \sigma_Z \otimes I_{2^k},$$

where $\sigma_{X,Y,Z}$ are the Pauli matrices

$$\sigma_X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_Y = \begin{bmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{bmatrix}, \quad \text{and} \quad \sigma_Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

For example, we have $F_1^{(1)} = \sigma_X$, $F_2^{(1)} = \sigma_Y$, and $F_3^{(1)} = \sigma_Z$ and

$$F_1^{(2)} = \sigma_X \otimes \sigma_X, \quad F_2^{(2)} = \sigma_X \otimes \sigma_Y, \quad F_3^{(2)} = \sigma_X \otimes \sigma_Z, \quad F_4^{(2)} = \sigma_Y \otimes I_2, \quad \text{and} \quad F_5^{(2)} = \sigma_Z \otimes I_2.$$

Remark VIII.7. Note that our construction differs from the one in Ref. 38 because we aim for the smallest dimension which contains g anti-commuting, self-adjoint, and unitary elements. This way, we obtain $d \geq 2^{\lceil (g-1)/2 \rceil}$ instead of $d \geq 2^{g-1}$ in the next theorem.

Theorem VIII.8 *Let $g \geq 2$, $d \geq 2^{\lceil (g-1)/2 \rceil}$ and consider $s \in \mathbb{R}_+^g$ such that for any spectrahedron $\mathcal{D}_A$ defined by traceless matrices $A_i \in \mathcal{M}_d$, $\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_A(1)$ implies $s \cdot \mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_A$. Then, $\sum_i s_i^2 \leq 1$. In terms of inclusion constants, we have*

$$\forall d \geq 2^{\lceil (g-1)/2 \rceil}, \qquad \Delta(g,d) \subseteq \Delta^0(g,d) \subseteq \mathrm{QC}_g.$$

Proof. Let us consider g anti-commuting, self-adjoint, unitary matrices $F_1, \ldots, F_g \in \mathcal{U}_d$ as in the construction following Lemma VIII.6; these matrices also enjoy the property of being traceless when $g \geq 2$. Let $\mathcal{D}_{\overline{F}}$ be the spectrahedron defined by the matrices $\overline{F_i}$, where $\overline{F_i}$ is the entry-wise complex conjugate of F_i. Since the matrices $\overline{F_i}$ are unitary, it is clear that $\mathcal{D}_{\diamond,g}(1) \subseteq \mathcal{D}_{\overline{F}}(1)$.

Assume now that $s \cdot \mathcal{D}_{\diamond,g} \subseteq \mathcal{D}_{\overline{F}}$ for some non-negative g-tuple s. Put $\hat{s} := s/\|s\|_2$. We claim that $(\hat{s}_i F_i)_{i=1}^g \in \mathcal{D}_{\diamond,g}$. Indeed, for any choice of signs ε_i, we have

$$\left\| \sum_{i=1}^g \varepsilon_i \hat{s}_i F_i \right\|_\infty = \left\| \left(\sum_{i=1}^g \varepsilon_i \hat{s}_i F_i \right)^2 \right\|_\infty^{1/2} = \left\| \sum_{i=1}^g \hat{s}_i^2 I + \sum_{i \neq j} \varepsilon_i \varepsilon_j \hat{s}_i \hat{s}_j F_i F_j \right\|_\infty^{1/2} = \left\| \sum_{i=1}^g \hat{s}_i^2 I \right\|_\infty^{1/2} = 1.$$

In the equality above, we have used the fact that the cross terms in the sum obtained by expanding the square vanish; it is this behavior of the matrices F_i that renders them useful in operator theory. From the hypothesis, it follows that $(s_i \hat{s}_i F_i)_{i=1}^g \in \mathcal{D}_{\overline{F}}$; in particular, we have

$$\left\| \sum_{i=1}^g \frac{s_i^2}{\|s\|_2} \overline{F_i} \otimes F_i \right\|_\infty = \frac{\sum_{i=1}^g s_i^2}{\|s\|_2} = \|s\|_2 \leq 1,$$

which is the conclusion we aimed for. In the equation above, we have used the following fact {see Ref. 38 [Eq. (5.4)] for the corresponding statement}: for non-negative scalars $a_1, \ldots, a_g$,

$$\left\| \sum_{i=1}^g a_i \overline{F_i} \otimes F_i \right\|_\infty = \sum_{i=1}^g a_i.$$

The fact that the left-hand side in the equality above is smaller than the right-hand side follows from the triangle inequality. The reverse inequality follows from taking the scalar product against the maximally entangled state,

$$\omega_d := \frac{1}{d} \sum_{i,j=1}^d (e_i \otimes e_i)(e_j \otimes e_j)^*,$$

for some orthonormal basis $\{e_i\}_{i=1}^d$ of $\mathbb{C}^d$. □

Putting together the result above with Theorem VII.7, we derive the following equality, one of the main results of this paper.

Corollary VIII.9. For any $g \geq 2$ *and any* $d \geq 2^{\lceil (g-1)/2 \rceil}$*, we have*

$$\Delta(g,d) = \Gamma(g,d) = \Delta^0(g,d) = \Gamma^0(g,d) = \mathrm{QC}_g.$$

Remark VIII.10. If the dimension bound $d \geq 2^{\lceil (g-1)/2 \rceil}$ *holds, the matrices* $(F_1 + I_d)/2 \oplus 0, \ldots, (F_g - I_d)/2 \oplus 0$ *considered in this section are* the most incompatible g-tuple of $d \times d$ quantum effects. *Indeed, for any direction* $\hat{s} \in \mathrm{QC}_g$, $\|\hat{s}\| = 1$*, it follows from Corollary VIII.9 that the g-tuple* $(t_1 \hat{s}_1 (F_1 + I_d)/2, \ldots, t_g \hat{s}_g (F_g + I_d)/2)$ *is compatible if and only if* $t_i \leq 1$ *for all* $i \in [g]$*. We would also like to point out that, for* $d = 2$ *and* $g = 3$, $g = 2$*, the claim above corresponds to the maximal incompatibility of the measurements corresponding to the Pauli observables.*

IX. DISCUSSION

In this final section, we would like to put the results obtained in this work in perspective and compare them with previously known bounds. We also list and discuss some questions left open in this work.

A. The shape of the different compatibility regions

We start by listing some previously known results on the different sets $\Gamma^{\#}$ considered in this work. Let us remind the reader that our primary focus was on the sets $\Gamma(g, d)$ because of their connection to the inclusion problem for free spectrahedra. In the quantum information community, the sets Γ^{all} and Γ^{lin} play a very important role because the most general type of trivial noise is allowed in the former case and because of the linear structure in the latter case. Previously, mainly the cases of small g, d have been considered in the literature. General lower bounds have been shown mostly using tools from symmetric approximate cloning, while upper bounds were rarely considered in the general case (we are considering here only the case of 2-outcome POVMs).

Let us first discuss the results in the literature for small g, d. Using an argument which connects joint measurability with a violation of the CHSH inequality,[47] it was shown in Ref. 11 that $\mathrm{QC}_2 \subseteq \Gamma(2, d)$ for all $d \in \mathbb{N}$. Furthermore, it was shown that also $\Gamma^{all}(2, 2) \subseteq \mathrm{QC}_2$ [see Ref. 10 (Proposition 3)] and $\Gamma^0(2, 2) \subseteq \mathrm{QC}_2$ [see Ref. 10 (Proposition 4)]. Therefore, for $d \geq 2$, an application of Proposition III.6 yields

$$\Gamma(2,d) = \Gamma^{all}(2,d) = \Gamma^0(2,d) = \mathrm{QC}_2.$$

Less was known in the $g \geq 3$ case since the connection to the CHSH inequality no longer holds.[4,29] From Ref. 9 [see also Ref. 12 (Sec. 14.4)], it follows that $\mathrm{QC}_3 \subseteq \Gamma^0(3, 2)$. Moreover, the authors of Ref. 7 show that $\Gamma^0(3, 2) \subseteq \mathrm{QC}_3$, hence $\Gamma^0(3, 2) = \mathrm{QC}_3$. This was improved in Ref. 37 (Sec. XI) to show also $\Gamma^{all}(3, 2) \subseteq \mathrm{QC}_3$. Using the results of this paper and combining them with the findings above, we can prove a stronger statement. An application of Theorem VII.7 together with Proposition III.6 yields

$$\Gamma(3,d) = \Gamma^{all}(3,d) = \Gamma^0(3,d) = \mathrm{QC}_3.$$

In the general case, the lower bounds came mainly from symmetric cloning;[23] see Proposition VI.2.

Let us now discuss the contributions of this paper to both the theory of joint measurability and free spectrahedra. As discussed in the Introduction, our main insight, the relation between the joint measurability of 2-outcome POVMs and the inclusion problem for the matrix diamond, allows us to translate the results from one field to the other. Arguably one of the main results in this work is the lower bound obtained in Theorem VII.7. Our theorem is based on the results about inclusion of free spectrahedra derived in Ref. 38, which can be transferred to the quantum setting. Together with the upper bound from Ref. 38 and the lower bound from Ref. 28, we obtain a much better understanding of the sets $\Gamma(g, d)$. We present in Fig. 3 our current picture of the sets $\Gamma(g, d)$ or, equivalently, of the

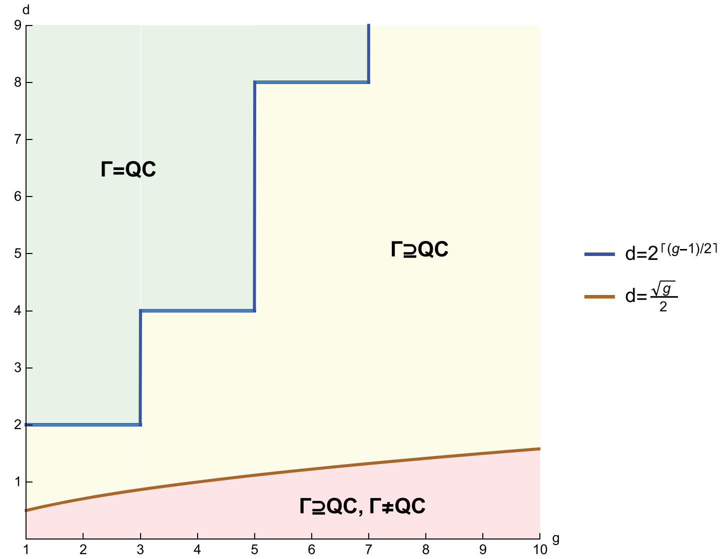

FIG. 3. The sets $\Gamma(g, d)$.

sets of inclusion constants $\Delta(g, d)$. The curves $d = 2^{\lceil (g-1)/2 \rceil}$ and $d = \sqrt{g}/2$ delimit three regions: above the first curve, we know that the set $\Gamma(g, d)$ is equal to QC_g, the positive part of the unit Euclidean ball, while below the second curve, we know the inclusion $\mathrm{QC}_g \subseteq \Gamma(g, d)$ to be strict. Below the curve $d = 2^{\lceil (g-1)/2 \rceil}$, the upper bound from Theorem VIII.8 does not apply, while below the second curve, $d = \sqrt{g}/2$, the lower bound $1/(2d)$ in the symmetric case is larger than the lower bound $1/\sqrt{g}$ coming from the quarter-circle QC_g. It is worthwhile to mention that the best lower bound for the sets $\Gamma(g, d)$ coming from symmetric cloning (second line in Proposition VI.2) is worse than the best of the two bounds coming from spectrahedron theory,

$$\frac{g + 2d}{g(1 + 2d)} \le \max\left\{\frac{1}{\sqrt{g}}, \frac{1}{2d}\right\}.$$

However, in the asymmetric regime, cloning gives non-trivial lower bounds since the $1/(2d)$ bound from Proposition VII.2 is not applicable for asymmetric tuples. We expect to obtain non-trivial results as soon as $\frac{g+2d}{g(1+2d)} > 1/\sqrt{g}$, that is, as soon as $g > 4d^2$. As an example, we plot in Fig. 4, for even $g = 2g_0$, the set of points of the form $(\underbrace{s, \ldots, s}_{g_0 \text{ times}}, \underbrace{t, \ldots, t}_{g_0 \text{ times}})$ belonging to $\Gamma(g, 4)$ and to QC_4.

The main question left open in this work is to compute the sets $\Gamma(g, d) = \Delta(g, d)$ for the range of parameters g, d where the upper and lower bounds from Ref. 38 do not agree.

Question IX.1. Compute, for $d < 2^{\lceil (g-1)/2 \rceil}$, the sets $\Gamma(g, d) = \Delta(g, d)$.

Regarding the sets $\Gamma^{lin}(g, d)$, the lower bounds coming from cloning (see Sec. VI) were already known in the literature; in particular, we recall that, in the symmetric case,

$$\frac{g + d}{g(1 + d)} \underbrace{(1, 1, \ldots, 1)}_{g \text{ times}} \subseteq \Gamma^{lin}(g, d).$$

Unfortunately, since there is no known inclusion of the Γ sets into the Γ^{lin} sets, we cannot use in this setting the very powerful lower bounds for $\Gamma(g, d)$ in Theorem VII.7; see also Remark III.5.

The upper bounds for the sets $\Gamma^{lin}(g, d)$ are new and come from Zhu's criterion (Corollaries VIII.4 and VIII.5)

$$\forall d \text{ prime power and } g \le d + 1, \qquad \Gamma^{lin}(g, d) \subseteq \sqrt{d - 1}\,\mathrm{QC}_g,$$
$$\forall d \in \{2, 3, \ldots, 20, 23, 27, 29, 30, \ldots\} \text{ and } g \le (d - 1)^2, \qquad \Gamma^{lin}(g, d) \subseteq \sqrt{d - 1}\,\mathrm{QC}_g$$

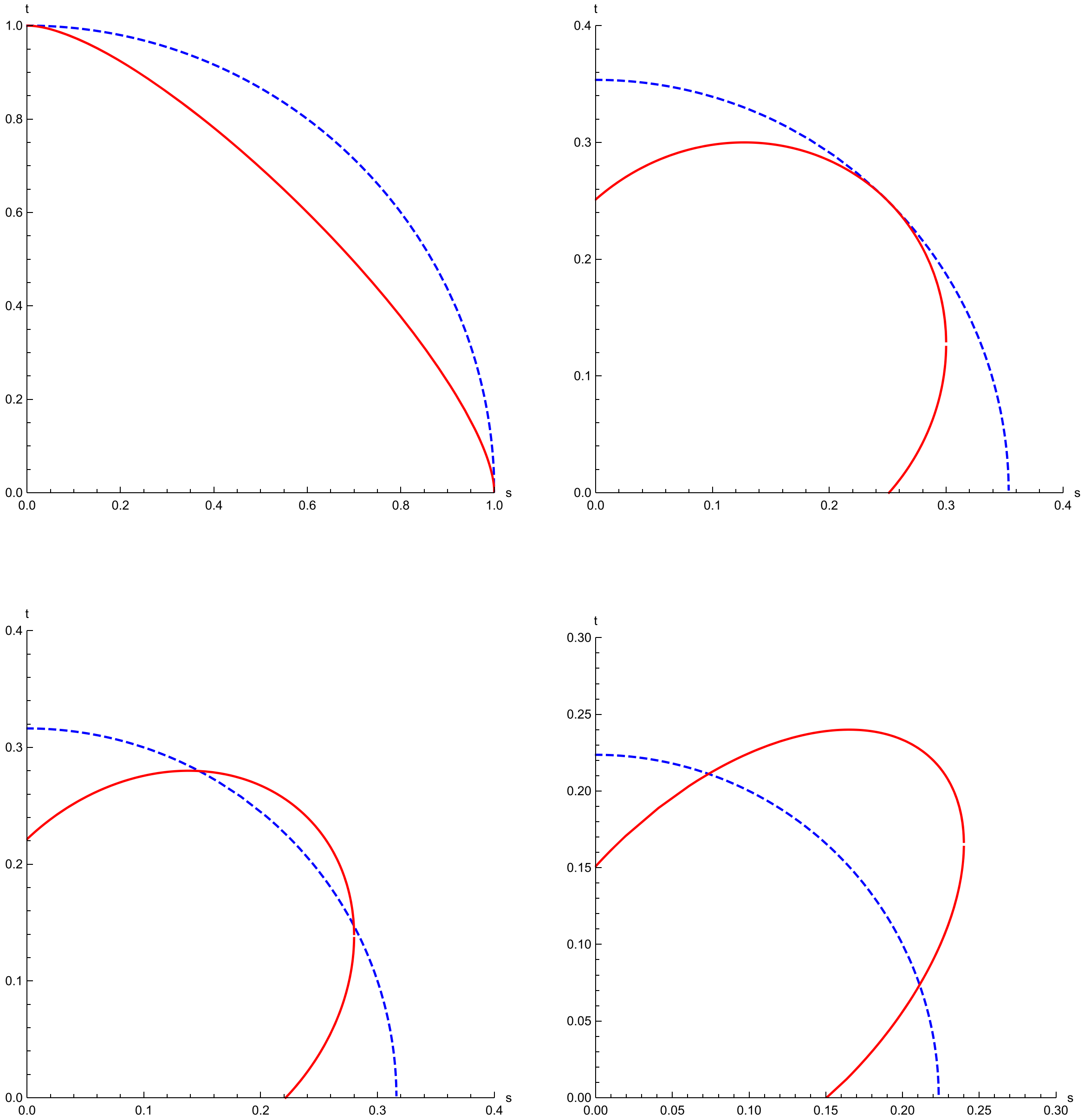

FIG. 4. The range of parameters (s, t) for which the point $(s, \ldots, s, t, \ldots, t) \in [0,1]^g$, having the same number of s and t, belongs to $\Gamma^{clone}(g, 4) \subseteq \Gamma(g, 2)$ (red curve) and QG_g (blue dashed curve) for $g = 2, 16, 20, 40$. Note that the bound from asymmetric cloning becomes better for $g > 4d^2 = 16$.

and from the work of Passer *et al.* (Theorem VIII.8 and Proposition III.4)

$$\forall d \geq 2^{\lceil (g-1)/2 \rceil}, \qquad \Gamma^{lin}(g,d) \subseteq \Gamma^0(g,d) \subseteq \mathrm{QC}_g.$$

Finally, regarding the sets Γ^{all}, allowing the most general type of noise, the new lower bounds obtained in this work are precisely the same as the ones for the sets Γ. Importantly, for all g, d, we have

$$\mathrm{QC}_g \subseteq \Gamma^{all}(g,d).$$

Note that the bound above was previously known only in the case $g = 2$. Moreover, in the symmetric case, we have

$$\max\left\{\frac{1}{\sqrt{g}}, \frac{1}{2d}\right\}\underbrace{(1,1,\ldots,1)}_{g \text{ times}} \subseteq \Gamma^{all}(g,d).$$

Upper bounds can be obtained via the map F from Proposition III.4 from upper bounds for Γ. For example, in the symmetric case, using Theorem VIII.8, we get

$$\forall d \geq 2^{\lceil (g-1)/2 \rceil}, \qquad s(\underbrace{1, 1, \ldots, 1}_{g \text{ times}}) \subseteq \Gamma^{all}(g, d) \implies s \leq \frac{2}{1 + \sqrt{g}},$$

which is roughly two times the lower bounds above.

B. The shape of the inclusion sets

Hitherto, we have discussed the implications of this work for quantum information theory. However, our results also shed new light on $\Delta(g, d)$ and $\Delta^0(g, d)$, the sets of inclusion constants for the matrix diamond.

As $\Delta^0(g, d) = \Gamma^0(g, d)$ by Theorem V.7, we have the lower bounds

$$\forall g, d \geq 1 \quad QC_g \subseteq \Gamma(g, d) \subseteq \Delta^0(g, d),$$
$$\forall g, d \geq 1 \quad \Gamma^{clone}(g, d) \subseteq \Delta^0(g, d).$$

Looking at the symmetric case, for which $s(1, \ldots, 1) \in \Gamma^{clone}$ if and only if $s \leq (g + d)/(g(1 + d))$, we see that this is larger than $1/\sqrt{g}$ if and only if $d \leq \sqrt{g}$. Therefore, both lower bounds are non-trivial. We remark that for all $d, g \geq 1$,

$$\frac{1}{2d} \leq \frac{g + d}{g(1 + d)}.$$

Therefore, the result from symmetric cloning is always stronger than the one from Ref. 28 (see Corollary VII.4). In terms of upper bounds, we only have that

$$\forall d \geq 2^{\lceil (g-1)/2 \rceil}, \qquad \Delta^0(g, d) \subseteq QC_g$$

from the work of Passer *et al.* (Theorem VIII.8).

Regarding the sets $\Delta(g, d)$, one obtains new lower bounds in the asymmetric setting using cloning and the inclusion $\Gamma^{clone}(g, 2d) \subseteq \Delta(g, d)$; see Fig. 4.

C. Outlook: POVMs with more outcomes

In this work, we have focused on binary measurements. However, our methods also work for measurements with more outcomes. For example, consider the case of a binary POVM $\{E, I - E\}$ and a POVM with three outcomes $\{F_1, F_2, I - F_1 - F_2\}$. Then, it can be shown that joint measurability is equivalent to the inclusion problem of the free spectrahedra defined by

$$A_1 = \frac{2}{3}\text{diag}[2, -1, -1, 2, -1, -1], \; A_2 = \frac{2}{3}\text{diag}[-1, 2, -1, -1, 2, -1], \; A_3 = \text{diag}[1, 1, 1, -1, -1, -1]$$

and

$$B_1 = 2E - I, \qquad B_2 = 2F_1 - \frac{2}{3}I, \qquad B_3 = 2F_2 - \frac{2}{3}I.$$

That is, $\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1)$ if and only if $E, I - E, F_1, F_2$, and $I - F_1 - F_2$ are quantum effects and $\mathcal{D}_A \subseteq \mathcal{D}_B$ if and only if $\{E, I - E\}$ and $\{F_1, F_2, I - F_1 - F_2\}$ are jointly measurable POVMs. Inclusion constants then correspond again to mixing with $I/2$ (for the binary POVM) and $I/3$ (for the three-outcome POVM), respectively. This idea is explored in detail in Ref. 5.

ACKNOWLEDGMENTS

A.B. would like to thank Michael M. Wolf for pointing out the connection between joint measurability and the CHSH inequality. A.B. acknowledges support from the ISAM Graduate Center at the Technische Universität München. I.N. would like to thank Teiko Heinosaari for many inspiring discussions about joint measurability; in particular, he pointed out to us the fact in Remark V.4.

I.N.'s research has been supported by the ANR projects StoQ (Grant No. ANR-14-CE25-0003-01) and NEXT (Grant No. ANR-10-LABX-0037-NEXT) and by the PHC Sakura program (Grant No. 38615VA). I.N. also acknowledges the hospitality of the Technische Universität München. A.B. and I.N. would like to thank the Institut Henri Poincaré in Paris for its hospitality and for hosting the trimester on "Analysis in Quantum Information Theory," during which part of this work was undertaken.

[1] Ambrozie, C.-G. and Gheondea, A., "An interpolation problem for completely positive maps on matrix algebras: Solvability and parametrization," Linear Multilinear Algebra **63**(4), 826–851 (2015).
[2] Appleby, M., Chien, T.-Y., Flammia, S., and Waldron, S., "Constructing exact symmetric informationally complete measurements from numerical solutions," J. Phys. A: Math. Theor. **51**(16), 165302 (2018).
[3] Ben-Tal, A. and Nemirovski, A., "On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty," SIAM J. Optim. **12**(3), 811–833 (2002).
[4] Bene, E. and Vértesi, T., "Measurement incompatibility does not give rise to Bell violation in general," New J. Phys. **20**(1), 013021 (2018).
[5] Bluhm, A. and Nechita, I., "Compatibility of quantum measurements and inlcusion constants for the matrix jewel," preprint arXiv:1809.04514 (2018).
[6] Bohr, N., "The quantum postulate and the recent development of atomic theory," Nature **121**(3050), 580–590 (1928).
[7] Brougham, T. and Andersson, E., "Estimating the expectation values of spin-1/2 observables with finite resources," Phys. Rev. A **76**, 052313 (2007).
[8] Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V., and Wehner, S., "Bell nonlocality," Rev. Mod. Phys. **86**, 419–478 (2014).
[9] Busch, P., "Unsharp reality and joint measurements for spin observables," Phys. Rev. D **33**(8), 2253 (1986).
[10] Busch, P. and Heinosaari, T., "Approximate joint measurements of qubit observables," Quantum Inf. Comput. **8**(8), 797–818 (2008).
[11] Busch, P., Heinosaari, T., Schultz, J., and Stevens, N., "Comparing the degrees of incompatibility inherent in probabilistic physical theories," EPL (Europhys. Lett.) **103**(1), 10002 (2013).
[12] Busch, P., Lahti, P., Pellonpää, J.-P., and Ylinen, K., *Quantum Measurement* (Springer, 2016).
[13] Cerf, N. J., "Asymmetric quantum cloning in any dimension," J. Mod. Opt. **47**(2-3), 187–209 (2000).
[14] Davidson, K. R., Dor-On, A., Shalit, O. M., and Solel, B., "Dilations, inclusions of matrix convex sets, and completely positive maps," Int. Math. Res. Not. **2017**(13), 4069–4130.
[15] Durt, T., Englert, B.-G., Bengtsson, I., and Życzkowski, K., "On mutually unbiased bases," Int. J. Quantum Inf. **8**(04), 535–640 (2010).
[16] Fine, A., "Hidden variables, joint probability, and the Bell inequalities," Phys. Rev. Lett. **48**(5), 291–295 (1982).
[17] Gudder, S., "Compatibility for probabilistic theories," preprint arXiv:1303.3647 (2013).
[18] Gurvits, L., "Classical deterministic complexity of Edmonds' problem and quantum entanglement," in *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing* (ACM, 2003), pp. 10–19.
[19] Hashagen, A.-L., "Universal asymmetric quantum cloning revisited," Quantum Inf. Comput., **17**(9-10), 0747–0778 (2017).
[20] Heinosaari, T., Jivulescu, M. A., Reeb, D., and Wolf, M. M., "Extending quantum operations," J. Math. Phys. **53**, 102208 (2012).
[21] Heinosaari, T., Kiukas, J., and Reitzner, D., "Noise robustness of the incompatibility of quantum measurements," Phys. Rev. A **92**, 022115 (2015).
[22] Heinosaari, T., Miyadera, T., and Ziman, M., "An invitation to quantum incompatibility," J. Phys. A: Math. Theor. **49**(12), 123001 (2016).
[23] Heinosaari, T., Schultz, J., Toigo, A., and Ziman, M., "Maximally incompatible quantum observables," Phys. Rev. A **378**(24-25), 1695–1699 (2014).
[24] Heinosaari, T. and Ziman, M., *The Mathematical Language of Quantum Theory* (Cambridge University Press, 2011).
[25] Heisenberg, W., "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," Z. Phys. **43**(3), 172–198 (1927).
[26] Helton, J. W., Klep, I., and McCullough, S., "Free convex algebraic geometry," in *Semidefinite Optimization and Convex Algebraic Geometry*, MOS-SIAM Series on Optimization (SIAM, 2013), pp. 341–405.
[27] Helton, J. W., Klep, I., and McCullough, S., "The matricial relaxation of a linear matrix inequality," Math. Programm. **138**(1-2), 401–445 (2013).
[28] Helton, J. W., Klep, I., McCullough, S. A., and Schweighofer, M., "Dilations, Linear matrix inequalities, the matrix cube problem and beta distributions," preprint arXiv:1412.1481v4 (2014).
[29] Hirsch, F., Quintino, M. T., and Brunner, N., "Quantum measurement incompatibility does not imply Bell nonlocality," Phys. Rev. A **97**, 012129 (2018).
[30] Hrubeš, P., "On families of anticommuting matrices," Linear Algebra Appl. **493**, 494–507 (2016).
[31] Iblisdir, S., Acín, A., Cerf, N. J., Filip, R., Fiurášek, J., and Gisin, N., "Multipartite asymmetric quantum cloning," Phys. Rev. A **72**(4), 042328 (2005).
[32] Kay, A., "Optimal universal quantum cloning: Asymmetries and fidelity measures," Quantum Inf. Comput. **16**(11-12), 0991–1028 (2016).
[33] Keyl, M., "Fundamentals of quantum information theory," Phys. Rep. **369**(5), 431–548 (2002).
[34] Newman, M. H. A., "Note on an algebraic theorem of Eddington," J. London Math. Soc. **7**(2), 93–99 (1932).
[35] Nielsen, M. A. and Chuang, I. L., *Quantum Computation and Quantum Information* (Cambridge University Press, 2010).
[36] We thank Teiko Heinosaari for pointing this out to us.

