Classicality and dense coding in the prepare-and-measure scenario

(Classicalidade e codificação superdensa no cenário de preparação e medição)

Carlos Augusto Belini de Gois

Supervisor:
Prof. Dr. Rafael Luiz da Silva Rabelo

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE FÍSICA "GLEB WATAGHIN"
DEP. DE FÍSICA DA MATÉRIA CONDENSADA

The Telescope, the Fluxions, the invention of Logarithms and the frenzy of multiplication, often for its own sake, that follow'd have for Emerson all been steps of an unarguable approach to God, a growing clarity,— Gravity, the Pulse of Time, the finite speed of Light present themselves to him as aspects of God's character. It's like becoming friendly with an erratic, powerful, potentially dangerous member of the Aristocracy. He holds no quarrel with the Creator's sovereignty, but is repeatedly appall'd at the lapses in Attention, the flaws in Design, the squand'rings of life and energy, the failures to be reasonable, or to exercise common sense,— first appall'd, then angry. We are taught,— we believe,— that it is love of the Creation which drives the Philosopher in his Studies. Emerson is driven, rather, by a passionate Resentment.

(Thomas Pynchon — Mason & Dixon)

Abstract

AAAAAAAAAAAAAAA

Resumo

BBBBBBBBBBBBBBB

Publications

Papers

- Carlos de Gois, George Moreno, Ranieri Nery, Samuraí Brito, Rafael Chaves, and Rafael Rabelo. General method for classicality certification in the prepare and measure scenario. *arXiv preprint arXiv:2101.10459*, 2021
- George Moreno, Ranieri Nery, Carlos de Gois, Rafael Rabelo, and Rafael Chaves. Semi-device-independent certification of entanglement in superdense coding. *Phys. Rev. A*, 103:022426, Feb 2021

Code

- C. de Gois. Code for "General method for classicality certification in the prepare and measure scenario" ([1]). `https://github.com/cgois/pam_classicality`, 2021

Acknowledgments

Contents

Introduction

In the widest sense, *quantum information* is the study of the encoding, processing and decoding of information by means of quantum systems. Its origins can be traced back to the first decades following the inception of quantum theory [4, 5, 6], and, most notably, to the now called Bell's theorem [7]. These earlier discussions were largely focused on conceptual puzzles arising from the theory; a similar motivation that now grounds the field we call quantum foundations. During the 80's and 90's, groundbreaking results in quantum computing and, most notably, quantum cryptography [8, 9], connected informational with foundational aspects of quantum theory, ultimately paving the way that turned quantum information into a vast and thriving research field.

Correlation scenarios are central to quantum information. In a correlation scenario, two or more parties locally interact with and read out information from their respective systems. Collecting the results obtained by each party, one can analyze whether they are independent or correlated. In the latter case, we can further ask *to what extent* such correlations are present. When the underlying systems are assumed to be classical, these correlations obey certain bounds that, strikingly, can be violated by some quantum systems. This puts a clear divide between the behaviors of classical, quantum, and even more exotic systems [10]. Because this disctinction is observable from measurement results alone, correlation scenarios can be used to device-independently certify quantum behaviors. In the *device independent* paradigm, each system is inside a black-box to which we impose no extra assumptions. All observations happen by providing inputs to the boxes and collecting their outputs. Even if no communication is allowed between the boxes, certain quantum systems are still certifiably so. Because of this no-signalling hypothesis, any correlation manifest between the systems must have been present from the start. Remarkably, it is possible to show that quantum behaviors in this so-called Bell nonlocality scenario require *entanglement* and *measurement incompatibility*, which are two structures at the core of the most surprising phenomena described by quantum theory. Thus, observing nonlocal behaviors, more than showing us the boxes must contain quantum systems, also tells us these systems must be entangled, and the measurements, albeit uncharacterized, have to be incompatible. Nonlocal behaviors are precisely the characteristic that allow for provably secure quantum key distribution protocols, randomness certification, and many other informational applications of device independent correlation scenarios [11]. From a more foundational perspective, stashing systems into black boxes allow us to hide implementation details, and thus examine classical, quantum and other theories for themselves. The question of what, exactly, is *quantum* in quantum theory has puzzled scientists for more than a century, and the device independent paradigm is closely linked to modern investigations on this topic [12, 13, 14].

An important relaxation on the nonlocality scenario is to allow for signalling between devices. These (semi-)device independent communication scenarios are attracting growing attention, as particularly in the case of *prepare-and-measure* scenarios. Part of this interest is due to the fact that they are useful for quantum key distribution and self-testing of quantum systems, and can be seen

as building blocks for quantum communication networks. In the simplest instance of a prepare-and-measure scenario, a preparation device receives an input, then prepares and communicates *some* system to a measurement device which, given another input by a second observer, outputs a result. Again, only observational data is collected by each of the two parties, which may later be collectively analyzed. Similarly to how we can tell quantum from classical behaviors apart in nonlocality scenarios, it is sometimes also possible to device independently certify that quantum — rather than classical —, communication happened in a prepare-and-measure experiment. A practical question arising in this context is whether some set $\mathcal{S}$ of quantum preparations can manifest nonclassical behaviors. Measurements play a crucial role in this regard. To necessarily and sufficiently certify the classicality of $\mathcal{S}$, one needs to test the preparations under all the infinitely many possible measurements. This is precisely the problem we tackle in chap. 4, where we devise and a general method to certify whether an arbitrary set of preparations always behaves in a way that classical systems also could. When this is the case, $\mathcal{S}$ is not useful for quantum enhancement in communication protocols. More than constructing the method, we also use it prove the existence of a quantum advantage activation phenomenon in random access coding, which is an important communication protocol. In that same chapter, we turn the question inside-out by asking whether some set $\mathcal{M}$ of quantum *measurements* is useful to reveal nonclassical behaviors. Through the method built for this second task, we prove there are incompatible measurements that nevertheless cannot give rise to nonclassical behaviors. This is in stark contrast to quantum