[37] Pal, R. and Ghosh, S., "Approximate joint measurement of qubit observables through an Arthur–Kelly model," J. Phys. A: Math. Theor. **44**(48), 485303 (2011).
[38] Passer, B., Shalit, O. M., and Solel, B., "Minimal and maximal matrix convex sets," J. Funct. Anal. **274**, 3197–3253 (2018).
[39] Paulsen, V., *Completely Bounded Maps and Operator Algebras*, Volume 78 of Cambridge Studies in Advanced Mathematics (Cambridge University Press, 2003).
[40] Pisier, G., *Introduction to Operator Space Theory*, Volume 294 of London Mathematical Society Lecture Note Series (Cambridge University Press, 2003).
[41] Scott, A. J. and Grassl, M., "Symmetric informationally complete positive-operator-valued measures: A new computer study," J. Math. Phys. **51**(4), 042203 (2010).
[42] Studziński, M., Ćwikliński, P., Horodecki, M., and Mozrzymas, M., "Group-representation approach to $1 \to N$ universal quantum cloning machines," Phys. Rev. A **89**(5), 052322 (2014).
[43] Uola, R., Budroni, C., Gühne, O., and Pellonpää, J.-P., "One-to-one mapping between steering and joint measurability problems," Phys. Rev. Lett. **115**(23), 230402 (2015).
[44] Uola, R., Moroder, T., and Gühne, O., "Joint measurability of generalized measurements implies classicality," Phys. Rev. Lett. **113**, 160403 (2014).
[45] Watrous, J., *The Theory of Quantum Information* (Cambridge University Press, 2018).
[46] Werner, R. F., "Optimal cloning of pure states," Phys. Rev. A **58**(3), 1827–1832 (1998).
[47] Wolf, M. M., Pérez-García, D., and Fernández, C., "Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory," Phys. Rev. Lett. **103**, 230402 (2009).
[48] Zhu, H., "Information complementarity: A new paradigm for decoding quantum incompatibility," Sci. Rep. **5**, 14317 (2015).
[49] Zhu, H., Hayashi, M., and Chen, L., "Universal steering criteria," Phys. Rev. Lett. **116**(7), 070403 (2016).

B Further articles

B.1 Dimensionality reduction of SDPs through sketching

Dimensionality reduction of SDPs through sketching

Andreas Bluhm and Daniel Stilck França

Semidefinite programs (SDPs) constitute an important class of problems in convex optimization with applications to engineering and quantum information theory. Although many problems of this class are known to be solvable in polynomial time, high-dimensional problems still pose a challenge both in terms of time and memory requirements. In this work, we investigate how sketching techniques can remedy these issues. We provide reductions both for feasibility problems for dual SDPs and optimization problems for primal SDPs. These procedures are black box reductions in the sense that they allow us to compress the original problem to a smaller instance of the same problem which can subsequently be solved using the same techniques as for the original problem.

The main tool we use in this work is Johnson-Lindenstrauss transforms (JLTs), which are defined in Definition 2.1. In Lemma 3.1, we show that a conjugation by JLTs preserves the Hilbert-Schmidt inner product of a set of Hermitian matrices up to a small error which scales with the trace norm of these matrices. While this idea is very simple, we show in Theorem 3.2 that the scaling of the error cannot be improved using linear sketches which are positive maps.

In Section 4, we use this knowledge to sketch feasibility problems of dual SDPs. Theorem 4.2 shows that a sketched version of the feasibility problem obtained by a conjugation with a JLT is infeasible with high probability if the original problem is infeasible. As feasibility of the original problem always implies feasibility of the sketched problem, this means that we can consider a potentially smaller problem to check feasibility of the original problem with high probability.

In Section 5, we consider primal SDPs and show how to find an approximate solution to such optimization problems through sketching techniques. Again, we conjugate the matrices specifying the SDP with a JLT and relax the constraints by a small amount to obtain what we call the sketched SDP in Definition 5.2, which is defined by matrices of potentially lower dimension. Theorems 5.3 and 5.5 show that the value of the sketched SDP is not too far from the original value with high probability under some assumptions on the original problem. In particular, we require that the trace norms of the matrices specifying the SDP and the trace norm of an optimial solution are constant in the problem size. Theorem 5.1 shows that sketching techniques cannot work for any SDP, thus demonstrating the limitations of this approach.

In Section 6 we discuss the time complexity and memory requirements of our approach and compare it to other methods to solve SDPs. We show that the bottleneck of our method is to compute the matrices obtained by conjugation with a JLT. For problems for which our methods are applicable, we obtain an exponential reduction in the memory requirements needed to solve the SDP.

The project's idea was motivated by discussions between Daniel Stilck França and me. Daniel Stilck França is the main author of this contribution, who also had the idea of using the conjugation by Johnson-Lindenstrauss transforms and considering the relaxed SDPs. In particular, he proved Lemma 3.1, Theorem 3.2, Theorem 5.1, Theorem 5.3 and Theorem 5.7 in the article and wrote the majority of the text for the first draft. I proved Theorem 5.5 and Proposition 5.6 and wrote Section 6. I was involved in all work with exception of the parts mentioned above.

Permission to include:

Andreas Bluhm and Daniel Stilck França
Dimensionality reduction of SDPs through sketching.
Linear Algebra and its Applications, 563, 461–475 (2019).

Personal use

Authors can use their articles, in full or in part, for a wide range of scholarly, non-commercial purposes as outlined below:

- Use by an author in the author's classroom teaching (including distribution of copies, paper or electronic)
- Distribution of copies (including through e-mail) to known research colleagues for their personal use (but not for Commercial Use)
- Inclusion in a thesis or dissertation (provided that this is not to be published commercially)
- Use in a subsequent compilation of the author's works
- Extending the Article to book-length form
- Preparation of other derivative works (but not for Commercial Use)
- Otherwise using or re-using portions or excerpts in other works

These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article. In all cases we require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.

Linear Algebra and its Applications 563 (2019) 461–475

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Dimensionality reduction of SDPs through sketching

Andreas Bluhm, Daniel Stilck França *

Department of Mathematics, Technical University of Munich, 85748 Garching, Germany

A R T I C L E I N F O

Article history:
Received 22 August 2017
Accepted 6 November 2018
Available online 12 November 2018
Submitted by H. Rauhut

MSC:
primary 90C22
secondary 15A39, 15B48, 68W20

Keywords:
Semidefinite programming
Sketching
Dimensionality reduction
Johnson–Lindenstrauss transforms

A B S T R A C T

We show how to sketch semidefinite programs (SDPs) using positive maps in order to reduce their dimension. More precisely, we use Johnson–Lindenstrauss transforms to produce a smaller SDP whose solution preserves feasibility or approximates the value of the original problem with high probability. These techniques allow to improve both complexity and storage space requirements. They apply to problems in which the Schatten 1-norm of the matrices specifying the SDP and also of a solution to the problem is constant in the problem size. Furthermore, we provide some results which clarify the limitations of positive, linear sketches in this setting.

1. Introduction

Semidefinite programs (SDPs) are a prominent class of optimization problems [1]. They have applications across different areas of science and mathematics, such as discrete optimization [14] or control theory [2].

* Corresponding author.
E-mail addresses: andreas.bluhm@ma.tum.de (A. Bluhm), dsfranca@mytum.de (D. Stilck França).

https://doi.org/10.1016/j.laa.2018.11.012

However, although there are many different algorithms that solve an SDP up to an error ϵ in a time that scales polynomially with the dimension and logarithmically with ϵ^{-1} [4], solving large instances of SDPs still remains a challenge. This is not only due to the fact that the number and cost of the iterations scale superquadratically with the dimension for most algorithms to solve SDPs, but also due to the fact that the memory required to solve large instances is beyond current capabilities. This has therefore motivated research on algorithms that can solve SDPs, or at least obtain an approximate solution, with less memory requirements. One example are the so called first order methods, which were developed to remedy the high memory requirements of interior point methods; see [11] and references therein. Another such example is the recent [18], where ideas similar to ours were applied to achieve optimal storage requirements necessary to solve a certain class of SDPs. While the latter work proposes a new way to solve an SDP using linear sketches, our approach relies on standard convex optimization methods.

In this work, we develop algorithms to estimate the value of an SDP with linear inequality constraints and to determine if a given linear matrix inequality (LMI) is feasible or not. These algorithms convert the original problem to one of the same type, but of smaller dimension, which we call the sketched problem. Subsequently, this new problem can be solved with the same techniques as the original one, but potentially using less memory and achieving a smaller runtime. Therefore, we call this a black box algorithm. With high probability an optimal solution to the sketched problem allows us to obtain a good approximation of the value or to test the feasibility of the problem.

In the case of LMIs, if the sketched problem is infeasible, we obtain a certificate that the original problem is also infeasible. If the sketched problem is feasible, we are able to infer that the original problem is either "close to feasible" or feasible with high probability, under some technical assumptions.

In the case of estimating the value of SDPs, we are able to give an upper bound that holds with high probability and a lower bound on the value of the SDP from the value of the sketched problem, again under some technical assumptions. For a certain class of SDPs, which includes the so-called semidefinite packing problems [9], we are able to find a feasible point of the original problem which is close to the optimal point and most technical aspects simplify significantly.

Our algorithms work by conjugating the matrices that define the constraints of the SDP with Johnson–Lindenstrauss transforms [17], thereby preserving the structure of the problem. Similar ideas have been proposed to reduce the memory usage and complexity of solving linear programs [13]. While those techniques aim to reduce the number of constraints, our goal is to reduce the dimension of the matrices involved.

Unfortunately, the dimension of the sketch needed to have a fixed error with high probability scales with the Schatten 1-norm of both the constraints and of an optimal solution to the SDP, which significantly restricts the class of problems to which these methods can be applied. We are able to show that one cannot significantly improve this scaling and that one cannot sketch general SDPs using linear maps.

This paper is organized as follows: in Section 2, we fix our notation and recall some basic notions from matrix analysis, Johnson–Lindenstrauss transforms, semidefinite programs and convex analysis which we will need throughout the paper. We then proceed to show how to sketch the Hilbert–Schmidt scalar product with positive maps in Section 3. We apply these techniques in Section 4 to show how to certify that certain LMIs are infeasible by showing the infeasibility of an LMI of smaller dimension. In Section 5, we apply similar ideas to estimate the value of an SDP with linear inequality constraints by solving an SDP of lower dimension. We conclude with a discussion of the possible gains in the complexity of solving these problems and for the memory requirements in Section 6.

2. Preliminaries

We begin by fixing our notation. For brevity, we will write the set $\{1, \dots, d\}$ as $[d]$. The positive vectors will denoted by $\mathbb{R}^m_+ := \{x \in \mathbb{R}^m : x_i \geq 0\}$. The set of real $d \times D$ matrices will be written as $\mathcal{M}_{d,D}$ and just $\mathcal{M}_d$ if $d = D$. We will denote by $\mathcal{M}_d^{sym}$ the set of symmetric $d \times d$ matrices. For $A \in \mathcal{M}_d$, A^T will denote the transpose of A. To avoid cumbersome notation and redundant theorems, we will the statements only for real matrices. However, note that all statements translate to the complex case in a straightforward fashion. For $A \in \mathcal{M}_d^{sym}$ we will write $A \geq 0$ if A is positive semidefinite. We will denote the cone of $d \times d$ positive semidefinite matrices by $\mathcal{S}_d^+$ and its interior, the positive definite matrices, by $\mathcal{S}_d^{++}$. For the Schatten p-norm for $p \in [1, \infty)$ of a matrix $A \in \mathcal{M}_d$ we will write

$$\|A\|_p := \mathrm{Tr}[(A^T A)^{\frac{p}{2}}]^{\frac{1}{p}}.$$

The $p = \infty$ norm is the usual operator norm. The Schatten-2 norm is often called the Hilbert–Schmidt (HS) norm and is induced by the Hilbert–Schmidt scalar product, which is given by $\langle A, B\rangle_{HS} = \mathrm{Tr}\left(A^T B\right)$.

A linear map $\Phi : \mathcal{M}_D \to \mathcal{M}_d$ is called positive if $\Phi(\mathcal{S}_D^+) \subseteq \mathcal{S}_d^+$. We will mostly consider maps of the form $\Phi(X) = SXS^T$ with $S \in \mathcal{M}_{d,D}$.

The following families of matrices will play a crucial role for our purposes:

Definition 2.1 *(Johnson–Lindenstrauss transform).* A random matrix $S \in \mathcal{M}_{d,D}$ is a Johnson–Lindenstrauss transform (JLT) with parameters (ϵ, δ, k) if with probability at least $1 - \delta$, for any k-element subset $V \subseteq \mathbb{R}^D$, for all $v, w \in V$ it holds that

$$|\langle Sv, Sw\rangle - \langle v, w\rangle| \leq \epsilon \|v\|_2 \|w\|_2 .$$

Note that one usually only demands that the norm of the vectors involved is distorted by at most ϵ in the definition of JLTs, but this is equivalent to the definition we chose by the polarization identity. There are many different examples of JLTs in the literature and we refer to [17] and references therein for more details. Most of the constructions of JLTs focus on real matrices, but the generalization to complex matrices is straightforward. One simple example are random matrices $S = \frac{1}{\sqrt{d}}R \in \mathcal{M}_{d,D}$ with R having i.i.d. standard Gaussian random variables, which can be shown to be (ϵ, δ, k)-JLT if $d = \Omega(\epsilon^{-2} \log(k\delta^{-1}))$ [17, Lemma 2.12].

It will later be of advantage to our algorithm to consider JLTs with a desired sparsity s and we mention the following almost optimal result. We refer to [10, Section 1.1] for a proof and remark that the proof is constructive.

Theorem 2.2 *(Sparse JLT [10, Section 1.1]). There is an (ϵ, δ, k)-JLT $S \in \mathcal{M}_{d,D}$ with $d = \mathcal{O}\left(\epsilon^{-2} \log(k\delta^{-1})\right)$ and $s = \mathcal{O}(\epsilon^{-1} \log(k\delta^{-1}))$ nonzero entries per column.*

Given some JLT $S \in \mathcal{M}_{d,D}$, the positive map $\Phi : \mathcal{M}_D \to \mathcal{M}_d$, $X \mapsto SXS^T$ will be called the sketching map and d the sketching dimension.

We will now fix our notation for semidefinite programs. Semidefinite programs are a class of optimization problems in which a linear functional is optimized under linear constraints over the set of positive semidefinite matrices. We refer to [1] for an introduction to the topic. There are many equivalent ways of formulating SDPs. In this work, we will assume w.l.o.g. that the SDPs are given in the following form:

Definition 2.3 *(Sketchable SDP).* Let $A, B_1, \ldots, B_m \in \mathcal{M}_D^{\text{sym}}$ and $\gamma_1, \ldots, \gamma_m \in \mathbb{R}$. We will call the constrained optimization problem

$$\begin{aligned} \text{maximize} \quad & \operatorname{Tr}(AX) \\ \text{subject to} \quad & \operatorname{Tr}(B_i X) \leq \gamma_i, \qquad i \in [m] \\ & X \geq 0, \end{aligned} \tag{1}$$

a sketchable SDP.

Sometimes we will also refer to a sketchable SDP as the original problem. We will see later how to approximate the value of these SDPs. SDPs have a rich duality theory [1]. The dual problem of a sketchable SDP is given by the following:

$$\begin{aligned} \text{minimize} \quad & \langle c, \gamma \rangle \\ \text{subject to} \quad & \sum_{i=1}^{m} c_i B_i - A \geq 0 \\ & c \in \mathbb{R}_+^m, \end{aligned} \tag{2}$$

where $\gamma \in \mathbb{R}^m$ is the vector with coefficients γ_i. SDPs and LMIs will be called feasible if there is at least one point satisfying all the constraints, otherwise we will call them infeasible. A sketchable SDP will be called strictly feasible if there is a point $X > 0$ such that all the constraints in (1) are satisfied with strict inequality. Under some conditions, such as Slater's condition [1], the primal problem (1) and the dual problem (2) have the same value. This is called strong duality.

We will need some standard concepts from convex analysis. Given $a_1, \dots, a_n \in V$ for a vector space V, we denote by $\mathrm{conv}\{a_1, \dots, a_n\}$ the convex hull of the points. By $\mathrm{cone}\{a_1, \dots, a_n\}$ we will denote the cone generated by these elements and a convex cone C will be called pointed if $C \cap -C = \{0\}$.

3. Sketching the Hilbert–Schmidt product with positive maps

One of our main ingredients to sketch an SDP or LMI will be a random positive map $\Phi : \mathcal{M}_D \to \mathcal{M}_d$ that preserves the Hilbert–Schmidt scalar product with high probability. We demand positivity to assure that the structure of the SDP or LMI is preserved. Below, we first consider the example $\Phi(X) = SXS^T$ with S a JLT. A similar estimate was proved in [12] for a different application.

Lemma 3.1. *Let $B_1, \dots, B_m \in \mathcal{M}_D^{\mathrm{sym}}$ and $S \in \mathcal{M}_{d,D}$ be an (ϵ, δ, k)-JLT with $\epsilon \leq 1$ and k such that*

$$k \geq \sum_{i=1}^{m} \mathrm{rank} B_i.$$

Then

$$\mathbb{P}\left[\forall i, j \in [m] : \left|\mathrm{Tr}\left(SB_iS^TSB_jS^T\right) - \mathrm{Tr}\left(B_iB_j\right)\right| \leq 3\epsilon \left\|B_i\right\|_1 \left\|B_j\right\|_1\right] \geq 1 - \delta. \quad (3)$$

Proof. Observe that the eigenvectors of the B_i corresponding to nonzero eigenvalues of the B_i form a subset of cardinality at most k of $\mathbb{R}^D$. Let $A, B \in \{B_1, \dots, B_m\}$. As S is an (ϵ, δ, k)-JLT, with probability at least $1 - \delta$ we have for all normalized eigenvectors a_i of A and b_j of B that

$$\left|\left|\langle Sa_i, Sb_j\rangle\right| - \left|\langle a_i, b_j\rangle\right|\right| \leq \epsilon \quad (4)$$

by the reverse triangle inequality. We also have that for any a_i, b_j

$$\left\|Sa_i\right\|_2 \leq \sqrt{1+\epsilon}, \quad \left\|Sb_j\right\|_2 \leq \sqrt{1+\epsilon},$$

again by the fact that S is a JLT. As $\epsilon \leq 1$ and by the Cauchy–Schwarzc inequality, it follows that

$$\left|\langle Sa_i, Sb_j\rangle\right| + \left|\langle a_i, b_j\rangle\right| \leq 3 \quad (5)$$

and hence, by multiplying (5) with (4),

$$\left| |\langle Sa_i, Sb_j\rangle|^2 - |\langle a_i, b_j\rangle|^2 \right| \leq 3\epsilon. \tag{6}$$

Now let λ_i and μ_j be the eigenvalues of A and B, respectively. We have:

$$\begin{aligned}\left|\mathrm{Tr}\left(SAS^TSBS^T\right) - \mathrm{Tr}\left(AB\right)\right| &= \left|\sum_{i,j=1}^{D} \lambda_i\mu_j(|\langle Sa_i, Sb_j\rangle|^2 - |\langle a_i, b_j\rangle|^2)\right| \\ &\leq 3\epsilon \sum_{i,j=1}^{D} |\lambda_i||\mu_j| = 3\epsilon \left\|A\right\|_1 \left\|B\right\|_1\end{aligned}$$

with probability at least $1-\delta$. As A, B were arbitrary, the claim follows. □

The scaling of the error with the Schatten 1-norm of the matrices involved in Lemma 3.1 is highly undesirable, as the norm might grow linearly with the dimension. Applying JLTs for the Hilbert space $\mathcal{M}_D^{\mathrm{sym}}$ would give a scaling of the error with the Schatten 2-norm, but it would not necessarily preserve positivity of the matrices. The next theorem shows that a scaling of the error with the Schatten 2-norm of the matrices involved is not possible with positive maps if we want to achieve a non-trivial compression. Therefore, we cannot hope for a much better error dependence even with more advanced tools than the crude estimates which we have used.

Theorem 3.2. *Let $\Phi : \mathcal{M}_D \to \mathcal{M}_d$ be a random positive map such that with strictly positive probability for any $Y_1, \dots Y_{D+1} \in \mathcal{M}_D$ and $0 < \epsilon < \frac{1}{4}$ we have*

$$|\mathrm{Tr}\left(\Phi(Y_i)^T\Phi(Y_j)\right) - \mathrm{Tr}\left(Y_i^TY_j\right)| \leq \epsilon\|Y_i\|_2\|Y_j\|_2. \tag{7}$$

Then $d = \Omega(D)$.

Proof. Let $\{e_i\}_{1\leq i\leq D}$ be an orthonormal basis of $\mathbb{R}^D$ and define $X_i = e_ie_i^T$. As Equation (7) is satisfied with positive probability, there must exist a positive map $\Phi : \mathcal{M}_D \to \mathcal{M}_d$ such that Equation (7) is satisfied for $Y_i = X_i$, $i \in [D]$, and $Y_{D+1} = \mathbb{1}$. As the X_i are orthonormal w.r.t. the Hilbert Schmidt scalar product and Φ is positive we have for $i, j \in [D]$

$$\mathrm{Tr}\left(\Phi(X_i)\Phi(X_j)\right) \in \begin{cases}[0, \epsilon], & \text{for } i \neq j \\ [1-\epsilon, 1+\epsilon], & \text{for } i = j.\end{cases} \tag{8}$$

Define the matrix $A \in \mathcal{M}_D$ with $(A)_{ij} = \mathrm{Tr}\left(\Phi(X_i)\Phi(X_j)\right)$ for $i, j \in [D]$. It is clear that A is symmetric and that its entries are positive. We have

$$\sum_{i,j\in[D]} A_{ij} = \mathrm{Tr}\left(\Phi(\mathbb{1})\Phi(\mathbb{1})\right) \in \left[(1-\epsilon)D, (1+\epsilon)D\right].$$

As $A_{ii} \geq (1-\epsilon)$, it follows that

$$\sum_{i \neq j} A_{ij} \leq 2\epsilon D. \tag{9}$$

Let

$$J = \{(i,j) \in [D] \times [D] | i \neq j, A_{ij} \leq \frac{1}{D}\}.$$

It follows from Equation (9) that $|\{(i,j) \in [D] \times [D] | i \neq j, (i,j) \notin J\}| \leq 2D^2\epsilon$ and so

$$|J| \geq \left((1-2\epsilon)D^2 - D\right).$$

Since for $(i,j) \in J$ also $(j,i) \in J$, we can write $J = (I \times I) \backslash \{(i,i) | i \in I\}$ for $I \subseteq [D]$. Thus,

$$|J| = |I|(|I|-1) \geq ((1-2\epsilon)D^2 - D) \geq \left(\frac{1}{2} - 2\epsilon\right) D^2,$$

for $D \geq 2$. From this it follows that

$$|I|^2 \geq |I|(|I|-1) \geq \left(\frac{1}{2} - 2\epsilon\right) D^2,$$

and we finally obtain

$$|I| \geq \sqrt{1/2 - 2\epsilon} D. \tag{10}$$

Notice that it follows from Equation (8) that we may rescale all the X_i to X_i' such that $\mathrm{Tr}\left(\Phi(X_i')^2\right) = 1$ and the pairwise scalar product still satisfies $\mathrm{Tr}\left(\Phi(X_i')\Phi(X_j')\right) \leq \frac{1}{D(1-\epsilon)}$ for $(i,j) \in J$. If there is an $N \in \mathbb{N}$ such that $d > \sqrt{1/2 - 2\epsilon} D$ for all $D \geq N$, the claim follows. We therefore now suppose that $d \leq \sqrt{1/2 - 2\epsilon} D$. Hence, $d \leq |I|$ by Equation (10). By the positivity of Φ and the fact that the X_i' are positive semidefinite, we have that $\Phi(X_i')$ is positive semidefinite. In [16, Proposition 2.7] it is shown that for any set $\{P_i\}_{i \in I}$ of $|I| \geq d$ positive semidefinite matrices in $\mathcal{M}_d$ such that $\mathrm{Tr}\left(P_i^2\right) = 1$ we have that

$$\sum_{i \neq j} \mathrm{Tr}\left(P_i P_j\right)^2 \geq \frac{(|I|-d)^2 |I|}{(|I|-1)d^2}.$$

By the definition of the set J, we have that

$$\sum_{(i,j) \in J} \mathrm{Tr}\left(\Phi(X_i')\Phi(X_j')\right)^2 \leq \frac{|J|}{(1-\epsilon)^2 D^2} \leq \frac{1}{(1-\epsilon)^2},$$

as $|J| \leq D^2$. From Equation (10) it follows that

$$\frac{1}{(1-\epsilon)^2} \geq \left(\frac{\sqrt{1/2 - 2\epsilon}D}{d} - 1\right)^2$$

and after some elementary computations we finally obtain

$$d \geq \frac{(1-\epsilon)\sqrt{1/2-2\epsilon}}{2-\epsilon}D. \quad \square$$

It remains open if one could achieve a better compression for a sublinear number of matrices. We also note that other theorems that restrict the possibility of dimensionality reduction using positive maps were proved in [8], although their results are restricted to maps that are in addition trace preserving and they also demand that the distribution of maps is highly symmetric.

4. Sketching linear matrix inequality feasibility problems

In this section we will show how to use JLTs to certify that certain linear matrix inequalities (LMI) are infeasible by showing that an LMI of smaller dimension is infeasible. The following lemma is similar in spirit to the well-known Farkas' lemma.

Lemma 4.1. *Let* $A, B_1, \ldots, B_m \in \mathcal{M}_D^{sym} \backslash \{0\}$ *such that*

$$\sum_{i=1}^{m} c_i B_i - A \not\geq 0 \tag{11}$$

for all $c \in \mathbb{R}_+^m$. *Suppose further that*

$$\Lambda = \text{cone}\{B_1, \ldots, B_m\}$$

is pointed and $\Lambda \cap S_D^+ = \{0\}$. *Then there exists a* $\rho \in \mathcal{S}_D^+$ *such that for all* $i \in [m]$

$$\text{Tr}(\rho B_i) < 0, \quad \text{Tr}(-A\rho) < 0 \quad \textit{and} \quad \text{Tr}(\rho) = 1. \tag{12}$$

Proof. Let $E = \text{conv}\{-A, B_1, \ldots, B_m\}$. We will show that $S_D^+ \cap E = \emptyset$. Suppose there exists an $X = -p_0 A + \sum_{i=1}^{m} p_i B_i \in S_D^+ \cap E$ with $p \in [0,1]^{m+1}$. If $p_0 > 0$, we could rescale X by p_0^{-1} and obtain a feasible point for (11), a contradiction. If $p_0 = 0$ and $X \neq 0$, this would in turn contradict $\Lambda \cap S_D^+ = \{0\}$. And if $X = 0$, the cone Λ would not be pointed. From these arguments it follows that $0 \notin E$. The set E is therefore closed, convex, compact and disjoint from the convex and closed set $\mathcal{S}_D^+$. We may thus find a hyperplane that strictly separates $\mathcal{S}_D^+$ from E. That is, a $\rho \in \mathcal{M}_D^{\text{sym}}$ such that w.l.o.g.

$\mathrm{Tr}(\rho X) \geq 0$ for all $X \in \mathcal{S}_D^+$, as $0 \in \mathcal{S}_D^+$, and $\mathrm{Tr}(Y\rho) < 0$ for all $Y \in E$. As $\mathrm{Tr}(\rho X) \geq 0$ for all $X \geq 0$, it follows that ρ is positive semidefinite and it is clear that by normalizing ρ we may choose ρ with $\mathrm{Tr}(\rho) = 1$. □

The main idea is now to show that under these conditions we may sketch the hyperplane in a way that it still separates the set of positive semidefinite matrices and the sketched version of the set $\{\sum_{i=1}^m \gamma_i B_i - A | \gamma_i \geq 0\}$.

Theorem 4.2. *Let $A, B_1, \ldots, B_m \in \mathcal{M}_D^{sym} \backslash \{0\}$ such that they satisfy the assumptions of Lemma 4.1. Moreover, let $\rho \in \mathcal{S}_D^+$ be as in Equation (12). Set*

$$\epsilon = \frac{1}{6} \min \left\{ \left| \frac{\mathrm{Tr}(\rho B_1)}{\|B_1\|_1} \right|, \ldots, \left| \frac{\mathrm{Tr}(\rho B_m)}{\|B_m\|_1} \right|, \left| \frac{\mathrm{Tr}(\rho A)}{\|A\|_1} \right| \right\}$$

and take $S \in \mathcal{M}_{d,D}$ to be an (ϵ, δ, k)-JLT. Here,

$$k \geq \mathrm{rank} A + \mathrm{rank} \rho + \sum_{i=1}^m \mathrm{rank} B_i.$$

Then

$$\sum_{i=1}^m c_i S B_i S^T - S A S^T \not\geq 0 \tag{13}$$

for all $c \in \mathbb{R}_+^m$, with probability at least $1 - \delta$.

Proof. It should first be noted that ρ exists and $|\mathrm{Tr}(A\rho)| > 0$, $|\mathrm{Tr}(B_i\rho)| > 0$ for all $i \in [m]$ by Lemma 4.1. Therefore, also $\epsilon > 0$. The matrix ρ defines a hyperplane that strictly separates the set

$$E = \left\{ \sum_{i=1}^m c_i B_i - A \middle| c \in \mathbb{R}_+^m \right\}$$

and S_D^+. We will now show that $S\rho S^T$ strictly separates the sets

$$E_S = \left\{ \sum_{i=1}^m c_i S B_i S^T - S A S^T \middle| c \in \mathbb{R}_+^m \right\}$$

and S_d^+ with probability at least $1 - \delta$, from which the claim follows. Note that by our choice of ρ and ϵ, it follows from Lemma 3.1 that we have

$$\mathrm{Tr}\left(S\rho S^T S B_i S^T\right) \leq \mathrm{Tr}(\rho B_i) + 3\epsilon \|B_i\|_1 < 0$$

with probability at least $1-\delta$ and similarly for $-A$ instead of B_i. Therefore, it follows that $\mathrm{Tr}(ZS\rho S^T) < 0$ for all $Z \in E_S$. As $S\rho S^T$ is a positive semidefinite matrix, it follows that $\mathrm{Tr}\left(YS\rho S^T\right) \geq 0$ for all $Y \in \mathcal{S}_d^+$. We have therefore found a strictly separating hyperplane for E_S and S_D^+ and the LMI (13) is infeasible. □

Theorem 4.2 suggests a way of sketching feasibility problems of the form

$$\sum_{i=1}^{m} c_i B_i - A \geq 0, \qquad c \in \mathbb{R}_+^m. \tag{14}$$

To obtain more concrete bounds on the probability that the original problem is infeasible although the sketched problem is feasible, one would need to know the parameter ϵ, which is not possible in most applications.

5. Approximating the value of semidefinite programs through sketching

We will now show how to approximate with high probability the value of a sketchable SDP by first conjugating both the target matrix and the matrices that describe the constraints with JLTs and subsequently solving a smaller SDP. The next theorem shows that in general it is not possible to approximate with high probability the value of a sketchable SDP using linear sketches.

Theorem 5.1. *Let $\Phi : \mathcal{M}_{2D} \to \mathbb{R}^d$ be a random linear map such that for all sketchable SDPs there exists an algorithm which allows us to estimate the value of an SDP up to a constant factor $1 \leq \tau < \frac{2}{\sqrt{3}}$ given the sketch $\{\Phi(A), \Phi(B_1), \ldots, \Phi(B_m)\}$ with probability at least $9/10$. Then $d = \Omega(D^2)$.*

Proof. It is well-known that the operator norm of a matrix $G \in \mathcal{M}_D$ can be computed via an SDP. A linear sketch of the constraints of this SDP would thus allow to approximately compute the operator norm with high probability. However, in [17, Theorem 6.5] it was shown that any algorithm that estimates the operator norm of a matrix from a linear sketch with probability larger than $9/10$ must have sketch dimension $\Omega(D^2)$. □

The above result remains true even if we restrict to SDPs that have optimal points with small Schatten 1-norm and low rank. However, we will see below that sketching becomes possible if the matrices that define the constraints and the target function have a small Schatten 1-norm.

Definition 5.2 *(Sketched SDP).* Let $A, B_1, \ldots, B_m \in \mathcal{M}_D^{\mathrm{sym}}$, $\eta, \gamma_1, \ldots, \gamma_m \in \mathbb{R}$ and $\epsilon > 0$. Let $X^* \in \mathcal{S}_D^+$ be an optimal point of the sketchable SDP defined through these matrices. Given that $\mathrm{Tr}\left(X^*\right) \leq \eta$ and given a random matrix $S \in \mathcal{M}_{d,D}$, we call the optimization problem

$$
\begin{aligned}
\text{maximize} \quad & \operatorname{Tr}\left(SAS^TY\right) \\
\text{subject to} \quad & \operatorname{Tr}\left(SB_iS^TY\right) \leq \gamma_i + \mu \left\|B_i\right\|_1, \qquad i \in [m] \\
& Y \geq 0
\end{aligned}
\tag{15}
$$

with $\mu = 3\epsilon\eta$ the sketched SDP.

The motivation for defining the sketched SDP is given by the following theorem, which follows directly form Lemma 3.1.

Theorem 5.3. *Let $A, B_1, \ldots, B_m \in \mathcal{M}_D^{\mathrm{sym}}$, $\eta, \gamma_1, \ldots, \gamma_m \in \mathbb{R}$ and $\epsilon > 0$. Denote by α the value of the sketchable SDP and assume it is attained at an optimal point X^* which satisfies $\operatorname{Tr}(X^*) \leq \eta$. Moreover, let $S \in \mathcal{M}_{d,D}$ be an (ϵ, δ, k)-JLT, with*

$$
k \geq \operatorname{rank} X^* + \operatorname{rank} A + \sum_{i=1}^{m} \operatorname{rank} B_i.
$$

Let α_S be the value of the sketched SDP defined by A, B_i and S. Then

$$
\alpha_S + 3\epsilon\eta \left\|A\right\|_1 \geq \alpha
$$

with probability at least $1 - \delta$.

Note that Theorem 5.3 does not rule out the possibility that the value of the sketched problem is much larger than that of the sketchable SDP. To investigate this issue, we introduce the following:

Definition 5.4 *(Relaxed SDP).* Let $A, B_1, \ldots, B_m \in \mathcal{M}_D^{\mathrm{sym}}$, $\eta, \gamma_1, \ldots, \gamma_m \in \mathbb{R}$ and $\epsilon > 0$. Given that an optimal point X^* of the sketchable SDP defined through these matrices satisfies $\operatorname{Tr}(X^*) \leq \eta$, we call the optimization problem

$$
\begin{aligned}
\text{maximize} \quad & \operatorname{Tr}(AX) \\
\text{subject to} \quad & \operatorname{Tr}(B_iX) \leq \gamma_i + \tilde{\epsilon}_i, \quad i \in [m] \\
& X \geq 0
\end{aligned}
\tag{16}
$$

with $\tilde{\epsilon}_i = 3\epsilon\eta \left\|B_i\right\|_1$ the relaxed SDP.

We will obtain lower bounds on the value of the sketchable SDP in terms of the value of the sketched SDP through continuity bounds on the relaxed SDP. The method of using duality to derive perturbation bounds for a convex optimization problem used here is standard and we refer to [5, Section 5.6] for a similar derivation. We denote by $\mathcal{A}(\tilde{\epsilon})$ the feasible set of the relaxed SDP as in Definition 5.4 for some $\tilde{\epsilon} \in \mathbb{R}_+^m$. With this notation, $\mathcal{A}(0)$ is the feasible set of the sketchable SDP. Analogously, we denote by $\alpha(\tilde{\epsilon})$ and $\alpha(0)$

the optimal value of the relaxed problem and of the sketchable SDP, respectively. Note that the following result is not probabilistic and holds regardless of the sketching matrix S used.

Theorem 5.5. *We are in the setting of Definition 2.3. Assume that there exists an $X_0 > 0$ such that all the constraints of the sketchable SDP are strictly satisfied and that the dual problem is feasible. Then, the value of the sketched SDP α_S is bounded by*

$$\alpha_S \leq \alpha(0) + C \|y^*\|_1 .$$

Here y^* is an optimal solution to the dual problem and

$$C = \max\{3\epsilon\eta \|B_i\|_1 |i \in [m]\},$$

where $\eta \geq \mathrm{Tr}\,(X^*)$ for an optimal point X^* of the sketchable SDP.

Proof. By Slater's condition [15, Theorem 2.2], strong duality holds and there is a $y^* \geq 0$ which achieves the optimal value. Note that, given a feasible point Y to the sketched SDP, $S^T Y S$ is a feasible point for the relaxed problem by the cyclicity of the trace. Thus, the relaxed SDP gives an upper bound for the sketched SDP. Hence, for any $X \geq 0$,

$$\begin{aligned} \alpha(0) &\geq \sum_{j=1}^{m} y_j^* \gamma_j - \mathrm{Tr}\left(\left[\sum_{i=1}^{m} y_i^* B_i - A\right] X\right) \\ &= \mathrm{Tr}\,(AX) - \sum_{i=1}^{m} y_i^* \left[\mathrm{Tr}\,(B_i X) - \gamma_i\right]. \end{aligned} \tag{17}$$

The first line holds by duality. If we take the supremum over $X \in \mathcal{A}(\epsilon)$, we obtain

$$\alpha(\tilde{\epsilon}) \leq \alpha(0) + \langle \tilde{\epsilon}, y^* \rangle,$$

from $y_i^* \geq 0$. Here, $\tilde{\epsilon}_i = 3\eta\epsilon \|B_i\|_1$, $i \in [m]$. The assertion then follows by an application of Hölder's inequality. □

Combining Theorem 5.3 and Theorem 5.5 it is possible to pick ϵ small enough to have an arbitrarily small additive error under some structural assumptions on the SDP. That is, we need bounds on the Schatten 1-norms both of A and B_i and we need a bound on the Schatten 1-norm of an optimal solution to the sketchable SDP. Moreover, we need a bound on the 1-norm of a dual solution as in [3]. The following proposition provides a generic bound of this kind.

Proposition 5.6. *Assume that there exists $X_0 \in \mathcal{A}(0)$ such that $X_0 > 0$ and the constraints are strictly satisfied. Then the value of the sketched SDP α_S is bounded by*

$$\alpha_S \leq \alpha(0) + \epsilon C_1 \left(\alpha(0) - \mathrm{Tr}\left(AX_0\right)\right)/C_2.$$

Here,

$$C_1 = \max\{3\eta \left\|B_i\right\|_1 |i \in [m]\},$$
$$C_2 = \min\{(\gamma_i - \mathrm{Tr}\left(B_i X_0\right)) |i \in [m]\},$$

where $\eta \geq \mathrm{Tr}\left(X^*\right)$ *for an optimal point* X^* *of the sketchable SDP.*

Proof. We need to bound $\|y^*\|_1$ in Theorem 5.5. From Equation (17) and $[\mathrm{Tr}\left(B_i X_0\right) - \gamma_i] < 0$, it follows that

$$\sum_{i=1}^{m} y_i^* \leq (\alpha(0) - \mathrm{Tr}\left(AX_0\right))/ \min_{i\in[m]} \left[\gamma_i - \mathrm{Tr}\left(B_i X_0\right)\right].$$

With $y^* \geq 0$, the assertion follows from Theorem 5.5. □

In the case that all the $\gamma_i > 0$ for a sketchable SDP we may obtain a bound on the value and an approximate solution to it in a much simpler way. This class includes the so-called semidefinite packing problems [9]. These are defined as problems in which all $B_i \geq 0$, and so also $\gamma_i \geq 0$. Note that we may set all $\gamma_i = 1$ w.l.o.g. by dividing B_i by γ_i. We then obtain:

Theorem 5.7. *For a sketchable SDP with* $\gamma_i = 1$ *and* $\nu = 3\epsilon\eta \max_{i\in[m]} \|B_i\|_1$, *we have that*

$$\frac{\alpha_S}{1+\nu} \leq \alpha. \tag{18}$$

Moreover, denoting by X_S^* *an optimal point of the sketched SDP, we have that* $\frac{1}{1+\nu} S^T X_S^* S$ *is a feasible point of the sketchable SDP that attains this lower bound.*

Proof. The lower bound in Equation (18) follows immediately from the cyclicity of the trace, as $\frac{1}{1+\nu} S^T X_S^* S$ is a feasible point of the sketchable SDP. □

6. Complexity and memory gains

In this section, we will discuss how much we gain by considering the sketched SDP instead of the sketchable SDP. We focus on the results of Section 5, but the discussion carries over to the results of Section 4. Throughout this section we will assume that we are guaranteed that the Schatten 1-norms both of an optimal solution to our SDP and of the matrices that define the constraints are $\mathcal{O}(1)$. We will suppose that upper bounds on the Schatten 1-norm of both an optimal solution and the constraints are given.

To generate the sketched SDP, we need to compute $m+1$ matrices of the form SBS^T, where $B \in \mathcal{M}_D$. Each of this computations needs $\mathcal{O}(\max\{\mathrm{nnz}(B), Dd\}\epsilon^{-1}\log(k\delta^{-1}))$ operations. In the worst case, when all matrices $\{A, B_1, \ldots, B_m\}$ are dense and have full rank, this becomes $\mathcal{O}(mD^2\log(mD))$ operations to generate the sketched SDP for fixed ϵ and δ. We obtain from these considerations and Theorem 5.3:

Proposition 6.1. *Let $A, B_1, \ldots, B_m \in \mathcal{M}_D^{\mathrm{sym}}$, $\gamma_1, \ldots, \gamma_m \in \mathbb{R}$ of a sketchable SDP be given. Furthermore, let $z := \max\{\mathrm{nnz}(A), \mathrm{nnz}(B_1), \ldots, \mathrm{nnz}(B_m)\}$ and $\mathrm{SDP}(m, d, \zeta)$ be the complexity of solving a sketchable SDP (up to accuracy ζ) of dimension d. Then a number of*

$$\mathcal{O}(\max\{z, D\epsilon^{-2}\log(k\delta^{-1})\}\epsilon^{-1}m\log(k\delta^{-1}) + \mathrm{SDP}(m, \epsilon^{-2}\log(k\delta^{-1}), \zeta))$$

operations is needed to generate and solve the sketched SDP, where k is defined as in Theorem 5.3.

Typically, the costs of forming the sketched matrices SB_iS^T dominates the overall complexity. To compare the above result to other methods for solving SDPs, let us fix ϵ, δ and ζ. Then, the ellipsoid method [7, Chapter 3] needs $\mathcal{O}(\max\{m, D^2\}D^6)$ operations to solve the sketchable SDP, whereas using interior point methods we need $\mathcal{O}(\max\{m^3, D^2m^2, mD^\omega\}D^{0.5}\log(D))$ operations [6, Chapter 5]. Here, ω is the exponent of matrix multiplication. Compared to that, forming the sketched problem and then solving it requires $\mathcal{O}(mD^2\log(mD))$ operations.

Another advantage is that using our methods, we need store much smaller matrices.

Proposition 6.2. *Let $A, B_1, \ldots, B_m \in \mathcal{M}_D^{\mathrm{sym}}$, $\gamma_1, \ldots, \gamma_m \in \mathbb{R}$ be a sketchable SDP. Then we need only to store $\mathcal{O}(m\epsilon^{-4}\log(k/\delta)^2)$ entries for the sketched problem, where k is defined as in Theorem 5.3.*

Numerical experiments with random instances of SDPs and LMIs that satisfy our requirements indicate that our methods may decrease the runtime of SDPs by one order of magnitude. Moreover, they allow us to solve problems in dimensions that are larger by one order of magnitude.

Acknowledgements

We would like to thank Ion Nechita for helpful discussions. A.B. acknowledges support from the ISAM Graduate Center at Technical University of Munich. D.S.F. acknowledges support from the graduate program TopMath of the Elite Network of Bavaria, the TopMath Graduate Center of TUM Graduate School at Technical University of Munich. D.S.F. is supported by the Technical University of Munich – Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763.

References

[1] M.F. Anjos, J.B. Lasserre, Handbook on Semidefinite, Conic and Polynomial Optimization, International Series in Operations Research & Management Science Series, Springer, 2016.

[2] S. Boyd, L. El Ghaoui, E. Feron, V. Balakrishnan, Linear Matrix Inequalities in System and Control Theory, Studies in Applied Mathematics, vol. 15, SIAM, 1994.

[3] F.G.S.L. Brandao, K.M. Svore, Quantum speed-ups for solving semidefinite programs, in: FOCS, 2017, pp. 415–426.

[4] S. Bubeck, Convex optimization: algorithms and complexity, Found. Trends Mach. Learn. 8 (3–4) (2015) 231–357.

[5] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[6] E. de Klerk, Aspects of Semidefinite Programming: Interior Point Algorithms and Selected Applications. Applied Optimization, Springer, US, 2002.

[7] M. Grötschel, L. Lovász, A. Schrijver, Geometric Algorithms and Combinatorial Optimization, Springer, 1988.

[8] A.W. Harrow, A. Montanaro, A.J. Short, Limitations on quantum dimensionality reduction, in: ICALP, 2011.

[9] G. Iyengar, D.J. Phillips, C. Stein, Approximation algorithms for semidefinite packing problems with applications to maxcut and graph coloring, in: IPCO, Springer, 2005, pp. 152–166.

[10] D.M. Kane, J. Nelson, Sparser Johnson–Lindenstrauss transforms, J. ACM 61 (1) (2014) 1–23.

[11] James Renegar, Efficient first-order methods for linear programming and semidefinite programming, arXiv:1409.5832, 2014.

[12] C.J. Stark, A.W. Harrow, Compressibility of positive semidefinite factorizations and quantum models, IEEE Trans. Inform. Theory 62 (5) (2016) 2867–2880.

[13] K. Vu, P.-L. Poirion, L. Liberti, Using the Johnson–Lindenstrauss lemma in linear and integer programming, arXiv:1507.00990, 2015.

[14] H. Wolkowicz, M.F. Anjos, Semidefinite programming for discrete optimization and matrix completion problems, Discrete Appl. Math. 123 (1–3) (2002) 513–577.

[15] J. Watrous, Semidefinite programs for completely bounded norms, Theory Comput. 5 (2009) 217–238.

[16] M.M. Wolf, Quantum channels and operations: guided tour, Lecture notes, available at http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf, 2012.

[17] D.P. Woodruff, Sketching as a tool for numerical linear algebra, Found. Trends Theor. Comput. Sci. 10 (1–2) (2014) 1–157.

[18] A. Yurtsever, M. Udell, J.A. Tropp, V. Cevher, Sketchy decisions: convex low-rank matrix optimization with optimal storage, arXiv:1702.06838, 2017.

B.2 Compatibility of quantum measurements and inclusion constants for the matrix jewel

Compatibility of quantum measurements and inclusion constants for the matrix jewel

Andreas Bluhm and Ion Nechita

While Core Article II proves that the compatibility of binary quantum measurements corresponds to an inclusion problem of free spectrahedra involving the matrix diamond, this work shows that the connection still holds for measurements with an arbitrary number of outcomes if one replaces the matrix diamond with a generalization thereof which we call the matrix jewel.

In Section 3 of this paper, we review some results from convex analysis together with results on the inclusion of free spectrahedra and on the compatibility of quantum measurements. Most notably, we introduce the direct sum of free spectrahedra in Section 3.3 and study its properties. In Lemma 3.15, we show that the direct sum of two maximal free spectrahedra for two polytopes yields the maximal spectrahedron for the direct sum of the polytopes. The matrix jewel is defined in Definition 4.1 as the direct sum of smaller free spectrahedra which we call the matrix jewel bases. Lemma 3.15 shows that the matrix jewel is the maximal free spectrahedron for a direct sum of simplices, where each simplex in $k-1$ dimensions corresponds to a measurement with k outcomes.

In Section 5, we make the connection between the compatibility of quantum measurements and the inclusion of the matrix jewel into a free spectrahedron defined by the measurements. As in Theorem V.3 of Core Article II, the three parts of Theorem 5.2 show that the different levels of the inclusion are in one-to-one correspondence with different degrees of compatibility. Theorem 5.3 then plays the role of Theorem V.7 in Core Article II. It states that the balanced compatibility region of g measurements in dimension d with k_i outcomes each, $i \in [g]$, corresponds to the inclusion set of the matrix jewel with the same parameters which is defined in Definition 3.10.

Since the matrix jewel has not been studied before in the literature on free spectrahedra, we use results from the study of mutually unbiased bases and approximate cloning reviewed in Section 6.1 and 6.2 to obtain bounds on the inclusion set of the matrix diamond. In Section 7, we use a symmetrization technique to bound the inclusion set of the matrix jewel using the inclusion set for the matrix diamond, which has been studied before. Using the correspondence in Theorem 5.3 of the inclusion set to the balanced compatibility region, the results of Theorems 7.1 and 7.2 translate to new bounds for quantum incompatibility. The new bounds both on the inclusion set for the matrix jewel and on the balanced compatibility region are discussed in Section 10.

Finally, we introduce in Section 8 the notion of an incompatibility witness for binary measurements, which we generalize in Section 9 to measurements with an arbitrary number of outcomes. For binary measurements, incompatibility witnesses are defined in Definition 8.1. They can be used to certify that a given set of measurements is incompatible. Proposition 8.3 shows that incompatibility witnesses are related to another free spectrahedron, the matrix cube. Since the latter is well-studied, this gives a numerically efficient method to check the compatibility of binary quantum measurements as discussed in Remark 8.5.

I was significantly involved in finding the ideas and carrying out the scientific work of all parts of this article. Furthermore, I was in charge of the writing of all parts with the exception of Sections 8 and 9.

Permission to include:

Andreas Bluhm and Ion Nechita
Compatibility of quantum measurements and inclusion constants for the matrix jewel.
arXiv preprint arXiv:1809.04514 (2018).
Submitted to *SIAM Journal on Applied Algebra and Geometry.*

Society for Industrial and Applied Mathematics (SIAM)

Consent to Publish

SIAM ("Publisher") requires Authors of articles in SIAM publications to provide a formal written Consent to Publish. The Author must sign the agreement except, in the case of "work-for-hire", when the Author's employer may sign as the party that has the right to grant rights to the Publisher. If there are multiple Authors of the material governed by this document, the term "Author" as used here refers to each and all of them, jointly and severally. [1]

Title of Contribution ("Work"): Compatibility of quantum measurements\\and inclusion constants for the matrix jewel
Authors: Andreas Bluhm and Ion Nechita
Name of Journal: SIAM Journal on Applied Algebra and Geometry
Manuscript Number: M123837

1. Author's Warranty

By signing this Consent, the Author warrants all of the following: The Work has not been published before in any form except as a preprint, unless explicitly noted as a footnote to the title. The Work is not being concurrently submitted to and is not under consideration by another publisher. The names listed above as authors appear in the manuscript itself, no author entitled to credit has been omitted, and there are no unnamed authors. The Author has the right to make the grants made to the Publisher complete and unencumbered. The Author also warrants that the Work does not libel anyone, violate anyone's privacy or publicity rights, infringe anyone's copyright, trademark, or trade secrets, or otherwise violate anyone's statutory or common law rights.

2. Author's Rights

A1. The Author may reproduce and distribute the Work (including derivative works) in connection with the Author's teaching, technical collaborations, conference presentations, lectures, or other scholarly works and professional activities as well as to the extent the fair use provisions of the U.S. Copyright Act permit. If the copyright is granted to the Publisher, then the proper notice of the Publisher's copyright should be provided.
A2. The Author may post the final draft of the Work, as it exists immediately prior to editing and production by the Publisher, on noncommercial pre-print servers such as arXiv.org.
A3. The Author may post the final published version of the Work on the Author's personal web site and on the web server of the Author's institution, provided that proper notice of the Publisher's copyright is included and that no separate or additional fees are collected for access to or distribution of the work.

3. Publisher's Rights

Even if the Author does not transfer Copyright to the Publisher, the Author grants the Publisher the following rights in perpetuity.

P1. The Publisher has unlimited rights throughout the world to publish and distribute the final version of the Work in any form and in all media now known or hereafter discovered.
P2. The Publisher has unlimited rights throughout the world to translate the final version of the Work and exercise all rights in all media in the resulting

COMPATIBILITY OF QUANTUM MEASUREMENTS AND INCLUSION CONSTANTS FOR THE MATRIX JEWEL

ANDREAS BLUHM AND ION NECHITA

Abstract. In this work, we establish the connection between the study of free spectrahedra and the compatibility of quantum measurements with an arbitrary number of outcomes. This generalizes previous results by the authors for measurements with two outcomes. Free spectrahedra arise from matricial relaxations of linear matrix inequalities. A particular free spectrahedron which we define in this work is the matrix jewel. We find that the compatibility of arbitrary measurements corresponds to the inclusion of the matrix jewel into a free spectrahedron defined by the effect operators of the measurements under study. We subsequently use this connection to bound the set of (asymmetric) inclusion constants for the matrix jewel using results from quantum information theory and symmetrization. The latter translate to new lower bounds on the compatibility of quantum measurements. Among the techniques we employ are approximate quantum cloning and mutually unbiased bases.

Contents

1. Introduction 1
2. Main results 2
3. Preliminaries 4
3.1. Convex analysis 4
3.2. Free spectrahedra 6
3.3. The direct sum of free spectrahedra 7
3.4. Quantum information theory 9
4. The matrix jewel 11
5. The matrix jewel and joint measurability of POVMs 13
6. Compatibility results from quantum information theory 15
6.1. Upper bounds from MUBs 16
6.2. Lower bounds from cloning 17
7. Lower bounds from symmetrization 20
8. Incompatibility witnesses and the matrix cube 21
9. Incompatibility witnesses – the general case 25
10. Discussion 27
References 30

1. Introduction

Given the solution set of a linear matrix inequality, the question often arises whether the unit cube is contained in this set (see Section 1.5 of [HKMS19] and references therein). However, this problem, which is known as the matrix cube problem, is known to be NP-hard [BTN02]. Fortunately, there exists a tractable relaxation of this problem which checks inclusion of corresponding free spectrahedra, which are matricial relaxations of the original sets [BTN02, HKM13]. To give error bounds for this relaxation, it is necessary to know the following: if inclusion for the original

Date: March 26, 2019.

spectrahedra holds, how much do we have to shrink the smaller free spectrahedron such that inclusion also holds at the level of free spectrahedra? For the matrix cube, as well as for unit balls of ℓ_p spaces and other highly symmetric convex sets, these inclusion constants have been recently studied [HKMS19, DDOSS17, PSS18].

Recently, the authors have found that the inclusion constants for the free spectrahedral relaxation of the ℓ_1-ball, the matrix diamond [DDOSS17], are relevant for the joint measurability of binary quantum measurements [BN18]. The fact that not all observables can be measured at the same time is one of the most remarkable properties of quantum mechanics, the observables of position and momentum providing the best-known example of this behavior [Hei27, Boh28]. The notion of joint measurability (or compatibility) has been introduced to capture this property of non-classical theories (see [HMZ16] for a review). In this work, we model quantum measurements by Positive Operator Valued Measures (POVMs), see [HZ11, Section 3.1]. POVMs are jointly measurable if they arise as marginals from a common measurement. This property is of practical interest, since only POVMs which are not jointly measurable can violate Bell inequalities [Fin82] or can be used for some quantum information tasks [BCP+14].

The present work continues the line of research started in [BN18]. While the previous work focused on measurements with only two outcomes, we establish here the connection between the joint measurability of POVMs with an arbitrary number of outcomes and the inclusion of the *matrix jewel*. The matrix jewel is a free spectrahedron which generalizes the matrix diamond and is introduced in this work. We can subsequently use this connection to translate results on joint measurability into bounds on the inclusion constants for the matrix jewel. Some of the techniques used involve approximate cloning of quantum states and mutually unbiased bases. Moreover, we compare the matrix jewel to more symmetric free spectrahedra such as the matrix diamond to obtain lower bounds on the inclusion constants of the matrix jewel. These translate to new bounds on the compatibility of quantum measurements.

We also introduce the notion of *incompatibility witnesses*, which are tuples of self-adjoint matrices that allow, in a simple way, to show that some POVMs are not compatible (the terminology is borrowed from entanglement theory).

The paper is organized as follows. After presenting informally our main results in Section 2, we recall in Section 3 some facts from (matricial) convexity theory and quantum information theory; we also introduce at that point two new operations on free spectrahedra, the Cartesian product and the direct sum. Sections 4 and 5 are the core of the paper: we introduce the matrix jewel and we relate its inclusion properties to compatibility of POVMs. In Section 6, we use several results from quantum information theory and symmetrization to give lower and upper bounds on the inclusion sets of the matrix jewel. In Sections 8 and 9 we develop the theory of incompatibility witnesses. The final section contains a review of our main contributions, as well as some open questions and future research directions.

2. Main results

In this section, we will review the main results of the present work. It is a follow-up paper on the work undertaken in [BN18]. We continue investigating the connection between free spectrahedral inclusion problems and joint measurability of quantum effects.

Quantum measurements are identified with *positive operator valued measures* (POVMs). Those are k-tuples of positive semidefinite matrices of fixed dimension which sum to the identity. Here, k is the number of measurement outcomes the quantum measurement has. Given a g-tuple of POVMs $E^{(1)}, \dots, E^{(g)}$, where the i-th POVM has k_i outcomes, we can ask the question whether these POVMs are *jointly measurable*. Joint measurability means that there is a joint POVM $\{ G_{i_1,\dots,i_g} \}$ with $i_j \in [k_j]$ from which the POVMs $E^{(j)}$ arise as marginals. Although not all measurements in quantum theory are compatible, they can be made compatible if we add a sufficient amount of noise. In this work, we focus on balanced noise, i.e. the elements of the j-th POVM become

$$\tilde{E}_i^{(j)} = s_j E_i^{(j)} + (1 - s_j)\frac{1}{k_j}I, \tag{1}$$

where $s_j \in [0, 1]$. This means, that with probability s_j we measure the original POVM $E^{(j)}$ whereas with probability $1-s_j$, we output a measurement outcome uniformly at random, independent of the system under study. The set of g-tuples s with the property that, for any g-tuple of d-dimensional POVMs $E^{(j)}$ with k_j outcomes, the noisy POVMs $\tilde{E}^{(j)}$ from (1) are compatible, will be written as $\Gamma(g, d, (k_1, \ldots, k_g))$, and will be called the *balanced compatibility region.*

A *free spectrahedron* is a special type of matrix convex set which arises as matricial relaxation of an ordinary linear matrix inequality. The free spectrahedron $\mathcal{D}_A$ for the self-adjoint matrix g-tuple A is the set of self-adjoint matrix g-tuples X of arbitrary dimension which fulfill the matrix inequality

$$\sum_{i=1}^{g} A_i \otimes X_i \leq I.$$

For scalar X, we recover the solution set $\mathcal{D}_A(1)$ of the linear matrix inequality defined by A. The free spectrahedral inclusion problem is to determine for which $s \in \mathbb{R}_+^g$ the implication

$$\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1) \implies s \cdot \mathcal{D}_A \subseteq D_B \tag{2}$$

is true. We will be interested in the case where the object on the left hand side is the *matrix jewel.* Consider the free spectrahedron given by the diagonal matrices $\mathrm{diag}[v_j]$, $j \in [k-1]$, where

$$v_j(\varepsilon) = -\frac{2}{k} + 2\delta_{\varepsilon,j} \qquad \forall \varepsilon \in [k].$$

We call this spectrahedron the matrix jewel base $\mathcal{D}_{\diamond,k}$. The matrix jewel $\mathcal{D}_{\diamond,(k_1,\ldots,k_g)}$ is then the direct sum of the $\mathcal{D}_{\diamond,k_i}$. We define the direct sum of free spectrahedra arising from polytopes as the maximal spectrahedron which has the direct sum of these polytopes at the scalar level. The matrix jewel is a generalization of the matrix diamond introduced in [DDOSS17] and considered in relation to quantum effect compatibility in [BN18]. We are interested in the vectors of the form $s = (s_1^{\times(k_1-1)}, \ldots, s_g^{\times(k_g-1)})$ for which the implication in Equation (2) is true for $\mathcal{D}_A = \mathcal{D}_{\diamond,(k_1,\ldots,k_g)}$ and any self-adjoint tuple B on the right hand side; we are using the notation

$$(s_1^{\times(k_1-1)}, \ldots, s_g^{\times(k_g-1)}) := (\underbrace{s_1, \ldots, s_1}_{k_1-1 \text{ times}}, \ldots, \underbrace{s_g, \ldots, s_g}_{k_g-1 \text{ times}}).$$

We call the set of these vectors the *inclusion set for the matrix jewel* $\Delta(g, d, (k_1, \ldots, k_g))$

The main contribution of this work is then the connection of the free spectrahedral inclusion problem to the problem of joint measurability. In Theorem 5.2 we find

Theorem. *For a fixed matrix dimension d, consider g tuples of self-adjoint matrices $E^{(i)} \in (\mathcal{M}_d^{sa})^{k_i-1}$, $k_i \in \mathbb{N}$, $i \in [g]$. Define $E_{k_i}^{(i)} := I_d - E_1^{(i)} \ldots - E_{k_i-1}^{(i)}$, set $\mathbf{k} = (k_1, \ldots, k_g)$, and write*

$$\begin{aligned}\mathcal{D}_E &:= \mathcal{D}_{(2E^{(1)} - \frac{2}{k_1}I, \ldots, 2E^{(g)} - \frac{2}{k_g}I)} \\ &= \bigsqcup_{n=1}^{\infty} \left\{ X \in (\mathcal{M}_n^{sa})^{\sum_{i=1}^{g}(k_i-1)} : \sum_{i=1}^{g}\sum_{j=1}^{k_i-1} \left(2E_j^{(i)} - \frac{2}{k_i}I\right) \otimes X_{i,j} \leq I_{dn} \right\}.\end{aligned}$$

Then

(1) $\mathcal{D}_{\diamond,\mathbf{k}}(1) \subseteq \mathcal{D}_E(1)$ *if and only if* $\left\{ E_1^{(i)}, \ldots, E_{k_i}^{(i)} \right\}$, $i \in [g]$ *are POVMs.*

(2) $\mathcal{D}_{\diamond,\mathbf{k}} \subseteq \mathcal{D}_E$ *if and only if* $\left\{ E_1^{(i)}, \ldots, E_{k_i}^{(i)} \right\}$, $i \in [g]$ *are jointly measurable POVMs.*

(3) $\mathcal{D}_{\diamondsuit,\mathbf{k}}(l) \subseteq \mathcal{D}_E(l)$ *for* $l \in [d]$ *if and only if for any isometry* $V : \mathbb{C}^l \hookrightarrow \mathbb{C}^d$*, the tuples* $\left\{ V^* E_1^{(i)} V, \ldots, V^* E_{k_i}^{(i)} V \right\}$, $i \in [g]$ *are jointly measurable POVMs.*

This extends [BN18, Theorem V.3] from binary measurements to measurements with k_i outcomes each. We find that the different levels of spectrahedral inclusion correspond to different degrees of joint measurability. Furthermore, we show in Theorem 5.3 that the balanced compatibility region and the inclusion set for the matrix jewel can be identified; again, this is a generalization of [BN18, Theorem V.7] for an arbitrary number of outcomes.

Theorem. *Let* d, $g \in \mathbb{N}$ *and* $(k_1, \ldots, k_g) \in \mathbb{N}^g$*. Then,*

$$\Gamma(g, d, (k_1, \ldots, k_g)) = \Delta(g, d, (k_1, \ldots, k_g)).$$

This identification allows to use results on one set to characterize the other. In [BN18], we mostly adapted results from the study of free spectrahedral inclusion to characterize the balanced compatibility region in quantum information theory. This was possible, since the matrix diamond (the matrix jewel for $k_i = 2$ for all i) is a highly symmetric object and has already been studied in the literature. The matrix jewel does not have these symmetries and has not been studied in the algebraic convexity literature. Therefore, we adapt results from quantum information theory in Section 6, which we subsequently use in Section 10 to give upper and lower bounds on $\Delta(g, d, (k_1, \ldots, k_g))$. The lower bounds come from asymmetric approximate cloning of quantum states and from two different symmetrization procedures. The latter yield new lower bounds on the balanced compatibility region of quantum measurements. General upper bounds can be imported from the case of binary POVMs, since more outcomes shrink the compatibility regions and therefore also the corresponding inclusion sets. For the case of $k_i = d$ and g not too large, we get better bounds from the study of measurements arising from mutually unbiased bases (MUBs).

We also introduce in this paper the notion of *incompatibility witnesses*, both in the case of binary POVMs (Section 8) and general POVMs (Section 9). As in the case of compatibility conditions, the theory in the binary case is simpler and the corresponding free spectrahedra have already been studied extensively in the mathematical literature. For these reasons, let us focus here on binary POVMs.

A g-tuple of self-adjoint matrices $X \in (\mathcal{M}_n^{sa})^g$ is called an incompatibility witness if X is an element of the matrix diamond $\mathcal{D}_{\diamond,g}$, i.e. if $\sum_{i=1}^g \varepsilon_i X_i \leq I_n$ for all sign vectors $\varepsilon \in \{\pm 1\}^g$. An incompatibility witness X can certify that g given effects are incompatible: If the matrix inequality

$$\sum_{i=1}^{g} (2E_i - I_d) \otimes X_i \leq I_{dn}$$

does not hold, the effects $E_1, \ldots, E_g$ are incompatible. There is a strong connection between incompatibility witnesses and the matrix cube (arguably the most studied class of free spectrahedra): X is an incompatibility witness if and only if $\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_X(1)$. Using the inclusion constants for the (complex) matrix cube, one can obtain tractable relaxations for the two equivalent conditions above (which otherwise require checking an exponential number of matrix inequalities).

3. Preliminaries

This section contains some facts from (algebraic) convexity and quantum information theory which will be needed in the following sections. The material here is for the most part well known, with the exception of Section 3.3.

3.1. Convex analysis.

Before we move on to the main topic of this section, let us fix some basic notation. We will often write $[n] := \{ 1, \ldots, n \}$ for brevity, where $n \in \mathbb{N}$. Furthermore, we will use $\mathbb{R}_+^g := \{ x \in \mathbb{R}^g : x_i \geq 0 \ \forall i \in [g] \}$, where $g \in \mathbb{N}$. Let n, $m \in \mathbb{N}$. Then, $\mathcal{M}_{n,m}$ is the set of complex $n \times m$ matrices and we will write just $\mathcal{M}_n$ if $m = n$. For the self-adjoint matrices, we will write

$\mathcal{M}_n^{sa}$. By $\mathcal{U}(d)$ we will denote the unitary $d \times d$ matrices. Moreover, we will write $I_n \in \mathcal{M}_n$ for the identity matrix, where we will often omit the subscript if the dimension is clear from the context. The *operator system* generated by the g-tuple $A \in (\mathcal{M}_d^{sa})^g$ is defined as

$$\mathcal{OS}_A := \operatorname{span}\{\, I_d, A_i : i \in [g]\,\}.$$

Furthermore, we will often write for such g-tuples $2A - I := (2A_1 - I_d, \ldots, 2A_g - I_d)$ and $V^*AV := (V^*A_1V, \ldots, V^*A_gV)$ with $V \in \mathcal{M}_{d,k}$, $k \in \mathbb{N}$.

We start with two standard objects in convex analysis, polytopes and polyhedra (c.f. [Bar02, Definition I.2.2]).

Definition 3.1. *The convex hull of a finite set of points in $\mathbb{R}^d$, $d \in \mathbb{N}$, is called a* polytope. *Let $c_1, \ldots, c_m$ be vectors in $\mathbb{R}^d$ and let $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$. The set*

$$\mathcal{P} := \left\{\, x \in \mathbb{R}^d : \langle c_i, x\rangle \le \alpha_i \quad \forall i \in [m]\,\right\}$$

is called a polyhedron.

By the Weyl-Minkowski theorem, a convex subset of $\mathbb{R}^d$ is a polytope if and only if it is a bounded polyhedron [Bar02, Corollary II.4.3]. We will need the following lemma, which follows easily from convexity:

Lemma 3.2 ([Bar02, Section IV.1])**.** *Let $\mathcal{P} = \operatorname{conv}(\{\, v_1, \ldots v_m\,\}) \subset \mathbb{R}^d$, $m \in \mathbb{N}$. Then, its polar dual can be written as $\mathcal{P}^\circ = \{\, x \in \mathbb{R}^d : \langle v_i, x\rangle \le 1 \quad \forall i \in [m]\,\}$.*

There are several ways of constructing new convex sets from a collection of given ones. One way is the Cartesian product:

Definition 3.3. *Let $\mathcal{P}_1 \subseteq \mathbb{R}^{k_1}$, $\mathcal{P}_2 \subseteq \mathbb{R}^{k_2}$ be two convex sets. Then, their* Cartesian product *is*

$$\mathcal{P}_1 \times \mathcal{P}_2 := \left\{\, (x, y) \in \mathbb{R}^{k_1+k_2} : x \in \mathcal{P}_1, y \in \mathcal{P}_2\,\right\}.$$

Another one is the direct sum:

Definition 3.4. *Let $\mathcal{P}_1 \subseteq \mathbb{R}^{k_1}$, $\mathcal{P}_2 \subseteq \mathbb{R}^{k_2}$ be two convex sets. Then, their* direct sum *is*

$$\mathcal{P}_1 \oplus \mathcal{P}_2 := \operatorname{conv}\left(\left\{\, (x, 0) \in \mathbb{R}^{k_1+k_2} : x \in \mathcal{P}_1\,\right\} \cup \left\{\, (0, y) \in \mathbb{R}^{k_1+k_2} : y \in \mathcal{P}_2\,\right\}\right).$$

Remark 3.5. *In particular, the above definition shows that the direct sum of two polytopes is again a polytope, because it is the convex hull of their respective extreme points embedded into a higher dimensional space.*

We can find a useful expression for the direct sum of two polytopes in terms of the Cartesian product and taking polars. We include a short proof for convenience.

Lemma 3.6 ([Bre97, Lemma 2.4])**.** *Let $\mathcal{P}_1 \subseteq \mathbb{R}^{k_1}$, $\mathcal{P}_2 \subseteq \mathbb{R}^{k_2}$ be two polytopes and such that $0 \in \mathcal{P}_1$, $\mathcal{P}_2$. Then,*

$$\mathcal{P}_1 \oplus \mathcal{P}_2 = (\mathcal{P}_1^\circ \times \mathcal{P}_2^\circ)^\circ.$$

Proof. Using Lemma 3.2, we may write

$$(\mathcal{P}_1 \oplus \mathcal{P}_2)^\circ = \left\{\, (x_1, x_2) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} : \langle p_i, x_i\rangle \le 1 \quad \forall p_i \in \mathcal{P}_i,\ i \in [2]\,\right\}.$$

Comparing this with the definition of $\mathcal{P}_i^\circ$, we find that $(\mathcal{P}_1 \oplus \mathcal{P}_2)^\circ = \mathcal{P}_1^\circ \times \mathcal{P}_2^\circ$. As the $\mathcal{P}_i$ are polytopes, they are compact and thus also $\mathcal{P}_1 \oplus \mathcal{P}_2$ is compact. As this set furthermore contains 0 by assumption, an application of the Bipolar Theorem [Bar02, Theorem IV.1.2] yields the claim. □

Later, we shall need the following result on the faces of the Cartesian product.

Lemma 3.7 ([Bre97, Lemma 2.3])**.** *Let $\mathcal{P}_1 \subseteq \mathbb{R}^{k_1}$, $\mathcal{P}_2 \subseteq \mathbb{R}^{k_2}$ be two polytopes. Then, the l-dimensional faces of $\mathcal{P}_1 \times \mathcal{P}_2$, for $0 \le l \le k_1 + k_2$ are the $\mathcal{F}_1 \times \mathcal{F}_2$, where $\mathcal{F}_i$ is a j_i-dimensional face of $\mathcal{P}_i$ and $j_1 + j_2 = l$.*

3.2. **Free spectrahedra.** In this section, we will review some basic results from the theory of free spectrahedra. The theory we will need for this work can be found in [HKM13, HKMS19, DDOSS17].

Let $A \in (\mathcal{M}_d^{sa})^g$ be a g-tuple of self-adjoint matrices. The *free spectrahedron at level* n defined by A is the set

$$\mathcal{D}_A(n) := \left\{ X \in (\mathcal{M}_n^{sa})^g : \sum_{i=1}^{g} A_i \otimes X_i \leq I_{nd} \right\}.$$

The *free spectrahedron* corresponding to A is then the union of all these levels, i.e.

$$\mathcal{D}_A := \bigsqcup_{n \in \mathbb{N}} \mathcal{D}_A(n).$$

Let $\mathcal{C} \subseteq \mathbb{R}^g$ be a convex set. In general, there are many free spectrahedra $\mathcal{D}_A$ with $\mathcal{D}_A(1) = \mathcal{C}$. If $\mathcal{C}$ is a polyhedron with 0 in its interior, we can find a maximal such free spectrahedron [DDOSS17, Definition 4.1]:

$$\mathcal{W}_{max}(\mathcal{C})(n) := \tag{3}$$
$$\left\{ X \in (\mathcal{M}_n^{sa})^g : \sum_{i=1}^{g} c_i X_i \leq \alpha I, \quad \forall\, c \in \mathbb{R}^g, \forall \alpha \in \mathbb{R} \ \text{ s.t. } \ \mathcal{C} \subseteq \{\, x \in \mathbb{R}^g : \langle c, x \rangle \leq \alpha \,\} \right\}.$$

Note that $\mathcal{W}_{max}(\mathcal{C})(1) = \mathcal{C}$, as claimed above.

Remark 3.8. *It is clear that the above is indeed a free spectrahedron, since polyhedra are defined as the intersection of* finitely *many hyperplanes (see Definition 3.1). The defining matrices can thus be chosen diagonal and of finite dimension. The fact that* 0 *is an interior point guarantees that we can always choose* $\alpha = 1$.

Remark 3.9. *The definition above can be used to define matrix convex sets for any convex set* $\mathcal{C}$*. If* $\mathcal{C}$ *is not a polyhedron or* 0 *not in the interior, however, the corresponding* $\mathcal{W}_{max}(\mathcal{C})$ *is not necessarily a free spectrahedron. See* [DDOSS17] *for details.*

In this work, we will be concerned with inclusion constants, i.e. constants for which the implication

$$\mathcal{D}_A(1) \subseteq \mathcal{D}_B(1) \implies s \cdot \mathcal{D}_A \subseteq \mathcal{D}_B$$

holds, where A, B are both g-tuples of self-adjoint matrices. Here, the *(asymmetrically) scaled* free spectrahedron is

$$s \cdot \mathcal{D}_A := \{\, (s_1 X_1, \ldots, s_g X_g) : X \in \mathcal{D}_A \,\}.$$

Definition 3.10. *Let* $D \in \mathbb{N}$ *and* $\mathcal{D}_A$ *be the free spectrahedron defined by* $A := (A^{(1)}, \ldots, A^{(g)})$*, where* $A^{(j)} \in (\mathcal{M}_D^{sa})^{k_j - 1}$*,* $k_j \in \mathbb{N}$*,* $j \in [g]$*. Let* $\mathbf{k} = (k_1, \ldots, k_g)$*. The* inclusion set *is defined as*

$$\Delta_{\mathcal{D}_A}(g, d, \mathbf{k}) :=$$
$$\left\{ s \in \mathbb{R}_+^g : \forall B \in (\mathcal{M}_d^{sa})^{\sum_{i=1}^{g}(k_i - 1)},\, \mathcal{D}_A(1) \subseteq \mathcal{D}_B(1) \implies (s_1^{\times (k_1 - 1)}, \ldots, s_g^{\times (k_g - 1)}) \cdot \mathcal{D}_A \subseteq \mathcal{D}_B \right\}.$$

If $\mathcal{D}_A$ *is the matrix jewel* $\mathcal{D}_{\diamond, \mathbf{k}}$ *in Definition 4.1, we will write* Δ *instead of* $\Delta_{\mathcal{D}_A}$.

This definition generalizes [BN18, Definition IV.1], which is recovered for $\mathbf{k} = (2, \ldots, 2)$. Note that the $(k_i - 1)$-tuples in the inclusion sets are scaled in the same way inside each group, where the size of the groups are determined by the vector $\mathbf{k}$. By the same argument as in [BN18, Proposition IV.3], these sets are convex.

The inclusion of free spectrahedra can be related to positivity properties of the map between the matrices defining them. Let $A \in (\mathcal{M}_D^{sa})^g$ and $B \in (\mathcal{M}_d^{sa})^g$. Let $\Phi : \mathcal{OS}_A \to \mathcal{M}_d$ be the unital map defined by

$$\Phi : A_i \mapsto B_i \qquad \forall i \in [g].$$

The following theorem has been proven in [HKM13, Theorem 3.5] for real matrices. See [BN18, Lemma IV.4] for a very similar proof in the complex case.

Lemma 3.11. *Let $A \in (\mathcal{M}_D^{sa})^g$ and $B \in (\mathcal{M}_d^{sa})^g$. Furthermore, let $\mathcal{D}_A(1)$ be bounded. Then, $\mathcal{D}_A(n) \subseteq \mathcal{D}_B(n)$ holds if and only if Φ as given above is n-positive. In particular, $\mathcal{D}_A \subseteq \mathcal{D}_B$ if and only if Φ is completely positive.*

3.3. **The direct sum of free spectrahedra.** In this section we introduce the operation of *direct sum* for free spectrahedra and relate it to the direct sum of polytopes. We derive some simple properties of this construction which will be used later in the paper. Here, we will identify $\mathbb{R}^d$ with the diagonal $d \times d$ matrices with real entries.

Definition 3.12. *Let $A \in (M_{d_1}^{sa})^{k_1}$, $B \in (M_{d_2}^{sa})^{k_2}$ be tuples of self-adjoint matrices.The* direct sum *of the corresponding free spectrahedra is defined as follows:*

$$\mathcal{D}_A \hat{\oplus} \mathcal{D}_B(n) := \left\{ X \in (\mathcal{M}_n^{sa})^{k_1+k_2} : \sum_{i=1}^{k_1} (A_i \otimes I_{d_2}) \otimes X_i + \sum_{j=1}^{k_2} (I_{d_1} \otimes B_j) \otimes X_{k_1+j} \leq I_{d_1 d_2 n} \right\}.$$

The following proposition shows that the above operation is indeed defined on free spectrahedra and not on tuples of self-adjoint matrices.

Proposition 3.13. *Let $A \in (M_{d_1}^{sa})^{k_1}$, $B \in (M_{d_2}^{sa})^{k_2}$, $\tilde{A} \in (M_{d_1'}^{sa})^{k_1}$ and $\tilde{B} \in (M_{d_2'}^{sa})^{k_2}$ be tuples of self-adjoint matrices such that $\mathcal{D}_A = \mathcal{D}_{\tilde{A}}$ and $\mathcal{D}_B = \mathcal{D}_{\tilde{B}}$. Furthermore, let $\mathcal{D}_A(1)$ and $\mathcal{D}_B(1)$ be bounded. Then,*

$$\mathcal{D}_A \hat{\oplus} \mathcal{D}_B = \mathcal{D}_{\tilde{A}} \hat{\oplus} D_{\tilde{B}}.$$

Proof. By Lemma 3.11, there exists a bijective map $\Phi : \mathcal{OS}_A \to \mathcal{OS}_{\tilde{A}}$ such that $\Phi(A_i) = \tilde{A}_i$ for all $i \in [k_1]$ which is unital and completely positive. In the same vein, there exists a bijective map $\Psi : \mathcal{OS}_B \to \mathcal{OS}_{\tilde{B}}$ with $\Psi(B_i) = \tilde{B}_i$ for all $i \in [k_2]$ which is also unital and completely positive. Let

$$\mathcal{A} = \operatorname{span}\{ I_{d_1 d_2}, A_i \otimes I_{d_2}, I_{d_1} \otimes B_j : i \in [k_1], j \in [k_2] \}$$

be the operator system corresponding to $\mathcal{D}_A \hat{\oplus} \mathcal{D}_B$ and $\tilde{\mathcal{A}}$ the same set with $\tilde{A}_i$, $\tilde{B}_j$, d_1' and d_2'. Then, $\Phi \otimes \Psi : \mathcal{A} \to \tilde{\mathcal{A}}$ is a unital completely positive map (this can be seen by extending Φ and Ψ to the full matrix algebras by Arveson's extension theorem, taking the tensor product and restricting again) which is bijective. Thus, another application of Lemma 3.11 yields the assertion. □

Remark 3.14. *It is not clear whether the above proposition is still true without the assumption that $\mathcal{D}_A(1)$ and $\mathcal{D}_B(1)$ are bounded. So in this case, strictly speaking, we define the operation on tuples of self-adjoint matrices. However, in this work, all relevant spectrahedra will be bounded.*

The following result connects the definition of the direct sum at the level of free spectrahedra with the usual definition for convex sets, Definition 3.4.

Lemma 3.15. *Let $\mathcal{P}_1$, $\mathcal{P}_2$ be two polytopes such that $0 \in \operatorname{int}(\mathcal{P}_i)$, $i \in [2]$. Then $\mathcal{W}_{max}(\mathcal{P}_1 \oplus \mathcal{P}_2) = \mathcal{W}_{max}(\mathcal{P}_1) \hat{\oplus} \mathcal{W}_{max}(\mathcal{P}_2)$.*

Proof. By a refined version of the Weyl-Minkowski theorem, [Bar02, Lemma VI.1.5], there exist $c_s^{(i)} \in \mathbb{R}^{k_i}$, $\alpha_s^{(i)} \in \mathbb{R}$ such that

$$\mathcal{P}_i = \left\{ x \in \mathbb{R}^{k_i} : \langle c_s^{(i)}, x \rangle \leq \alpha_s^{(i)} \quad \forall s \in [m_i] \right\},$$

where $m_i \in \mathbb{N}$. Furthermore, $\mathcal{F}_s^{(i)} = \left\{ p_i \in \mathcal{P}_i : \langle c_s^{(i)}, p_i \rangle = \alpha_s^{(i)} \right\}$ are the facets of $\mathcal{P}_i$. By assumption, $0 \in \operatorname{int}(\mathcal{P}_i)$, and thus $\alpha_s^{(i)} > 0$. Therefore, we can write

$$\mathcal{P}_i = \left\{ x \in \mathbb{R}^{k_i} : \sum_{j=1}^{k_i} x_j P_j^{(i)} \leq I_{m_i} \right\} = \left\{ x \in \mathbb{R}^{k_i} : \langle h_s^{(i)}, x \rangle \leq 1 \quad \forall s \in [m_i] \right\},$$

where $h_s^{(i)} = c_s^{(i)}/\alpha_s^{(i)}$ and $P_j^{(i)} \in \mathbb{R}^{m_i}$ such that $P_j^{(i)}(s) = h_s^{(i)}(j)$. Combining Lemma 3.7 and the fact that facets of a polytope correspond to extreme points of its polar [Bar02, Theorem VI.1.3], we find that the extreme points of $\mathcal{P}_1^\circ \times \mathcal{P}_2^\circ$ are $(h_{s_1}^{(1)}, h_{s_2}^{(2)})$, $s_i \in [m_i]$, $i \in [2]$. Using Lemma 3.6 and Lemma 3.2, we obtain

$$\mathcal{P}_1 \oplus \mathcal{P}_2 = \left\{ (x_1, x_2) \in \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} : \langle (h_{s_1}^{(1)}, h_{s_2}^{(2)}), (x_1, x_2) \rangle \leq 1 \quad \forall s_i \in [m_i],\, i = 1, 2 \right\}.$$

Thus, we find that the $(h_{s_1}^{(1)}, h_{s_2}^{(2)})$ are the hyperplanes defining $\mathcal{P}_1 \oplus \mathcal{P}_2$. Moreover, we can again write this in spectrahedral form,

$$\mathcal{P}_1 \oplus \mathcal{P}_2 = \left\{ x \in \mathbb{R}^{k_1+k_2} : \sum_{j=1}^{k_1+k_2} x_j Q_j \leq I_{m_1 m_2} \right\}.$$

Here, $Q_j \in \mathbb{R}^{m_1 m_2}$, where $Q_j(s_1, s_2) := (h_{s_1}^{(1)}, h_{s_2}^{(2)})_j$. Hence, by the definition of the maximal spectrahedron,

$$\mathcal{W}_{max}(\mathcal{P}_1 \oplus \mathcal{P}_2)(n) = \left\{ X \in (\mathcal{M}_n^{sa})^{k_1+k_2} : \sum_{j=1}^{k_1+k_2} Q_j \otimes X_j \leq I_{n m_1 m_2} \right\}.$$

Evaluating the expression for the Q_j further, we infer

$$\begin{aligned} Q_j(s_1, s_2) &= \begin{cases} h_{s_1}^{(1)}(j) = P_j^{(1)}(s_1) & 1 \leq j \leq k_1 \\ h_{s_2}^{(2)}(j - k_1) = P_{j-k_1}^{(2)}(s_2) & k_1 + 1 \leq j \leq k_1 + k_2 \end{cases} \\ &= \begin{cases} (P_j^{(1)} \otimes I_{k_2})(s_1, s_2) & 1 \leq j \leq k_1 \\ (I_{k_1} \otimes P_{j-k_1}^{(2)})(s_1, s_2) & k_1 + 1 \leq j \leq k_1 + k_2 \end{cases}. \end{aligned}$$

This proves the assertion. □

Remark 3.16. *The assumption* $0 \in \operatorname{int}(\mathcal{P})$ *is needed to ensure that the polytope* $\mathcal{P}$ *can be written as a linear matrix inequality.*

The next result shows that level-1 inclusion of the direct sum of two polytopes *into* a spectrahedron amounts to individual inclusion of each polytope into the corresponding part of the spectrahedron.

Lemma 3.17. *Let* $A^{(i)} \in (\mathcal{M}_d^{sa})^{k_i}$, $k_i \in \mathbb{N}$, $i = 1, 2$ *be two tuples of matrices and* $\mathcal{P}_j \subset \mathbb{R}^{k_j}$, $j = 1, 2$ *two polytopes. Then,*

$$\mathcal{P}_1 \oplus \mathcal{P}_2 \subseteq \mathcal{D}_{(A^{(1)}, A^{(2)})}(1) \iff \mathcal{P}_i \subseteq \mathcal{D}_{A^{(i)}}(1) \quad i = 1, 2.$$

Proof. Let $\left\{ w_j^{(i)} \right\}_{j=1}^{m_i} \subset \mathbb{R}^{k_i}$ be the set of extreme points of $\mathcal{P}_i$ with $m_i \in \mathbb{N}$. Then, the set of extreme points of $\mathcal{P}_1 \oplus \mathcal{P}_2$ is $\left\{ (w_{j_1}^{(1)}, 0), (0, w_{j_2}^{(2)}) : j_i \in [m_i],\, i = 1, 2 \right\}$. This can easily be seen from the definition. Since the inclusion of polytopes can be checked at the extreme points, the assertion follows. □

In a similar fashion, one can define the *Cartesian product* of two free spectrahedra as

(4)
$$(\mathcal{D}_A \hat{\times} \mathcal{D}_B)(n) := \left\{ X \in (\mathcal{M}_n^{sa})^{k_1+k_2} : \sum_{i=1}^{k_1} (A_i \oplus 0_{d_2}) \otimes X_i + \sum_{j=1}^{k_2} (0_{d_1} \oplus B_j) \otimes X_{k_1+j} \leq I_{(d_1+d_2)n} \right\}.$$

It is easy to check that, at level $n = 1$,

$$(\mathcal{D}_A \hat{\times} \mathcal{D}_B)(1) = \mathcal{D}_A(1) \times D_B(1).$$

3.4. **Quantum information theory.** We will conclude this section with a short review of some concepts from quantum information theory which we will use. For an introduction to the mathematics of quantum mechanics, see e.g. [HZ11] or [Wat18]. A quantum mechanical system is given as a *state* $\rho \in \mathcal{S}(\mathcal{H})$. Here, $\mathcal{H}$ is the Hilbert space of the system and

$$\mathcal{S}(\mathcal{H}) := \{ \rho \in \mathcal{B}(\mathcal{H}) : \rho \geq 0, \mathrm{tr}[\rho] = 1 \}.$$

In the present work, we will only deal with finite-dimensional Hilbert spaces. A state is *pure* if it has rank one. Valid transformations between quantum systems are given in terms of completely positive maps. Let $\mathcal{H}$, $\mathcal{K}$ be two Hilbert spaces and $\mathcal{T} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{K})$ be a linear map. This map is k-positive if the map $T \otimes \mathrm{Id}_k : \mathcal{B}(\mathcal{H}) \otimes \mathcal{M}_k \to \mathcal{B}(\mathcal{K}) \otimes \mathcal{M}_k$ is positive for $k \in \mathbb{N}$. It is completely positive if $\mathcal{T}$ is k-positive for all $k \in \mathbb{N}$. For $\mathcal{T}$ to be a *quantum channel*, we require additionally that the map is trace preserving.

Quantum mechanical measurements are described using effect operators, i.e.

$$\mathrm{Eff}_d := \{ E \in \mathcal{M}_d^{sa} : 0 \leq E \leq I \}.$$

A measurement then corresponds to a *positive operator valued measure* (POVM). Let Σ be the set of measurement outcomes, which we assume to be finite for simplicity. The corresponding POVM is then a set of effects $\{ E_i \}_{i \in \Sigma}$, $E_i \in \mathrm{Eff}_d$ for all $i \in \Sigma$, such that

$$\sum_{i \in \Sigma} E_i = I_d.$$

Since the actual measurement outcomes are not important for us, we will write $\Sigma = [m]$ for some $m \in \mathbb{N}$.

The main concept for the rest of this work is the notion of joint measurability. A collection of POVMs is jointly measurable if they arise as marginals from a joint POVM (see [HMZ16] for an introduction).

Definition 3.18 (Jointly measurable POVMs). *Let $\left\{ E_i^{(j)} \right\}_{i \in [m_j]}$ be a collection of d-dimensional POVMs, where $m_j \in \mathbb{N}$ for all $j \in [g]$, $g \in \mathbb{N}$. The POVMs are* jointly measurable *(often also called* compatible*) if there is a d-dimensional joint POVM $\left\{ R_{i_1,\ldots,i_g} \right\}$ with $i_j \in [m_j]$ such that for all $u \in [g]$ and $v \in [m_u]$,*

$$E_v^{(u)} = \sum_{\substack{i_j \in [m_j] \\ j \in [g] \setminus \{ u \}}} R_{i_1,\ldots,i_{u-1},v,i_{u+1},\ldots i_g}.$$

There is an equivalent definition of joint measurability [HMZ16, Equation 16], formulated in terms of post-processing, which will sometimes be useful. Measurements are compatible if and only if they arise through post-processing from a common measurement.

Lemma 3.19. *Let $E^{(j)} \in (M_d^{sa})^{k_j}$, $j \in [g]$, be a collection of POVMs. These POVMs are jointly measurable if and only if there is some $m \in \mathbb{N}$ and a POVM $M \in (M_d^{sa})^m$ such that*

$$E_i^{(j)} = \sum_{x=1}^{m} p_j(i|x) M_x$$

for all $i \in [k_j]$, $j \in [g]$ and some conditional probabilities $p_j(i|x)$.

Not all measurements in quantum mechanics are compatible, but they can be made compatible if we add enough noise. By adding noise we mean taking the convex combination of a POVM and a trivial measurement, i.e a POVM in which all effects are proportional to the identity. These are called trivial, because they do not depend on the state of the system. With this idea, we can define several compatibility regions, i.e. sets of noise parameters for which any collection of a fixed number of measurements in fixed dimension and with a fixed number outcomes is compatible. For the first such set, we restrict to balanced noise.

Definition 3.20. *Let* $\mathbf{k} \in \mathbb{N}^g$*,* d*,* $g \in \mathbb{N}$*. Then, we call*

$$\Gamma(g, d, \mathbf{k}) := \left\{ s \in [0,1]^g : s_j E^{(j)} + (1 - s_j) I/k_j \text{ compatible } \forall \text{ POVMs } E^{(j)} \in (M_d^{sa})^{k_j} \right\}$$

the balanced compatibility region *for* g *POVMs in* d *dimensions with* k_j *outcomes,* $j \in [g]$*.*

Sometimes it is desirable that the noise is linear in the effect operators. Such noise arises in the framework of quantum steering [UMG14, HKR15].

Definition 3.21. *Let* $\mathbf{k} \in \mathbb{N}^g$*,* d*,* $g \in \mathbb{N}$*. Then, we call*

$$\Gamma^{lin}(g, d, \mathbf{k}) := \left\{ s \in [0,1]^g : s_j E^{(j)} + (1 - s_j) \frac{\operatorname{tr}\left[E^{(j)}\right]}{d} I \text{ compatible } \forall \text{ POVMs } E^{(j)} \in (M_d^{sa})^{k_j} \right\}$$

the linear compatibility region *for* g *POVMs in* d *dimensions with* k_j *outcomes,* $j \in [g]$*.*

Let us prove a lemma which shows that coarse graining, i.e. grouping several outcomes together, does not destroy joint measurability.

Lemma 3.22. *Let* $E^{(j)} \in (M_d^{sa})^{k'_j}$*,* $k'_j \in \mathbb{N}$*,* $j \in [g]$*, be a collection of jointly measurable POVMs. Then, also* $E^{(j)}$*,* $j \in [g] \setminus \{ l \}$ *and* $\tilde{E}^{(l)}$ *are jointly measurable, where*

$$\tilde{E}^{(l)} = (E_1^{(l)}, \ldots, E_{k_l}^{(l)}, E_{k_l+1}^{(l)} + \ldots + E_{k'_l}^{(l)})$$

and $l \in [g]$*,* $k_l \in \mathbb{N}$*,* $k_l \leq k'_l$*.*

Proof. Let $G_{i_1,\ldots,i_g}$, $i_j \in k'_j$, $j \in [g]$ be a joint POVM for the $E^{(j)}$. Then, we can define a new POVM as

$$\tilde{G}_{i_1,\ldots,i_g} = \begin{cases} G_{i_1,\ldots,i_g} & i_l \leq k_l \\ \sum_{j=k_l+1}^{k'_l} G_{i_1,\ldots,i_{l-1},j,i_{l+1},\ldots,i_g} & i_l = k_l + 1 \end{cases}.$$

Note that on the left hand side, $i_j \in k'_j$ for $j \in [g] \setminus \{ l \}$ and $i_l \in [k_l + 1]$. It can easily be verified that this POVM is a joint POVM for the $E^{(j)}$ (with $j \neq l$) and $\tilde{E}^{(l)}$. ☐

Proposition 3.23. *Consider two* g*-tuples of integers* $\mathbf{k}, \mathbf{k}'$ *such that* $\mathbf{k}' \geq \mathbf{k}$ *(coordinatewise, i.e.* $k'_i \geq k_i$*,* $\forall i \in [g]$*). Let* $\# \in \{ \emptyset, lin \}$*. Then,*

$$\Gamma^{\#}(g, d, \mathbf{k}') \subseteq \Gamma^{\#}(g, d, \mathbf{k}).$$

Proof. Fix $s \in \Gamma(g, d, \mathbf{k}')$. Let furthermore $E^{(j)} \in (M_d^{sa})^{k_j}$, $j \in [g]$ be a collection of POVMs. Let $\tilde{E}^{(j)} \in (M_d^{sa})^{k'_j}$ be the POVM which is equal to $E^{(j)}$ in the first k_j entries and 0 for the rest. Then, the $s_j \tilde{E}^{(j)} + (1 - s_j) I/k'_j$ are jointly measurable. Let

$$F^{(j)} = \left(s_j \tilde{E}_1^{(j)} + (1 - s_j) I/k'_j, \ldots, s_j \tilde{E}_{k_j}^{(j)} + (1 - s_j) I/k'_j, (1 - s_j) \frac{k'_j - k_j}{k'_j} I \right).$$

An iterative application of Lemma 3.22 shows that also the $F^{(j)}$ are jointly measurable. Let $\mathbf{i} = (i_1, \ldots, i_g) \in [k_1 + 1] \times \ldots \times [k_g + 1]$. Define

$$p_j(l|\mathbf{i}) = \begin{cases} 1 & i_j = l \text{ and } l \leq k_j \\ \frac{1}{k_j} & i_j = k_j + 1 \text{ and } l \leq k_j \\ 0 & l > k_j \end{cases}.$$

These are conditional probabilities and it holds that

$$\sum_{\mathbf{i}\in\times_{j=1}^{g}[k_j+1]} p_j(l|\mathbf{i})G_{\mathbf{i}} = \sum_{\substack{\mathbf{i}\in\times_{j=1}^{g}[k_j+1]\\ i_j=l}} G_{\mathbf{i}} + \frac{1}{k_j}\sum_{\substack{\mathbf{i}\in\times_{j=1}^{g}[k_j+1]\\ i_j=k_j+1}} G_{\mathbf{i}}$$

$$= s_j E_l^{(j)} + (1-s_j)\frac{I}{k_j'} + \frac{1}{k_j}\frac{k_j'-k_j}{k_j'}(1-s_j)I$$

$$= s_j E_l^{(j)} + (1-s_j)\frac{I}{k_j}.$$

From Lemma 3.19, it follows that $s \in \Gamma(g,d,\mathbf{k})$. The assertion for Γ^{lin} follows directly from extending the POVMs by zeroes. □

The following proposition generalizes [BN18, Proposition III.4(6)].

Proposition 3.24. *Let $\mathbf{k} \in \mathbb{N}^g$. Furthermore, let $k_{max} = \max_{j\in[g]} k_j$. Then,*

$$\Gamma^{lin}(g, k_{max}d, k_{max}^{\times g}) \subseteq \Gamma(g,d,\mathbf{k}).$$

Proof. From Proposition 3.23, it follows that

$$\Gamma(g,d,k_{max}^{\times g}) \subseteq \Gamma(g,d,\mathbf{k}),$$

so it is enough to prove

$$\Gamma^{lin}(g, k_{max}d, k_{max}^{\times g}) \subseteq \Gamma(g,d,k_{max}^{\times g}).$$

Pick g POVMs $E^{(j)}$ of dimension d and with k_{max} outcomes each, $j \in [g]$. Let

$$F_i^{(j)} = E_i^{(j)} \oplus E_{i+1}^{(j)} \oplus \ldots \oplus E_{i+(k_{max}-1)}^{(j)} \qquad \forall i \in [k_{max}], \forall j \in [g].$$

Above, we are considering the addition operation modulo k_{max}. Clearly, $F_i^{(j)} \geq 0$ and $\sum_{i=1}^{k_{max}} F_i^{(j)} = I_{dk_{max}}$ for any $j \in [g]$, so the $F^{(j)}$ again are POVMs. Let $s \in \Gamma^{lin}(g, k_{max}d, k_{max}^{\times g})$. Then, the $s_jF^{(j)} + (1-s_j)I_{k_{max}d}/k_{max}$ are jointly measurable POVMs, because $\operatorname{tr}\left[F_i^{(j)}\right]/(k_{max}d) = 1/k_{max}$. Applying an isometry onto the first block of the direct sum ascertains that the $s_jE^{(j)} + (1-s_j)I_d/k_{max}$ are jointly measurable as well. Since the POVMs we picked were arbitrary, the assertion follows. □

4. The matrix jewel

In the following, we identify the subalgebra of $d \times d$ diagonal matrices with $\mathbb{C}^d$.

Definition 4.1 (Matrix jewel)**.** *Consider the vectors $v_1^{(k)}, \ldots, v_{k-1}^{(k)} \in \mathbb{C}^k$ defined as*

$$v_j^{(k)}(\varepsilon) := -\frac{2}{k} + 2\delta_{\varepsilon,j}, \qquad \forall j \in [k-1], \forall \varepsilon \in [k].$$

The free spectrahedron defined by

$$\mathcal{D}_{\diamondsuit,k}(n) := \left\{ X \in (\mathcal{M}_n^{sa})^{k-1} : \sum_{j=1}^{k-1} v_j^{(k)} \otimes X_j \leq I_{kn} \right\}.$$

is called the matrix jewel base. *For a g-tuple of non-negative integers $\mathbf{k} = (k_1, \ldots, k_g)$, we define the* matrix jewel *$\mathcal{D}_{\diamondsuit,\mathbf{k}}$ to be the free spectrahedron*

$$\mathcal{D}_{\diamondsuit,\mathbf{k}} := \mathcal{D}_{\diamondsuit,k_1} \hat{\oplus} \mathcal{D}_{\diamondsuit,k_2} \hat{\oplus} \cdots \hat{\oplus} \mathcal{D}_{\diamondsuit,k_g},$$

where the direct sum operation $\hat{\oplus}$ *for free spectrahedra was introduced in Section 3.3. In other words, we have*

(5)

$$\mathcal{D}_{\diamondsuit,\mathbf{k}}(n) = \left\{ X \in (\mathcal{M}_n^{sa})^{\sum_{i=1}^{g}(k_i-1)} : \sum_{i=1}^{g} \sum_{j=1}^{k_i-1} \left[I^{\otimes(i-1)} \otimes v_j^{(k_i)} \otimes I^{\otimes(g-i)} \right] \otimes X_{i,j} \leq I_{(\prod_{s=1}^{g} k_i)n} \right\}.$$

Remark 4.2. *The matrix jewel is the maximal matrix convex set (in the sense of* [DDOSS17, Section 4]*, see also Equation* (3)*) built on top of the direct sum of simplices*

$$\mathcal{D}_{\diamondsuit,k_1}(1) \oplus \mathcal{D}_{\diamondsuit,k_2}(1) \oplus \cdots \oplus \mathcal{D}_{\diamondsuit,k_g}(1).$$

At level one, the matrix jewel base is isomorphic to a simplex, for which we can identify the extremal points.

Lemma 4.3. *The extremal points of the jewel base* $\mathcal{D}_{\diamondsuit,k}(1) \subseteq \mathbb{R}^{k-1}$ *are*

$$x_i^{(k)} := -k/2e_i, \qquad \textit{for } i \in [k-1]$$
$$x_k^{(k)} := k/2(\underbrace{1,\ldots,1}_{k-1 \textit{ times}}),$$

where e_i *are the elements of the standard orthonormal basis in* $\mathbb{R}^{k-1}$*.*

Proof. Since $\mathcal{D}_{\diamondsuit,k}(1)$ is a polyhedron and since the hyperplanes $(v_1(\varepsilon),\ldots v_{k-1}(\varepsilon))_{\varepsilon=1}^{k}$ are such that each $k-1$ of them linearly span $\mathbb{R}^{k-1}$, [Bar02, Theorem II.4.2] implies that it is enough to check whether each point as above fulfills $k-1$ of the above constraints with equality (there is no point which fulfills all constraints with equality). We verify for fixed $\varepsilon \in [k]$:

$$\sum_{j=1}^{k-1} v_j(\varepsilon)(-k/2e_i)_j = 1 - k\delta_{\varepsilon,i}, \qquad i \in [k-1],$$

and

$$\sum_{j=1}^{k-1} v_j(\varepsilon)k/2(1,\ldots,1)_j = 1 - k\delta_{\varepsilon,k},$$

which proves the claim. □

At level 1, the matrix jewel base is, for $k=2$, the segment $[-1,1] \subseteq \mathbb{R}$. We display in Figure 1 the sets $\mathcal{D}_{\diamondsuit,k}(1)$, for $k=3,4$. The notion of matrix jewel generalizes the matrix diamond introduced

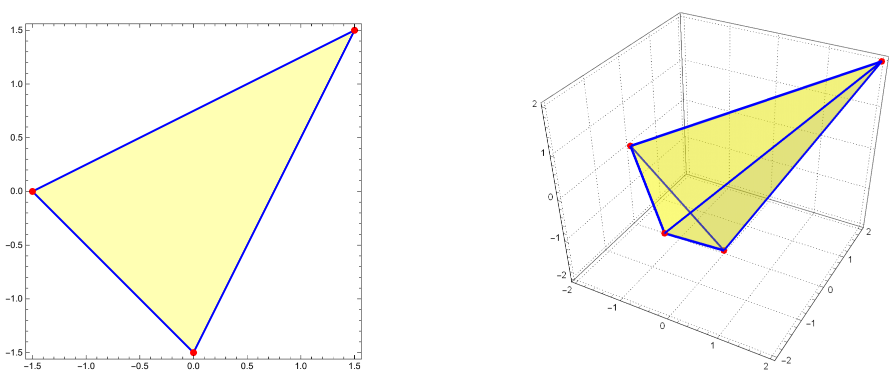

FIGURE 1. The spectrahedron level of the matrix jewel base $\mathcal{D}_{\diamondsuit,k}(1)$, for $k=3,4$.

in [DDOSS17]; indeed, with the notation of [BN18], the matrix diamond of size g is given by

$$\mathcal{D}_{\diamond,g} = \mathcal{D}_{\diamondsuit,(\underbrace{2,\ldots,2}_{g \text{ times}})} = \widehat{\bigoplus}_{i=1}^{g} \mathcal{D}_{\diamondsuit,2}.$$

In Figure 2, we print the first level of the matrix jewel, for vectors $\mathbf{k}$ equal to, respectively, $(2,2)$, $(2,2,2)$, and $(2,3)$.

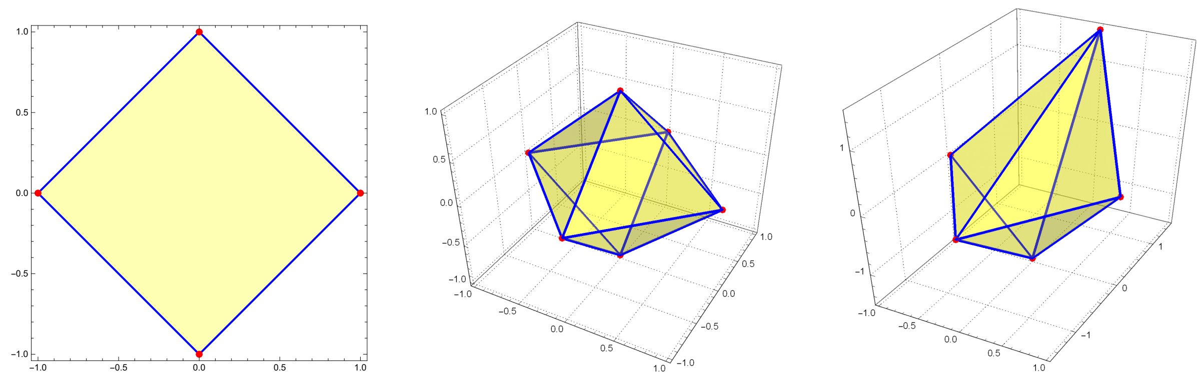

FIGURE 2. The spectrahedron level of the matrix jewels $\mathcal{D}_{\diamondsuit,(2,2)}(1)$, $\mathcal{D}_{\diamondsuit,(2,2,2)}(1)$, and $\mathcal{D}_{\diamondsuit,(2,3)}(1)$. The first two are in fact the matrix diamonds $\mathcal{D}_{\diamond,2}(1)$ and $\mathcal{D}_{\diamond,3}(1)$ from [BN18] (a square and an octahedron), while the last polyhedron is new.

5. THE MATRIX JEWEL AND JOINT MEASURABILITY OF POVMS

In this section, we establish an equivalence between the inclusion of the matrix jewel in a spectrahedron defined by a tuple of POVMs and the joint measurability of the POVMs. The inclusion at different levels will correspond to different notions of joint measurability. Our first result relates the inclusion of the matrix jewel base, at level 1, to the definition of a POVM.

Proposition 5.1. *Let $E \in (\mathcal{M}_d^{sa})^{k-1}$. Then, $\{ E_1, \ldots, E_{k-1}, I - E_1 - \ldots - E_{k-1} \}$ is a POVM if and only if*

$$\mathcal{D}_{\diamondsuit,k}(1) \subseteq \mathcal{D}_{2E-\frac{2}{k}I}(1).$$

Proof. Since the left hand side is a polytope, we only need to check the assertion on the extremal points $x_j^{(k)}$ from Lemma 4.3. We have

$$-\frac{k}{2}e_i \in \mathcal{D}_{2E-\frac{2}{k}I}(1) \iff -\frac{k}{2}(2E_i - \frac{2}{k}I) \leq I \iff E_i \geq 0$$

and

$$\frac{k}{2}(1,\ldots,1) \in \mathcal{D}_{2E-\frac{2}{k}I}(1) \iff \frac{k}{2}\sum_{i=1}^{k-1}(2E_i - \frac{2}{k}I) \leq I \iff \sum_{i=1}^{k-1} E_i \leq I.$$

This proves the assertion. □

The following theorem is one of our main results, connecting joint measurability of arbitrary POVMs to the inclusion of the matrix jewel. It is a generalization of [BN18, Theorem V.3] from the case of g binary (i.e. 2-outcome) POVMs to general POVMs (with an arbitrary number of outcomes).

Theorem 5.2. *For a fixed matrix dimension d, consider g tuples of self-adjoint matrices $E^{(i)} \in (\mathcal{M}_d^{sa})^{k_i-1}$, $k_i \in \mathbb{N}$, $i \in [g]$. Define $E_{k_i}^{(i)} := I_d - E_1^{(i)} \ldots - E_{k_i-1}^{(i)}$, set $\mathbf{k} = (k_1, \ldots, k_g)$, and write*

$$\begin{aligned}\mathcal{D}_E &:= \mathcal{D}_{(2E^{(1)}-\frac{2}{k_1}I,\ldots,2E^{(g)}-\frac{2}{k_g}I)} \\ &= \bigsqcup_{n=1}^{\infty}\left\{X \in (\mathcal{M}_n^{sa})^{\sum_{i=1}^g (k_i-1)} \,:\, \sum_{i=1}^{g}\sum_{j=1}^{k_i-1}\left(2E_j^{(i)} - \frac{2}{k_i}I\right)\otimes X_{i,j} \leq I_{dn}\right\}.\end{aligned}$$

Then

(1) $\mathcal{D}_{\diamondsuit,\mathbf{k}}(1) \subseteq \mathcal{D}_E(1)$ *if and only if* $\left\{ E_1^{(i)}, \ldots, E_{k_i}^{(i)} \right\}$, $i \in [g]$ *are POVMs.*

(2) $\mathcal{D}_{\diamondsuit,\mathbf{k}} \subseteq \mathcal{D}_E$ *if and only if* $\left\{ E_1^{(i)}, \ldots, E_{k_i}^{(i)} \right\}$, $i \in [g]$ *are jointly measurable POVMs.*

(3) $\mathcal{D}_{\diamondsuit,\mathbf{k}}(l) \subseteq \mathcal{D}_E(l)$ *for* $l \in [d]$ *if and only if for any isometry* $V : \mathbb{C}^l \hookrightarrow \mathbb{C}^d$, *the tuples* $\left\{ V^*E_1^{(i)}V, \ldots, V^*E_{k_i}^{(i)}V \right\}$, $i \in [g]$ *are jointly measurable POVMs.*

Proof. Since $\mathcal{D}_{\diamondsuit,k_i}(1)$ is a polytope for all $i \in [g]$ and $D_{\diamondsuit,k_i} = \mathcal{W}_{max}(D_{\diamondsuit,k_i}(1))$, the first assertion follows from Lemmas 3.15 and 3.17 together with Proposition 5.1.

For the second assertion, let us define, for $i \in [g]$ and $j \in [k_i - 1]$,

$$w_j^{(i)} := \underbrace{I \otimes \cdots \otimes I}_{i-1 \text{ times}} \otimes v_j^{(k_i)} \otimes \underbrace{I \otimes \cdots \otimes I}_{g-i \text{ times}}.$$

Here, the $v_j^{(k_i)}$ are (identified with) the diagonal matrices appearing in Definition 4.1, with the appropriate matrix dimension (k_i in the formula above). The free spectrahedral inclusion holds if and only if the unital map $\Phi : \mathcal{OS}_{\left\{ w_j^{(i)} \right\}_{i\in[g],j\in[k_i-1]}} \to \mathcal{M}_d$, defined as

$$\Phi : w_j^{(i)} \mapsto 2E_j^{(i)} - \frac{2}{k_i}I \qquad \forall i \in [g], \forall j \in [k_i - 1],$$

is completely positive, since $\mathcal{D}_{\diamondsuit,\mathbf{k}}(1)$ is a polytope and therefore bounded. By Arveson's extension theorem Φ is completely positive if and only if there is a completely positive extension $\tilde{\Phi} : \mathbb{C}^{k_1\cdots k_g} \to \mathcal{M}_d$ of Φ. As $\mathbb{C}^{k_1\cdots k_g}$ is a commutative matrix subalgebra, $\tilde{\Phi}$ is completely positive if and only if it is positive. Thus, it suffices to check the existence of a positive extension $\tilde{\Phi}$, which we will do now. Let $\varepsilon \in [\mathbf{k}] := \times_{i=1}^{g}[k_i]$. Then,

$$w_j^{(i)}(\varepsilon) = -\frac{2}{k_i} + 2\delta_{\varepsilon(i),j}. \tag{6}$$

Let $g_\eta \in \mathbb{C}^{k_1\cdots k_g}$, $\eta \in [\mathbf{k}]$ such that $g_\eta(\varepsilon) = \delta_{\varepsilon,\eta}$. These vectors form a basis of $\mathbb{C}^{k_1\cdots k_g}$. Hence, we can rewrite Equation (6) as

$$w_j^{(i)}(\varepsilon) = -\frac{2}{k_i}\sum_{\eta\in[\mathbf{k}]} g_\eta(\varepsilon) + 2\sum_{\substack{\eta\in[\mathbf{k}]\\ \eta(i)=j}} g_\eta(\varepsilon).$$

Let $G_\eta := \tilde{\Phi}(g_\eta)$. The map $\tilde{\Phi}$ is positive if and only if $G_\eta \geq 0$ for all $\eta \in [\mathbf{k}]$. By the definition of $\tilde{\Phi}$ and its unitality, we obtain

$$-\frac{2}{k_i}I + 2\sum_{\substack{\eta\in[\mathbf{k}]\\ \eta(i)=j}} G_\eta = 2E_j^{(i)} - \frac{2}{k_i}I \qquad \forall i \in [g], \forall j \in [k_i - 1].$$

Thus, there exist a positive extension $\tilde{\Phi}$ if and only if there exist $\{ G_\eta \}_{\eta \in [\mathbf{k}]}$ such that

$$\begin{aligned} G_\eta &\geq 0 && \forall \eta \in [\mathbf{k}] \\ I &= \sum_{\eta \in [\mathbf{k}]} G_\eta \\ E_j^{(i)} &= \sum_{\substack{\eta \in [\mathbf{k}] \\ \eta(i)=j}} G_\eta && \forall i \in [g], \forall j \in [k_i - 1]. \end{aligned}$$

This is equivalent to the G_η being a joint POVM for the $\left\{ E_1^{(i)} \dots E_{k_i}^{(i)} \right\}$, since the above conditions also imply

$$E_{k_i}^{(i)} = I - \sum_{j=1}^{k_i-1} E_j^{(i)} = \sum_{\eta \in [\mathbf{k}]} G_\eta - \sum_{j=1}^{k_i-1} \sum_{\substack{\eta \in [\mathbf{k}] \\ \eta(i)=j}} G_\eta = \sum_{\substack{\eta \in [\mathbf{k}] \\ \eta(i)=k_i}} G_\eta \qquad \forall i \in [g].$$

Finally, the third claim follows from the second one, using the standard argument in [BN18, Lemma V.2 and Corollary IV.6]. □

The correspondence in the theorem above also extends to the level of balanced compatibility regions / inclusion sets. The theorem below corresponds to [BN18, Theorem V.7] and is a generalization of the latter from binary POVMs to POVMs with an arbitrary number of outcomes.

Theorem 5.3. *Let d, $g \in \mathbb{N}$ and $\mathbf{k} \in \mathbb{N}^g$. Then,*

$$\Gamma(g, d, \mathbf{k}) = \Delta(g, d, \mathbf{k}).$$

Proof. Let $s \in \mathbb{R}_+^g$. It holds that $s \in \Gamma(g, d, \mathbf{k})$ if and only if $s_j E^{(j)} + (1 - s_j) I_d / k_j$, $j \in [g]$ are jointly measurable for any d-dimensional POVMs with k_j outcomes for the j-th POVM. Let $\mathcal{D}_E$ be as in Theorem 5.2. We find that

$$\begin{aligned} (s_1^{\times(k_1-1)}, \dots, s_g^{\times(k_g-1)}) \cdot \mathcal{D}_{◆,\mathbf{k}} \subseteq \mathcal{D}_E &\iff \mathcal{D}_{◆,\mathbf{k}} \subseteq \mathcal{D}_{(2s_1 E^{(1)} - \frac{2s_1}{k_1} I, \dots, 2s_g E^{(g)} - \frac{2s_g}{k_g} I)} \\ &\iff \mathcal{D}_{◆,\mathbf{k}} \subseteq \mathcal{D}_{(2F^{(1)} - \frac{2}{k_1} I, \dots, 2F^{(g)} - \frac{2}{k_g} I)}, \end{aligned}$$

where $F_i^{(j)} = s_j E_i^{(j)} + (1 - s_j) \frac{1}{k_j} I_d$ and $i \in [k_j - 1]$, $j \in [g]$. Hence, it follows from Theorem 5.2 that $s \in \Gamma(g, d, \mathbf{k})$ if and only if the implication

$$\mathcal{D}_{◆,\mathbf{k}}(1) \subseteq \mathcal{D}_E(1) \implies \left(s_1^{\times(k_1-1)}, \dots, s_g^{\times(k_g-1)}\right) \cdot \mathcal{D}_{◆,\mathbf{k}} \subseteq \mathcal{D}_E$$

is true for all $E = (2E^{(1)} - \frac{2}{k_1} I, \dots, 2E^{(g)} - \frac{2}{k_g} I)$. Moreover, $A \mapsto 2A - (2/k)I$ is a bijective map on $\mathcal{M}_d^{sa}$ for fixed $k \in \mathbb{N}$. Thus, $s \in \Gamma(g, d, \mathbf{k})$ is equivalent to $s \in \Delta(g, d, \mathbf{k})$. □

6. Compatibility results from quantum information theory

Having established in the previous section the close relation between the compatibility and inclusion sets, we gather next results from quantum information theory which provide upper and lower bounds on the sets Γ. Such bounds translate immediately, via Theorem 5.3, to the corresponding bounds for the sets Δ; we postpone this analysis until Section 10.

6.1. **Upper bounds from MUBs.** Mutually unbiased bases (MUBs) yield natural examples of POVMs which are very far from being compatible [WF89]. Recall that, in $\mathbb{C}^d$, a collection of g orthonormal bases $\left\{ \psi_i^{(j)} \right\}_{i=1}^{d}$, $j \in [g]$, is called *mutually unbiased* if

$$|\langle \psi_i^{(j)}, \psi_u^{(v)} \rangle|^2 = \frac{1}{d}$$

for $j \neq v$ and any i, $u \in [d]$. Let $E_i^{(j)} = \psi_i^{(j)} (\psi_i^{(j)})^*$ be the corresponding effect operators. In the case where we construct one MUB from another one by applying a Fourier transform, i.e.

$$\psi_k^{(2)} = \frac{1}{\sqrt{d}} \sum_{l=1}^{d} e^{2\pi \mathrm{i} \frac{lk}{d}} \psi_l^{(1)}$$

we will call these two MUBs *canonically conjugated*.

The maximal number of MUBs in dimension d is $d+1$ and it is known that this bound is attained if $d = p^r$ for a prime number p and $r \in \mathbb{N}$ [WF89]. Apart from that, very few examples are known, see [DEBŻ10] for a review. From [CHT12], we have the following results on two canonically conjugated MUBs:

Proposition 6.1 ([CHT12, Proposition 5, Example 1 and Proposition 6])**.** *Let $E^{(1)}$ and $E^{(2)}$ be the effect operators corresponding to two canonically conjugated MUBs. Then, $\lambda E^{(1)} + (1-\lambda) I/d$ and $\mu E^{(2)} + (1-\mu) I/d$ are jointly measurable if and only if*

$$\mu \leq \frac{1}{d}[(d-2)(1-\lambda) + 2\sqrt{(1-d)\lambda^2 + (d-2)\lambda + 1}]$$

(equivalently, we can exchange λ and μ). Another equivalent form is that the above POVMs are jointly measurable if and only if

$$\mu + \lambda \leq 1 \qquad \text{or} \qquad \lambda^2 + \mu^2 + \frac{2(d-2)}{d}(1-\mu)(1-\lambda) \leq 1.$$

In particular, for $\mu = \lambda$, this simplifies to

$$\lambda \leq \frac{1}{2}\left(1 + \frac{1}{1+\sqrt{d}}\right).$$

For more than two MUBs, there is a necessary criterion which generalizes the above in the symmetric case.

Proposition 6.2 ([DSFB18, Equation 10])**.** *Let $\lambda E^{(j)} + (1-\lambda) I/d$, $j \in [g]$ be jointly measurable. Then, it holds that*

$$\lambda \leq \frac{\sqrt{d} + g}{g(\sqrt{d}+1)}.$$

There is a different approach to finding necessary conditions for joint measurability developed by H. Zhu. While it is not restricted to MUBs, in seems to work best for these objects. We recall Zhu's incompatibility criterion [Zhu15, ZHC16]. Define, for any matrix A with $\mathrm{tr}[A] \neq 0$,

$$\overline{\mathcal{G}}(A) := \frac{|A^\circ\rangle\langle A^\circ|}{\mathrm{tr}[A]} \in \mathcal{M}_{d^2}^{sa},$$

where $A^\circ = A - \mathrm{tr}[A] I/d$ and $|A^\circ\rangle$ is a vectorization of A°. For a POVM E, we define

$$\overline{\mathcal{G}}(\{ E_i \}_{i \in [k]}) = \sum_{i=1}^{k} \overline{\mathcal{G}}(E_i).$$

Proposition 6.3 ([Zhu15, Equations (10,11)]). *Let $E^{(j)}$, $j \in [g]$ be a collection of compatible POVMs in $\mathcal{M}_d$. Then,*

$$1 + \min\left\{ \operatorname{tr}[H] : H \geq \overline{\mathcal{G}}(E^{(j)}), \forall j \in [g] \right\} \leq d.$$

If we are interested in the case of g MUBs, we obtain the following necessary criterion, which appears in [ZHC16]. We will provide a proof for convenience.

Proposition 6.4. *Let $\left\{ \psi_i^{(j)} \right\}_{i=1}^{d}$, $j \in [g]$ be a collection of MUBs with corresponding POVMs $E^{(j)}$. If $\lambda_j E^{(j)} + (1-\lambda_j)I/d$ are compatible for $\lambda_j \in [0,1]$, then*

$$\sum_{j=1}^{g} \lambda_j^2 \leq 1.$$

Proof. A straightforward calculation shows that $\operatorname{tr}\left[(E_i^{(j)})^\circ (E_u^{(v)})^\circ\right] = 0$ if and only if

$$\operatorname{tr}\left[E_i^{(j)} E_u^{(v)}\right] = \frac{\operatorname{tr}\left[E_i^{(j)}\right] \operatorname{tr}\left[E_u^{(v)}\right]}{d}.$$

The latter condition is fulfilled by the MUBs for $j \neq v$. Hence, the $\overline{\mathcal{G}}(E^{(j)})$ are pairwise orthogonal and the same holds for $\overline{\mathcal{G}}(\tilde{E}^{(j)})$, where $\tilde{E}_i^{(j)} = \lambda_j E_i^{(j)} + (1-\lambda_j)I/d$, $\lambda_j \in [0,1]$. This implies that

$$H \geq \sum_{j=1}^{g} \overline{\mathcal{G}}(\tilde{E}^{(j)})$$

in Proposition 6.3. Therefore,

$$\begin{aligned} d - 1 &\geq \sum_{j=1}^{g} \sum_{i=1}^{d} \operatorname{tr}\left[\overline{\mathcal{G}}(\tilde{E}_i^{(j)})\right] \\ &= \sum_{j=1}^{g} \sum_{i=1}^{d} \lambda_j^2 \operatorname{tr}\left[\overline{\mathcal{G}}(E_i^{(j)})\right] \\ &= d\frac{d-1}{d} \sum_{j=1}^{g} \lambda_j^2. \end{aligned}$$

This proves the claim. □

6.2. **Lower bounds from cloning.** In this section, we will review some results on asymmetric cloning, which will then translate into lower bounds on the inclusion sets for the matrix jewel. See [BN18, Section VI] for a more detailed discussion. Let us define the set of allowed parameters arising from cloning:

$$\begin{aligned} \Gamma^{clone}(g,d) := \Big\{ & s \in [0,1]^g : \exists \mathcal{T} : \mathcal{M}_d^{\otimes g} \to \mathcal{M}_d \text{ unital and completely positive linear map s.t.} \\ & \forall X \in \mathcal{M}_d, \forall i \in [g], \mathcal{T}\left(I^{\otimes(i-1)} \otimes X \otimes I^{\otimes(n-i)}\right) = s_i X + (1-s_i)\frac{\operatorname{tr}[X]}{d} I \Big\}. \end{aligned} \tag{7}$$

A cloning map $\mathcal{C}$ is a quantum channel from $\mathcal{M}_d$ to $\mathcal{M}_d^{\otimes g}$ which maps all pure states σ as close as possible to $\sigma^{\otimes g}$. Often, the worst case single copy fidelity F_i is used to quantify the error with respect to a perfect cloning device (which is impossible to implement). Here,

$$F_i(\mathcal{C}) := \inf_{\psi \in \mathcal{S}(\mathcal{H}) \text{ pure}} \operatorname{tr}\left[\mathcal{C}(\psi) I^{\otimes(i-1)} \otimes \psi \otimes I^{\otimes(g-i)}\right], \qquad \forall i \in [g].$$

The following proposition clarifies the connection between asymmetric cloning and our definition of $\Gamma^{clone}(g,d)$, by showing that, without any loss in single copy fidelities, any cloning map can be assumed to have depolarizing marginals. It uses ideas which can be found in [Wer98] (see also [Has17]).

Proposition 6.5. *Let $\mathcal{C} : \mathcal{M}_d \to \mathcal{M}_d^{\otimes g}$ be a quantum channel with $F_i(\mathcal{C}) = \eta_i$ $\forall i \in [g]$. Then, there is a channel $\tilde{\mathbb{C}} : \mathcal{M}_d \to \mathcal{M}_d^{\otimes g}$ such that*

$$\operatorname{tr}\left[\tilde{\mathcal{C}}(\psi) I^{\otimes(i-1)} \otimes \psi \otimes I^{\otimes(n-i)}\right] = \nu_i \geq \eta_i \qquad \forall \psi \in \mathcal{S}(\mathcal{C}^d) \text{ pure}, \forall i \in [g].$$

Moreover, $\tilde{\mathcal{C}}$ can be chosen such that

$$\tilde{\mathcal{C}}_i(A) = \operatorname{tr}_{i^c}[\tilde{\mathcal{C}}(A)] = \lambda_i A + (1-\lambda_i)\frac{\operatorname{tr}[A]}{d} I_d \qquad \forall A \in \mathcal{M}_d.$$

Here, $\lambda_i = (d\nu_i - 1)/(d-1) \in [0,1]$ $\forall i \in [g]$ and $\operatorname{tr}_{i^c}[\cdot]$ denotes the partial trace over all systems but the i-th one.

Proof. We claim that we can choose $\tilde{C}$ as a symmetrized version of $\mathcal{C}$, i.e.

$$\tilde{\mathcal{C}}(A) = \int_{\mathcal{U}(d)} (U^{\otimes g}) \mathcal{C}(U^* A U)(U^{\otimes g})^* d\mu(U) \qquad A \in \mathcal{M}_d.$$

Here, μ is the normalized Haar measure on the unitary group. The marginals of this map are

$$\begin{aligned}\tilde{\mathcal{C}}_i(A) &= \int_{\mathcal{U}(d)} \operatorname{tr}_{i^c}[(U^{\otimes g})\mathcal{C}(U^*AU)(U^{\otimes g})^*] d\mu(U) \\ &= \int_{\mathcal{U}(d)} U \mathcal{C}_i(U^* A U) U^* d\mu(U) \qquad \forall A \in \mathcal{M}_d,\end{aligned}$$

where we have written $\mathcal{C}_i(A) := \operatorname{tr}_{i^c}[\mathcal{C}(A)]$. We observe furthermore that for any $V \in \mathcal{U}(d)$ and $A \in \mathcal{M}_d$,

$$\begin{aligned}V\tilde{\mathcal{C}}_i(A)V^* &= V \int_{\mathcal{U}(d)} U\mathcal{C}_i(U^*AU)U^* d\mu(U) V^* \\ &= \int_{\mathcal{U}(d)} W \mathcal{C}_i(W^* V A V^* W) W^* d\mu(W) \\ &= \tilde{\mathcal{C}}_i(VAV^*),\end{aligned}$$

where we have used left-invariance of the Haar measure in the second line. Thus,

$$V\tilde{\mathcal{C}}_i(\cdot)V^* = \tilde{\mathcal{C}}_i(V \cdot V^*). \tag{8}$$

Let us compute the single copy fidelities.

$$\begin{aligned}\operatorname{tr}\left[\tilde{\mathcal{C}}(\psi) I^{\otimes(i-1)} \otimes \psi \otimes I^{\otimes(n-i)}\right] &= \int_{\mathcal{U}(d)} \operatorname{tr}\left[\mathcal{C}(U^*\psi U) I^{\otimes(i-1)} \otimes U^*\psi U \otimes I^{\otimes(n-i)}\right] d\mu(U) \\ &\geq \int_{\mathcal{U}(d)} \eta_i d\mu(U) = \eta_i.\end{aligned}$$

Here, we have used that $U^*\psi U$ is a pure state and that the Haar measure is positive. This shows the first assertion. Let us now prove the third assertion. Let τ_{0i} be the Choi matrix of $\tilde{\mathcal{C}}_i$, i.e. $\tau_{0i} = (\mathrm{Id}_d \otimes \tilde{\mathcal{C}}_i)(\Omega)$, where Ω is the maximally entangled state

$$\Omega := \frac{1}{d}\sum_{i,j=1}^{d}(e_i \otimes e_i)(e_j \otimes e_j)^*$$

and $\{\, e_i \,\}_{i=1}^d$ is an orthonormal basis of $\mathbb{C}^d$. Let $V \in \mathcal{U}(d)$. Then,

$$\begin{aligned}(\overline{V} \otimes V)\tau_{0i}(\overline{V} \otimes V)^* &= (\mathrm{Id}_d \otimes \tilde{\mathcal{C}}_i)((\overline{V} \otimes V)\Omega(\overline{V} \otimes V)^*)\\ &= (\mathrm{Id}_d \otimes \tilde{\mathcal{C}}_i)(\Omega) = \tau_{0i}\end{aligned}$$

where we have used Equation (8) and the well-known trick $(A \otimes I_d)\Omega = (I_d \otimes A^T)\Omega$ for any $A \in \mathcal{M}_d$. The above invariance implies that τ_{0i} is an isotropic state and is therefore of the form [Key02, Section 3.1.3]

$$\tau_{0i} = (1-\lambda_i)\frac{1}{d^2}I_{d^2} + \lambda_i\Omega.$$

By the Choi-Jamiołkowski isomorphism, this is equivalent to

$$\tilde{\mathcal{C}}_i(A) = \lambda_i A + (1-\lambda_i)\frac{\mathrm{tr}[A]}{d}I_d \qquad \forall A \in \mathcal{M}_d.$$

With this expression, we can explicitly compute the single copy fidelities

$$\mathrm{tr}\left[\tilde{\mathcal{C}}(\psi)I^{\otimes(i-1)} \otimes \psi \otimes I^{\otimes(n-i)}\right] = \mathrm{tr}\left[\tilde{\mathcal{C}}_i(\psi)\psi\right] = \lambda_i + \frac{1-\lambda_i}{d}.$$

This proves the second assertion as well as the expression for λ_i in terms of ν_i. □

Therefore, we can now use $\mathcal{T} = \tilde{\mathcal{C}}^*$ in Equation (7), which shows that $\Gamma^{clone}(g,d)$ indeed arises from optimal asymmetric cloning. The exact form of $\Gamma^{clone}(g,d)$ has been computed in [Kay16, ŚCHM14], using different methods. To obtain the theorem below from [Kay16], one needs to perform the necessary transform from ν_i to λ_i.

Theorem 6.6 ([Kay16, Theorem 1, Section 2.3])**.** *For any* $g,d \geq 2$

$$\begin{aligned}\Gamma^{clone}(g,d) = \Bigg\{ s \in [0,1]^g : (g+d-1)\left[g - d^2 + d + (d^2-1)\sum_{i=1}^{g} s_i\right] \\ \leq \left(\sum_{i=1}^{g}\sqrt{s_i(d^2-1)+1}\right)^2 \Bigg\}.\end{aligned}$$

In particular, for $s_1 = \ldots = s_g$*, the maximal value is*

$$s_{max} = \frac{g+d}{g(d+1)}.$$

In the symmetric case, the optimal cloning map is unique [Wer98, Key02]. The following proposition shows that cloning gives indeed a lower bound on the balanced compatibility region.

Proposition 6.7. *Let* $g,\ d \in \mathbb{N}$ *and* $\mathbf{k} \in \mathbb{N}^g$ *and* $k_{max} = \max_{j\in[g]} k_j$*. Then, it holds that*

$$\Gamma^{clone}(g, k_{max}d) \subseteq \Gamma(g,d,\mathbf{k}).$$

Proof. Using

$$G_{i_1,\ldots,i_g} = \mathcal{T}\left(E_{i_1}^{(1)} \otimes \ldots \otimes E_{i_g}^{(g)}\right)$$

as a joint POVM, where $\mathcal{T}$ is the map from Equation (7), it is clear that

$$\Gamma^{clone}(g, D) \subseteq \Gamma^{lin}(g, D, \mathbf{k}), \qquad \forall \mathbf{k} \in \mathbb{N}^g, \forall D \in \mathbb{N}. \tag{9}$$

The assertion follows then by Proposition 3.24. □

Remark 6.8. *Note that the left hand side of Equation (9) is independent of* $\mathbf{k}$*, since the cloning map is designed to clone states, not measurements, such that we can perform any kind of measurement on the approximate clones.*

7. Lower bounds from symmetrization

In this section, we give lower bounds on $\Delta(g, d, \mathbf{k})$ by considering its inclusion inside more symmetric spectrahedra. We start with a single point.

Theorem 7.1. *Let* g*,* $d \in \mathbb{N}$*,* $k_j \in \mathbb{N}$*,* $\mathbf{k} = (k_1, \ldots, k_g)$*. Then,*

$$\frac{1}{2d}\left(\frac{1}{k_1 - 1}, \ldots, \frac{1}{k_g - 1}\right) \in \Delta(g, d, \mathbf{k}).$$

Proof. We consider a symmetrization of the matrix jewel, which we denote as

$$\mathcal{D}_{S\diamondsuit,\mathbf{k}} := \mathcal{W}_{max}(\mathrm{conv}\{-\mathcal{D}_{\diamondsuit,\mathbf{k}}(1) \cup \mathcal{D}_{\diamondsuit,\mathbf{k}}(1)\}).$$

Since the matrix jewel is a polytope on the first level, $\mathcal{D}_{S\diamondsuit,\mathbf{k}}$ is indeed a free spectrahedron. It holds that

$$\mathcal{D}_{\diamondsuit,\mathbf{k}} \subseteq \mathcal{D}_{S\diamondsuit,\mathbf{k}},$$

since the inclusion holds at level 1 and $\mathcal{D}_{S\diamondsuit,\mathbf{k}}$ is a maximal spectrahedron (see also [DDOSS17, Remark 4.2]). Let $\lambda \in [0,1]^g$ be such that

$$\lambda \cdot \mathcal{D}_{S\diamondsuit,\mathbf{k}}(1) \subseteq \mathcal{D}_{\diamondsuit,\mathbf{k}}(1).$$

Then, for any $B \in (\mathcal{M}_d^{sa})^{\sum_{j=1}^{g}(k_j - 1)}$, the implication

$$\mathcal{D}_{\diamondsuit,\mathbf{k}}(1) \subseteq \mathcal{D}_B(1) \implies \frac{1}{2d}\lambda \cdot \mathcal{D}_{S\diamondsuit,\mathbf{k}} \subseteq \mathcal{D}_B$$

holds by [BN18, Proposition VII.2], which generalizes [HKMS19, Theorem 1.4] to the complex setting. We can apply this result to the asymmetrically scaled spectrahedron, since $\lambda \cdot \mathcal{D}_A(n) \subseteq \mathcal{D}_B(n)$ if and only if $\mathcal{D}_A(n) \subseteq \mathcal{D}_{\lambda \cdot B}(n)$ for any free spectrahedra $\mathcal{D}_A$, $\mathcal{D}_B$ and any $n \in \mathbb{N}$. Therefore, $\lambda/(2d) \in \Delta(g, d, \mathbf{k})$. We only need to find the largest valid λ. As can be seen from comparing the extreme points, the symmetrization carries through the direct sum construction of the matrix jewel,

$$\mathcal{D}_{S\diamondsuit,\mathbf{k}}(1) = \bigoplus_{i=1}^{g} \mathrm{conv}\{-\mathcal{D}_{\diamondsuit,k_i}(1) \cup \mathcal{D}_{\diamondsuit,k_i}(1)\}.$$

We note that $X \in \mathcal{D}_A$ if and only if $X \in \mathcal{D}_{A \otimes I}$, which are the elements appearing as summands in the direct sum of free spectrahedra. By Lemma 3.17, the conditions on λ reduce to

$$\lambda_i \mathrm{conv}\{-\mathcal{D}_{\diamondsuit,k_i}(1) \cup \mathcal{D}_{\diamondsuit,k_i}(1)\} \subseteq \mathcal{D}_{\diamondsuit,k_i}(1)$$

for each $i \in [k]$. We see that $\lambda_i = 1/(k_i - 1)$ is a valid choice, since $\mathcal{D}_{\diamondsuit,k_i}(1)$ has extreme points $-k_i/2e_j$, $j \in [k_i - 1]$ and $k_i/2(1, \ldots, 1)$ by Lemma 4.3. Therefore, $-k_i/2(1, \ldots, 1)$ and $k_i/2e_j \in (k_i - 1)\mathcal{D}_{\diamondsuit,k_i}(1)$ for all $j \in [k_i - 1]$. □

Furthermore, we can approximate the matrix jewel by sets for which we know the inclusion constants. A convenient choice for such a set is the matrix diamond. A similar idea has been used in [Pas18, Section 2].

Theorem 7.2. *Let g, $d \in \mathbb{N}$, $k_j \in \mathbb{N}$, $\mathbf{k} = (k_1, \ldots, k_g)$. Then,*

$$\left(\frac{1}{(k_1-1)^2}, \ldots, \frac{1}{(k_g-1)^2}\right) \cdot \Delta\left(g, d, 2^{\times \sum_{i=1}^{g}(k_i-1)}\right) \subseteq \Delta(g, d, \mathbf{k}).$$

In particular,

$$\left(\frac{1}{(k_1-1)^2}, \ldots, \frac{1}{(k_g-1)^2}\right) \cdot \mathrm{QC}_{\sum_{i=1}^{g}(k_i-1)} \subseteq \Delta(g, d, \mathbf{k}).$$

Proof. We observe that

$$\mathcal{D}_{\jewel,k_i}(1) \subseteq \frac{k_i(k_i-1)}{2} \cdot \mathcal{D}_{\diamond,k_i-1}(1).$$

This follows from the computation of the ℓ_1 norms of the extremal points of the jewel base found in Lemma 4.3. Moreover, the matrix diamond is the maximal spectrahedron for the ℓ_1-ball. Thus, together with Lemma 3.17,

$$\mathcal{D}_{\jewel,\mathbf{k}} \subseteq \left(\frac{k_1(k_1-1)}{2}, \ldots, \frac{k_g(k_g-1)}{2}\right) \cdot \mathcal{D}_{\diamond,\sum_{i=1}^{g}(k_i-1)}$$

(see again [DDOSS17, Remark 4.2]). Furthermore, we need to find the largest $\lambda_i \geq 0$ for $i \in [g]$ such that

$$\lambda_i \cdot \mathcal{D}_{\diamond,k_i-1}(1) \subseteq \mathcal{D}_{\jewel,k_i}(1).$$

The extreme point of the matrix diamond are $\pm e_j$ for $j \in [k_i - 1]$. It holds that $\pm\lambda_i e_j \subseteq \mathcal{D}_{\jewel,k_i}(1)$ if and only if

$$\pm\lambda_i e_j \in \left[-\frac{k_i}{2}, \frac{1}{k_i-1}\frac{k_i}{2}\right].$$

This follows directly from Lemma 4.3. Thus, $\lambda_i \leq k_i/(2(k_i-1))$. From Lemma 3.17, we infer that

$$\left(\frac{k_1}{2(k_1-1)}, \ldots, \frac{k_g}{2(k_g-1)}\right) \cdot \mathcal{D}_{\diamond,\sum_{j=1}^{g}(k_j-1)}(1) \subseteq \mathcal{D}_{\jewel,\mathbf{k}}(1).$$

Let $B \in (\mathcal{M}_d^{sa})^{\sum_{j=1}^{g}(k_j-1)}$. Now, by the previous reasoning, the implication

$$\left(\frac{k_1}{2(k_1-1)}, \ldots, \frac{k_g}{2(k_g-1)}\right) \cdot \mathcal{D}_{\diamond,\sum_{j=1}^{g}(k_j-1)}(1) \subseteq \mathcal{D}_{\jewel,\mathbf{k}}(1) \subseteq \mathcal{D}_B(1) \implies$$
$$\left(s_1\frac{1}{(k_1-1)^2}, \ldots, s_g\frac{1}{(k_g-1)^2}\right) \cdot \mathcal{D}_{\jewel,\mathbf{k}} \subseteq \left(s_1\frac{k_1}{2(k_1-1)}, \ldots, s_g\frac{k_g}{2(k_g-1)}\right) \cdot \mathcal{D}_{\diamond,\sum_{j=1}^{g}(k_j-1)} \subseteq \mathcal{D}_B$$

holds for all $s \in \Delta(g, d, 2^{\times \sum_{j=1}^{g}(k_j-1)})$. As B was arbitrary, this proves the first assertion. The second follows from [BN18, Theorem VII.7], which adapts results from [PSS18]. □

8. Incompatibility witnesses and the matrix cube

In this section we introduce the notion of *incompatibility witnesses* in the case of tuples of binary POVMs. The case of general POVMs will be treated in the next section. We would like to point out that a related notion was recently introduced by A. Jenčová in [Jen18]; see also [CHT18] for yet another notion of incompatibility witness.

Let us start with a simple calculation motivating the new definition. Recall from [BN18, Theorem V.3] (or from Theorem 5.2) that a g quantum effects $E_1, \ldots, E_g \in \mathcal{M}_d$ are compatible if and only if for all elements of the matrix diamond $X \in \mathcal{D}_{\diamond,g}$, it holds that

$$\sum_{i=1}^{g}(2E_i - I_d) \otimes X_i \leq I_{dn}. \tag{10}$$

Recall that for a g-tuple $(X_1, \ldots, X_g) \in \mathcal{M}_n^{sa}$ to be an element of the matrix diamond, it needs to satisfy the following conditions:

$$\forall \varepsilon \in \{\pm 1\}^g, \qquad \sum_{i=1}^{g} \varepsilon_i X_i \leq I_n.$$

Let us now show, by a simple and direct computation, why compatible effects $E_1, \ldots, E_g$ must satisfy condition (10), for any choice of X as above. We write, for a joint POVM G,

$$\begin{aligned}
\sum_{i=1}^{g}(2E_i - I_d) \otimes X_i &= \sum_{i=1}^{g}\left[\sum_{\substack{\eta \in \{0,1\}^g \\ \eta_i = 0}} G_\eta - \sum_{\substack{\eta \in \{0,1\}^g \\ \eta_i = 1}} G_\eta\right] \otimes X_i \\
&= \sum_{i=1}^{g} \sum_{\eta \in \{0,1\}^g} (-1)^{\eta_i} G_\eta \otimes X_i \\
&= \sum_{\eta \in \{0,1\}^g} G_\eta \otimes \left[\sum_{i=1}^{g}(-1)^{\eta_i} X_i\right] \\
&\leq \sum_{\eta \in \{0,1\}^g} G_\eta \otimes I_n \\
&= I_{dn}.
\end{aligned}$$

The computation above justifies the following definition.

Definition 8.1. *A g-tuple of self-adjoint matrices $X \in (\mathcal{M}_n^{sa})^g$ is called an* incompatibility witness *if one of the following equivalent conditions holds:*

(1) *X is an element of the matrix diamond $\mathcal{D}_{\diamond,g}$*
(2) *for all sign vectors $\varepsilon \in \{\pm 1\}^g$, $\sum_{i=1}^{g} \varepsilon_i X_i \leq I_n$*
(3) *for all sign vectors $\varepsilon \in \{\pm 1\}^g$, $\| \sum_{i=1}^{g} \varepsilon_i X_i \|_\infty \leq 1$.*

We can now restate the second claims in [BN18, Theorem V.3] and Theorem 5.2 (applied to binary POVMs) as follows.

Proposition 8.2. *A set of d-dimensional quantum effects $(E_1, \ldots, E_g)$ is jointly measurable if and only if, for any incompatibility witness X, condition (10) holds. Moreover, one can restrict the size of the incompatibility witness to be d.*

Deciding whether a g-tuple of operators is an incompatibility witness requires to check 2^g matrix inequalities of size n, a task which is computationally intractable for large g. We relate this question to another free spectrahedral inclusion problem, that of the *complex matrix cube*. Recall from [HKMS19] that the matrix cube is the free spectrahedron

$$\begin{aligned}
\mathcal{D}_{\square,g} &:= \bigsqcup_{n=1}^{\infty} \{X \in (\mathcal{M}_n^{sa})^g \; : \; \|X_i\|_\infty \leq 1, \, \forall i \in [g]\} \\
&= \bigsqcup_{n=1}^{\infty} \left\{X \in (\mathcal{M}_n^{sa})^g \; : \; \sum_{i=1}^{g} c_i \otimes X_i \leq I_{2gn}\right\},
\end{aligned}$$

where the vectors $c_1, \ldots c_g \in \mathbb{C}^{2g}$ are given by

$$c_i = (e_i, -e_i).$$

We have the following result.

Proposition 8.3. *A g-tuple $X \in (\mathcal{M}_d^{sa})^g$ is an incompatibility witness if and only if $\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_X(1)$. Moreover, we have*

$$\mathcal{D}_{\square,g} \subseteq \mathcal{D}_X \implies \mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_X(1) \implies \vartheta^{\mathbb{C}}_{g,d}\mathcal{D}_{\square,g} \subseteq \mathcal{D}_X, \tag{11}$$

where $\vartheta^{\mathbb{C}}_{g,d}$ are the symmetric inclusion constants for the complex *matrix cube.*

Proof. The convex set inclusion $\mathcal{D}_{\square,g}(1) \subseteq \mathcal{D}_X(1)$ can be checked at the level of extremal points of the cube $\mathcal{D}_{\square,g}(1)$, which are the 2^g sign vectors $\varepsilon \in \{\pm 1\}^g$. The resulting conditions are precisely the ones from Definition 8.1. Equation (11) follows from the definition of the inclusion constants. □

Remark 8.4. *The inclusion constants $\vartheta^{\mathbb{C}}_{g,d}$ above are the maximal elements $s \in \Delta_{\mathcal{D}_{\square,g}}(g, d, 2^{\times g})$ such that $s_1 = \ldots = s_g$. They are known to possess a dimension independent lower bound, $g^{-1/2} \leq \vartheta^{\mathbb{C}}_{g,d}$* [PSS18, Section 6]*, which is known to be tight for d large enough.*

Remark 8.5. *The chain of implications (11) suggests an efficient numerical procedure to determine, up to some precision, whether a given g-tuple of self-adjoint operators is an incompatibility witness. This is because the first and the last free spectrahedral inclusions can be formulated as an SDP, as follows:*

$$\begin{aligned} &\textit{maximize} && s \\ &\textit{subject to} && \exists \Phi : \mathbb{C}^{2g} \to \mathcal{M}_d \textit{ unital, completely positive} \\ & && s\Phi(c_i) = X_i \quad \forall i \in [g]. \end{aligned}$$

If the value s^* of the SDP above is such that $s^* \geq 1$, we conclude that the first inclusion in (11) holds, so X is an incompatibility witness. On the other hand, if the optimal value is such that $s^* < \vartheta^{\mathbb{C}}_{g,d}$, we conclude that X is not *an incompatibility witness. However, if $s^* \in [\vartheta^{\mathbb{C}}_{g,d}, 1)$, we cannot conclude anything. Finally, let us point out that the SDP above has $3g+1$ constraints of size d, hence it is more tractable than the original brute-force condition, requiring 2^g matrix inequalities.*

We end this section with an example of an application of the theory of incompatibility witnesses. We shall prove that the upper bound derived in [ULMH16] for the amount of noise needed to make a g-tuple of "planar" qubit POVMs jointly measurable can also be understood in the framework of incompatibility witnesses.

Recall that a *planar qubit POVM* is a binary qubit POVM with effects which depend on only two Pauli operators (we choose σ_X and σ_Y below). In the case of planar qubit POVMs defined by vectors in the complex plane with angles in arithmetic progression, we have the following result.

Lemma 8.6. *Let $X = (X_1, \ldots, X_g)$ where X_j are planar qubit observables*

$$X_j = \cos(j\pi/g)\sigma_X + \sin(j\pi/g)\sigma_Y, \qquad j \in [g]. \tag{12}$$

Then, λX is a incompatibility witness if and only if $|\lambda| \leq \sin(\pi/(2g))$.

Proof. Let $\varepsilon \in \{\pm 1\}^g$. The condition $\|\sum_j \varepsilon_j \lambda X_j\|_\infty \leq 1$ reduces in this case, using the Bloch ball picture, to

$$|\lambda| \left\| \left(\sum_{j=1}^g \varepsilon_j \cos(j\pi/g), \sum_{j=1}^g \varepsilon_j \sin(j\pi/g) \right) \right\|_2 \leq 1 \iff |\lambda| \left| \sum_{j=1}^g \varepsilon_j \omega^j \right| \leq 1,$$

where $\omega = \exp(2\pi i/(2g))$ is a $2g$-th root of unity. Note that choosing $\varepsilon \equiv 1$ gives

$$1 \geq |\lambda| \left| \sum_{j=1}^g \omega^j \right| = \frac{2|\lambda|}{|1-\omega|} = \frac{|\lambda|}{\sin(\pi/(2g))},$$

proving one direction of the conclusion. For the other direction, note that $-\omega^j = \omega^{g+j}$, hence the signed sum of roots of unity corresponds to a sum of a subset of size g of $2g$-roots of unity. The conclusion will follow from the following claim, proving that any optimizer must be a rotation of the $\varepsilon \equiv 1$ case.

Claim. The maximization problem

$$\max_{J \subseteq [2g]} \left| \sum_{j \in J} \omega^j \right|$$

is attained for a subset J_0 with cardinality g and such that the set $\{\omega^j\}_{j \in J_0}$ is contained in some half-plane of $\mathbb{C} = \mathbb{R}^2$. Furthermore, if $j \in J_0$, then $j + g \notin J_0$ (the sums are considered modulo $2g$).

Indeed, let J be any maximizer, and let $s := \sum_{j \in J} \omega^j$. We show that $\{\omega^j\}_{j \in J}$ lies in the half-plane $\{z \in \mathbb{R}^2 \,:\, \langle s, z \rangle \geq 0\}$. We will use the following fact: For two non-zero vectors a, b in a real Hilbert space, $\langle a, b \rangle \geq 0$ implies $|a + b| > |a|$. Assume that there is some $j \in J$ with $\langle s, \omega^j \rangle < 0$. If $g + j \notin J$, replacing j with $g + j$ (taken cyclically) would increase the modulus of the sum, contradicting maximality. This is true, because the sum s' after replacement can be written $s' = s - 2\omega^j$ and $\langle s, -\omega^j \rangle > 0$. Then $|s'| > s$ by the fact above. If $g + j \in J$, the two contributions cancel, and we can consider $J' = J \setminus \{j, g + j\}$ and iterate. So, there is no $j \in J$ such that $\langle s, \omega^j \rangle < 0$. Conversely, if $j \in [2g]$ such that $\langle s, \omega^j \rangle \geq 0$, then $j \in J$. If this was not the case, we would have $|s + \omega^j| > |s|$ which contradicts maximality. Hence, $|J| \geq g$. Assume $|J| > g$. Then, there is an $l \in J$ such that also $g + l \in J$. By the above, this implies $-\langle s, \omega^l \rangle \geq 0$ and thus $|s - \omega^l| > |s|$. Removing l from J would thus increase the modulus, contradicting maximality.

The above claim implies that $\sum_{j \in J_0} \omega^j = \omega^k \sum_{j \in [g]} \omega_j$ for some $k \in [2g]$, as there are no more than $g + 1$ consecutive ω^j in a half-space. This proves the assertion since $|\omega^k| = 1$. □

Proposition 8.7. *Let g be a fixed positive integer, and consider the quantum effects*

$$E_j = \frac{1}{2}(I_2 + t_j X_j), \qquad j \in [g],$$

for some $t_j \in [0, 1]$, where X_j have been defined in (12). If the above effects are jointly measurable, then

$$\sum_{j=1}^{g} t_j \leq \frac{1}{\sin(\pi/(2g))}.$$

Proof. From the previous lemma, we know that $\sin(\pi/(2g))X$ is an incompatibility witness, hence so is $\sin(\pi/(2g))X^\top$

$$\sin(\pi/(2g)) \sum_{j=1}^{g} t_j X_j \otimes X_j^\top \leq I_4.$$

Let $\Omega = 1/2 \sum_{i,j=1}^{2} (e_i \otimes e_i)(e_j \otimes e_j)^*$ be the maximally entangled state, where $\{ e_1, e_2 \}$ is the basis of $\mathbb{C}^2$ with respect to which we transpose. By taking the Hilbert-Schmidt inner product of the previous inequality with Ω, we obtain

$$\sin(\pi/(2g)) \sum_{j=1}^{g} t_j \leq 1,$$

proving the claim. Here, we have used $\operatorname{tr}[\Omega A \otimes B] = 1/2 \operatorname{tr}\left[B^T A\right]$ and $\operatorname{tr}[\sigma_X \sigma_Y] = 0$, $\sigma_X^2 = I_2 = \sigma_Y^2$, by which $\operatorname{tr}\left[X_j^2\right] = 2 \ \forall j \in [g]$. □

Corollary 8.8. *The proposition above implies the following upper bound for the balanced compatibility regions* Γ *introduced in* [BN18] *for binary POVMs: for all* $g \geq 2$,

$$\Gamma(g, 2, 2^{\times g}) \subseteq \left\{ s \in [0,1]^g \, : \, \sum_{j=1}^{g} s_j \leq \frac{1}{\sin(\pi/(2g))} \right\}.$$

Remark 8.9. *Very similar ideas were used in the proof of* [BN18, Theorem VIII.8]. *There, it was shown that if* $F_1, \ldots, F_g$ *are anti-commuting, self-adjoint, unitary* $d \times d$ *matrices, then the* g*-tuple* $(s_1 F_1, \ldots, s_g F_g)$ *is an incompatibility witness for any unit norm vector* s*. As above, this observation, together with the "maximally entangled state trick" yields upper bounds on the sets* $\Gamma(g, d, 2^{\times g})$.

9. Incompatibility witnesses – the general case

We generalize here the notion of incompatibility witnesses introduced in the previous section for binary POVMs to the case of POVMs with arbitrary number of outcomes.

Definition 9.1. *Given a* g*-tuple of integers* $\mathbf{k}$*, we call the elements of the matrix jewel* $\mathcal{D}_{\diamondsuit,\mathbf{k}}$ incompatibility witnesses*. An incompatibility witness* $X \in \mathcal{D}_{\diamondsuit,\mathbf{k}}(n)$ *has the property that for all compatible POVMs* $E^{(1)}, \ldots, E^{(g)}$ *having* k_i *outcomes, respectively, the following inequality is satisfied*

$$\sum_{i=1}^{g} \sum_{j=1}^{k_i - 1} \left(2E_j^{(i)} - \frac{2}{k_i} I \right) \otimes X_{ij} \leq I_{dn}.$$

In order to decide whether a given $\sum_{i=1}^{g}(k_i - 1)$-tuple X is an incompatibility witness, one has to check $\prod_{i=1}^{g}(k_i - 1)$ matrix inequalities (see Definitions 4.1 and 3.12). When g is large, this task becomes computationally difficult, so it useful to formulate the above membership question as a spectrahedral inclusion problem which can benefit from tractable relaxations. To do so, we need to consider the dual object to the matrix jewel (base), which we introduce next.

Definition 9.2. *Consider the vectors* $x_1^{(k)}, \ldots, x_k^{(k)} \in \mathbb{C}^{k-1}$ *from Lemma 4.3 and define the vectors* $y_1^{(k)}, \ldots, y_{k-1}^{(k)} \in \mathbb{C}^k$ *by* $y_j(i) = x_i(j)$*, for all* $i \in [k]$ *and* $j \in [k-1]$:

$$y_j^{(k)} = \frac{k}{2}(e_k - e_j), \qquad j \in [k-1].$$

The free spectrahedron defined by

$$\mathcal{D}_{\blacksquare,k}(n) := \left\{ X \in (\mathcal{M}_n^{sa})^{k-1} : \sum_{j=1}^{k-1} y_j^{(k)} \otimes X_j \leq I_{kn} \right\}.$$

is called the matrix cuboid base*. For a* g*-tuple of non-negative integers* $\mathbf{k} = (k_1, \ldots, k_g)$*, we define the* matrix cuboid $\mathcal{D}_{\blacksquare,\mathbf{k}}$ *to be the free spectrahedron*

$$\mathcal{D}_{\blacksquare,\mathbf{k}} := \mathcal{D}_{\blacksquare,k_1} \hat{\times} \mathcal{D}_{\blacksquare,k_2} \hat{\times} \cdots \hat{\times} \mathcal{D}_{\blacksquare,k_g},$$

where the Cartesian product operation $\hat{\times}$ *for free spectrahedra was introduced in Equation* (4).

The definition above generalizes the notion of incompatibility witness from Definition 8.1 to the setting of g POVMs with arbitrary number of outcomes: $\mathcal{D}_{\square,g} = \mathcal{D}_{\blacksquare,2^{\times g}}$. Note also that, at level $n = 1$, the matrix jewel base and the matrix cuboid base are dual sets; in particular, $\mathcal{D}_{\blacksquare,k}(1)$ is a simplex.

Remark 9.3. *The matrix cuboid is the maximal matrix convex set (in the sense of* [DDOSS17, Section 4]*, see also Equation* (3)*) built on top of the Cartesian product of simplices*

$$\mathcal{D}_{\blacksquare,k_1}(1) \times \mathcal{D}_{\blacksquare,k_2}(1) \times \cdots \times \mathcal{D}_{\blacksquare,k_g}(1).$$

We display in Figure 3 some examples of the $n = 1$ of matrix cuboids.

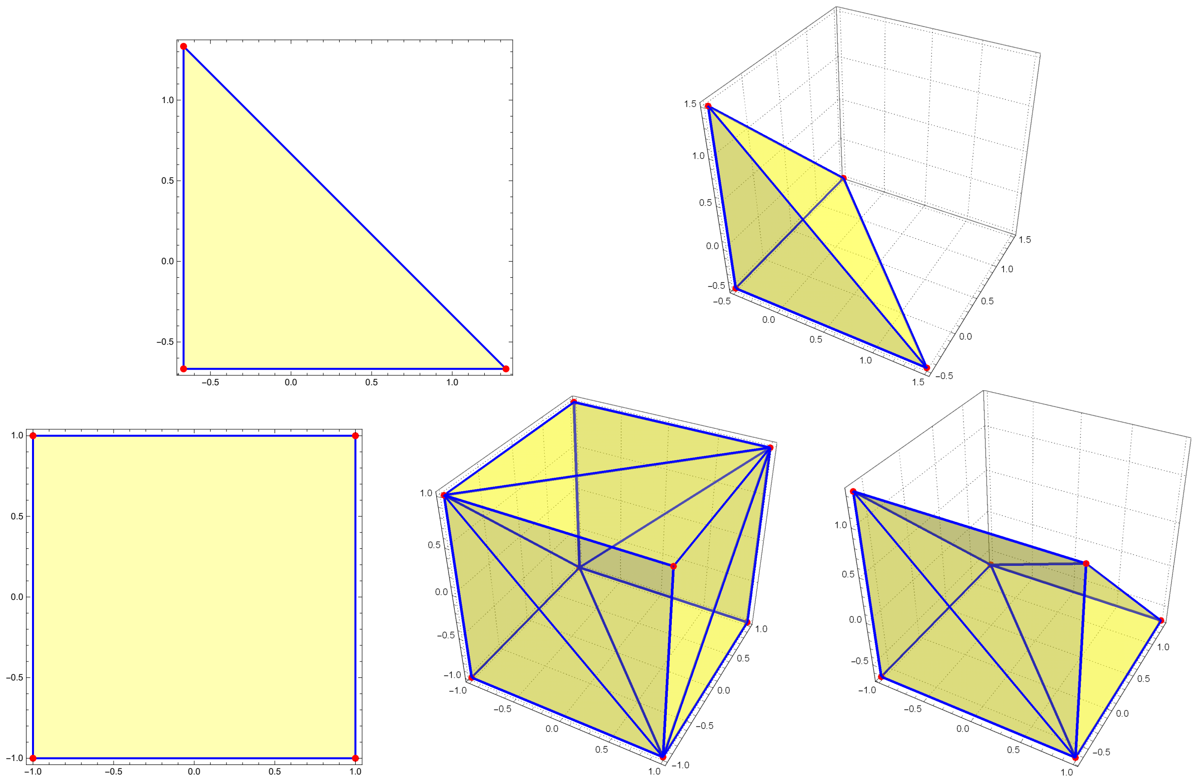

FIGURE 3. Top row: the spectrahedron level of the matrix cuboid base $\mathcal{D}_{\blacksquare,k}(1)$, for $k = 3, 4$. Bottom row: the spectrahedron level of the matrix cuboids $\mathcal{D}_{\blacksquare,(2,2)}(1)$, $\mathcal{D}_{\blacksquare,(2,2,2)}(1)$, and $\mathcal{D}_{\blacksquare,(2,3)}(1)$. The first two are in fact the matrix cubes $\mathcal{D}_{\square,2}(1)$ and $\mathcal{D}_{\square,3}(1)$ from [HKMS19] (a square and a cube), while the last polyhedron (a triangular prism, the Cartesian product of the triangle and the square) is new.

The relation between the notion of incompatibility witness and the matrix cuboid is given in the following result, which generalizes Proposition 8.3.

Proposition 9.4. *A g-tuple $X \in (\mathcal{M}_d^{sa})^{\sum_{i=1}^g (k_i-1)}$ is an incompatibility witness if and only if $\mathcal{D}_{\blacksquare,\mathbf{k}}(1) \subseteq \mathcal{D}_X(1)$.*

Proof. The condition in the statement can be checked at the level of the extreme points of $\mathcal{D}_{\blacksquare,\mathbf{k}}(1)$, which are Cartesian products of the extreme points

$$\operatorname{ext}\left(\mathcal{D}_{\blacksquare,k_i}(1)\right) = \{w_1, \ldots, w_k\},$$

where $w_i^{(k)} \in \mathbb{C}^{k-1}$ are given by

$$w_i^{(k)}(j) := -\frac{2}{k} + 2\delta_{i,j}, \qquad \forall i \in [k], \forall j \in [k-1].$$

Note that the vectors $w_i^{(k)}$ introduced above and the vectors $v_j^{(k)}$ from Definition 4.1 are related by $w_i^{(k)}(j) = v_j^{(k)}(i)$, for all $i \in [k]$ and $j \in [k-1]$. From the definition of the Cartesian product, it follows that the extremal points of the matrix jewel base are

$$\operatorname{ext}\left(\mathcal{D}_{❒,\mathbf{k}}(1)\right) = \{w_{\mathbf{i}}\}_{\mathbf{i}\in[\mathbf{k}]},$$

where

$$w_{\mathbf{i}}(s,j) = w_{i_s}^{(k_s)}(j), \qquad \forall j \in [k_s], \forall s \in [g].$$

The condition in the statement reads

$$\left(\forall \mathbf{i} \in [\mathbf{k}], \qquad \sum_{s=1}^{g} \sum_{j=1}^{k_s-1} w_{\mathbf{i}}(s,j) X_{s,j} \leq I\right) \iff \sum_{s=1}^{g} \sum_{j=1}^{k_s-1} v_{s,j} \otimes X_{s,j} \leq I,$$

where $v_{s,j}(\mathbf{i}) := w_{\mathbf{i}}(s,j) = w_{i_s}^{(k_s)}(j) = v_j^{(k_s)}(i_s)$. Equivalently, we have

$$v_{s,j} = \underbrace{I \otimes \cdots \otimes I}_{s-1 \text{ times}} \otimes v_j^{(k_s)} \otimes \underbrace{I \otimes \cdots \otimes I}_{k-s \text{ times}},$$

which are precisely the vectors defining the matrix jewel, see Equation (5). □

10. Discussion

In this section, we study the shape of the inclusion sets for the matrix jewel, before we conclude with some open questions. Contrary to the matrix diamond appearing in the study of binary measurements [BN18], the matrix jewel has not been studied in the literature on free spectrahedra. In algebraic convexity, the matrix convex sets having received the most attention are the matrix cube [BTN02, HKMS19], the different matricial notions of sphere [HKMS19, DDOSS17], and the maximal spectrahedra built upon ℓ_p spaces [PSS18]. These examples have symmetries that the matrix jewel lacks, rendering its structure more involved. Therefore, we only dispose of two kind of tools at this moment to study the structure of the matrix jewel. The first class are the results from quantum information theory presented in Section 6. The second class of results, derived in Section 7, compares the matrix jewel to more symmetric free spectrahedra.

In terms of lower bounds, we have shown in Proposition 6.7 that $\Gamma^{clone}(g, k_{max}d) \subseteq \Delta(g,d,\mathbf{k})$, where k_{max} is the maximal entry of $\mathbf{k}$. This implies in particular that for the balanced case in which $s_1 = \ldots = s_g$, we have

$$s_{max} \geq \frac{g + k_{max}d}{g(1+k_{max}d)}, \tag{13}$$

where s_{max} is the greatest balanced inclusion constant in $\Delta(g,d,\mathbf{k})$. We also obtain lower bounds from the symmetrization of the matrix jewel (see Theorem 7.1)

$$\left(\frac{1}{2d(k_1-1)}, \ldots, \frac{1}{2d(k_g-1)}\right) \in \Delta(g,d,\mathbf{k}) \tag{14}$$

and from the comparison with the matrix diamond (see Theorem 7.2)

$$\begin{aligned} \Delta(g,d,\mathbf{k}) &\supseteq \left(\frac{1}{(k_1-1)^2}, \ldots, \frac{1}{(k_g-1)^2}\right) \cdot \Delta\left(g, d, 2^{\times \sum_{i=1}^{g}(k_i-1)}\right) \\ &\supseteq \left(\frac{1}{(k_1-1)^2}, \ldots, \frac{1}{(k_g-1)^2}\right) \cdot \mathrm{QC}_{\sum_{i=1}^{g}(k_i-1)}. \end{aligned} \tag{15}$$

Lower Bounds		
cloning	$s \geq \frac{g+kd}{g(1+kd)}$	
symmetrization	$s \geq \frac{1}{2d(k-1)}$	
matrix diamond	$s \geq \frac{1}{(k-1)^2\sqrt{g(k-1)}}$	
Upper Bounds		
anti-commuting unitaries	$s \leq \frac{1}{\sqrt{g}}$	if $d \geq 2^{\lceil \frac{g-1}{2} \rceil}$
MUBs		if $g \leq$ max nb. of MUBs in $\mathbb{C}^d$ and $k = d$

TABLE 1. A comparison of all lower and upper bounds for the maximal s such that $(s, \ldots, s) \in \Delta(g, d, \mathbf{k})$, in the case where $\mathbf{k} = (k, \ldots, k)$.

Let g_d be the maximal number of MUBs which exist in a given dimension d. Then, the results gathered in Section 6.1 translate into upper bounds on $\Delta(g, d, d^{\times g})$, where $g \leq g_d$. For the balanced case, we know from [DSFB18] that

$$s_{max} \leq \frac{g + \sqrt{d}}{g(1 + \sqrt{d})}.$$

For the asymmetric case, we have from [Zhu15] that

$$\Delta(g, d, d^{\times g}) \subseteq \mathrm{QC}_g, \qquad g \leq g_d.$$

Here,

$$\mathrm{QC}_g = \left\{ s \in [0,1]^g : \sum_{i=1}^{g} s_i^2 \leq 1 \right\}$$

is the higher dimensional equivalent of the positive quarter of the unit circle in two dimensions. For $g = 2$, we have a tighter upper bound, namely the one from [CHT12] (see Proposition 6.1). Let

$$A = \left\{ s \in [0,1]^2 : s_1 + s_2 \leq 1 \right\} \cup \left\{ s \in [0,1]^2 : s_1^2 + s_2^2 + \frac{2(d-2)}{2}(1 - s_1)(1 - s_2) \leq 1 \right\}.$$

Then $A \subseteq \mathrm{QC}_2$ with equality for $d = 2$ and strict inclusion for $d > 2$ and

$$\Delta(2, d, d^{\times 2}) \subseteq A.$$

For more general bounds, we can use Proposition 3.23 together with Theorem 5.3. Let $\mathbf{k}$ such that $k_i \geq 2$ for all $i \in [g]$. Then,

$$\Delta(g, d, \mathbf{k}) \subseteq \Delta(g, d, 2^{\times g}).$$

The right hand side was studied in [BN18]. From [BN18, Theorem VIII.8], which uses results from [PSS18], we obtain

$$\Delta(g, d, \mathbf{k}) \subseteq \mathrm{QC}_g \qquad \forall d \geq 2^{\lceil \frac{g-1}{2} \rceil}.$$

Using the concept of inclusion witness, we can bound $\Delta(g, 2, 2^{\times g})$ for any g. We have seen in Corollary 8.8 that

$$\Delta(g, 2, 2^{\times g}) \subseteq \left\{ s \in [0,1]^g : \sum_{i=1}^{g} s_i \leq \frac{1}{\sin(\pi/(2g))} \right\}.$$

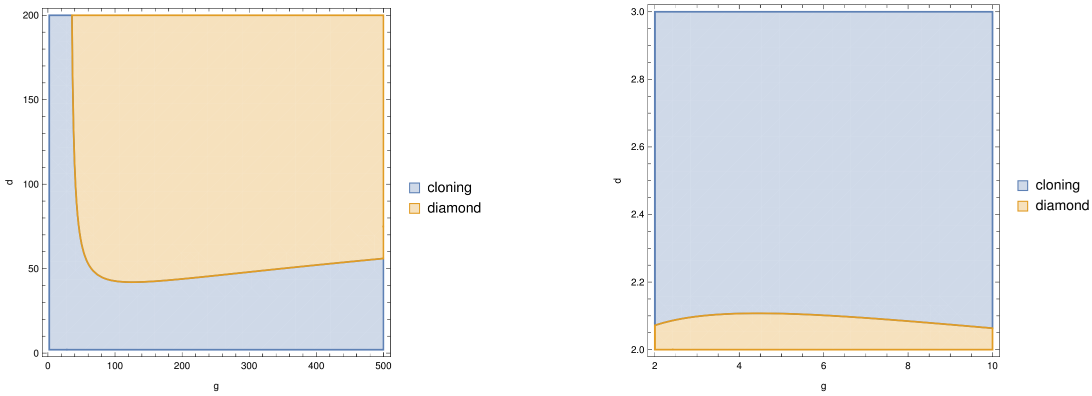

FIGURE 4. A comparison of the two lower bounds from equations (13) and (15), coming respectively from quantum cloning and from the comparison to the matrix diamond. On the left panel, we consider the case of g POVMs on $\mathbb{C}^d$ with $k = 3$ outcomes, while on the right panel we consider the case $k = d$. The regions correspond to the better (i.e. larger) lower bound.

We gather all these bounds in Table 1. In the case where POVMs have the same number of outcomes $k \geq 3$, it turns out that the bound (14) obtained by symmetrization is always weaker than the cloning bound (13). Note, however, that this is no longer the case for $\mathbf{k}$ in which not all entries are the same. We compare the cloning bound with the bound (15) coming from the comparison with the matrix diamond in Figure 4. It turns out that in the case where $k = d$ (the number of outcomes matches the dimension), the cloning bound always outperforms the diamond bound, except for qubits ($d = 2$).

For the balanced compatibility region $\Gamma(g, d, \mathbf{k})$, the lower bounds obtained via the symmetrization of the matrix jewel in Theorems 7.1 and 7.2 are new and improve over the lower bounds from asymmetric cloning for suitable choices of parameters (see Figure 4). The correspondence in Theorem 5.3 yields

$$\left(\frac{1}{2d(k_1 - 1)}, \ldots, \frac{1}{2d(k_g - 1)}\right) \in \Gamma(g, d, \mathbf{k})$$

and

$$\Gamma(g, d, \mathbf{k}) \supseteq \left(\frac{1}{(k_1 - 1)^2}, \ldots, \frac{1}{(k_g - 1)^2}\right) \cdot \mathrm{QC}_{\sum_{i=1}^{g}(k_i - 1)}.$$

As mentioned at the beginning of this section, all the bounds on the inclusion set for the matrix jewel we have obtained here stem either from quantum information theory or from some symmetrization technique. We leave it as an open question whether it is possible to obtain stronger bounds from the study of free spectrahedra, which would then have interesting consequences for quantum information theory. We also leave open the study of the matrix cuboid from Section 9, which can be seen as a generalization of the matrix cube. In particular, the inclusion constants for such free spectrahedra would allow to obtain, via Proposition 8.3, efficient criteria for deciding whether a tuple of matrices is an incompatibility witness for general POVMs.

Acknowledgments. A.B. acknowledges support from the ISAM Graduate Center at the Technische Universität München. I.N. and A.B. would like to thank Guillaume Aubrun for pointing us to the direct sum of convex sets. I.N.'s research has been supported by the ANR projects StoQ (grant number ANR-14-CE25-0003-01) and NEXT (grant number ANR-10-LABX-0037-NEXT), by the

PHC Sakura program (grant number 38615VA), and by the UEFISCDI (grant number PN-III-P1-1.1-MCT-2018-0015). Both authors also acknowledge the hospitality of M. Jivulescu and N. Lupa from the Universitatea Politehnică Timişoara, where most of this work was done.

References

[Bar02] Alexander Barvinok. *A course in convexity*, volume 54 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002. 5, 7, 8, 12

[BCP+14] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86:419–478, 2014. 2

[BN18] Andreas Bluhm and Ion Nechita. Joint measurability of quantum effects and the matrix diamond. *Journal of Mathematical Physics*, 59(11):112202, 2018. 2, 3, 4, 6, 7, 11, 13, 15, 17, 20, 21, 22, 25, 27, 28

[Boh28] Niels Bohr. The quantum postulate and the recent development of atomic theory. *Nature*, 121(3050):580–590, 1928. 2

[Bre97] David D. Bremner. On the complexity of vertex and facet enumeration for complex polytopes. *Ph.D. thesis, School of Computer Science, McGill University, Monréal, Canada*, 1997. 5

[BTN02] Aharon Ben-Tal and Arkadi Nemirovski. On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization*, 12(3):811–833, 2002. 1, 27

[CHT12] Claudio Carmeli, Teiko Heinosaari, and Alessandro Toigo. Informationally complete joint measurements on finite quantum systems. *Physical Review A*, 85:012109, Jan 2012. 16, 28

[CHT18] Claudio Carmeli, Teiko Heinosaari, and Alessandro Toigo. Quantum incompatibility witnesses. *arXiv preprint arXiv:1812.02985*, 2018. 21

[DDOSS17] Kenneth R. Davidson, Adam Dor-On, Orr Moshe Shalit, and Baruch Solel. Dilations, inclusions of matrix convex sets, and completely positive maps. *International Mathematics Research Notices*, 2017(13):4069–4130, 2017. 2, 3, 6, 12, 13, 20, 21, 26, 27

[DEBŻ10] Thomas Durt, Berthold-Georg Englert, Ingemar Bengtsson, and Karol Życzkowski. On mutually unbiased bases. *International Journal of Quantum Information*, 8(04):535–640, 2010. 16

[DSFB18] Sébastien Designolle, Paul Skrzypczyk, Florian Fröwis, and Nicolas Brunner. Quantifying measurement incompatibility of mutually unbiased bases. *arXiv preprint arXiv:1805.09609*, 2018. 16, 28

[Fin82] Arthur Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48(5):291–295, 1982. 2

[Has17] Anna-Lena Hashagen. Universal asymmetric quantum cloning revisited. *Quantum Information & Computation*, 17(9-10):0747–0778, 2017. 18

[Hei27] Werner Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3):172–198, 1927. 2

[HKM13] J. William Helton, Igor Klep, and Scott McCullough. The matricial relaxation of a linear matrix inequality. *Mathematical Programming*, 138(1-2):401–445, 2013. 1, 6, 7

[HKMS19] J. William Helton, Igor Klep, Scott McCullough, and Markus Schweighofer. Dilations, linear matrix inequalities, the matrix cube problem and beta distributions. *Memoirs of the American Mathematical Society*, 257(1232), 2019. 1, 2, 6, 20, 22, 26, 27

[HKR15] Teiko Heinosaari, Jukka Kiukas, and Daniel Reitzner. Noise robustness of the incompatibility of quantum measurements. *Physical Review A*, 92:022115, 2015. 10

[HMZ16] Teiko Heinosaari, Takayuki Miyadera, and Mário Ziman. An invitation to quantum incompatibility. *Journal of Physics A: Mathematical and Theoretical*, 49(12):123001, 2016. 2, 9

[HZ11] Teiko Heinosaari and Mário Ziman. *The Mathematical Language of Quantum Theory*. Cambridge University Press, 2011. 2, 9

[Jen18] Anna Jenčová. Incompatible measurements in a class of general probabilistic theories. *Physical Review A*, 98(1):012133, 2018. 21

[Kay16] Alastair Kay. Optimal universal quantum cloning: Asymmetries and fidelity measures. *Quantum Information & Computation*, 16(11 & 12):0991–1028, 2016. 19

[Key02] Michael Keyl. Fundamentals of quantum information theory. *Physics Reports*, 369(5):431–548, 2002. 19

[Pas18] Benjamin Passer. Shape, scale, and minimality of matrix ranges. *arXiv preprint arXiv:1803:09212*, 2018. 20

[PSS18] Benjamin Passer, Orr Moshe Shalit, and Baruch Solel. Minimal and maximal matrix convex sets. *Journal of Functional Analysis*, 274:3197–3253, 2018. 2, 21, 23, 27, 28

[SĆHM14] Michał Studziński, Piotr Ćwikliński, Michał Horodecki, and Marek Mozrzymas. Group-representation approach to $1 \to N$ universal quantum cloning machines. *Physical Review A*, 89(5):052322, 2014. 19

[ULMH16] Roope Uola, Kimmo Luoma, Tobias Moroder, and Teiko Heinosaari. Adaptive strategy for joint measurements. *Physical Review A*, 94(2):022109, 2016. 23

[UMG14] Roope Uola, Tobias Moroder, and Otfried Gühne. Joint measurability of generalized measurements implies classicality. *Physical Review Letters*, 113:160403, 2014. 10

[Wat18] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018. 9

[Wer98] Reinhard F. Werner. Optimal cloning of pure states. *Physical Review A*, 58(3):1827–1832, 1998. 18, 19

[WF89] William K. Wootters and Brian D. Fields. Optimal state-determination by mutually unbiased measurements. *Annals of Physics*, 191(2):363–381, 1989. 16

[ZHC16] Huangjun Zhu, Masahito Hayashi, and Lin Chen. Universal steering criteria. *Physical Review Letters*, 116(7):070403, 2016. 16, 17

[Zhu15] Huangjun Zhu. Information complementarity: A new paradigm for decoding quantum incompatibility. *Scientific reports*, 5:14317, 2015. 16, 17, 28

E-mail address: andreas.bluhm@ma.tum.de

Zentrum Mathematik, Technische Universität München, Boltzmannstrasse 3, 85748 Garching, Germany

E-mail address: nechita@irsamc.ups-tlse.fr

Laboratoire de Physique Théorique, Université de Toulouse, CNRS, UPS, France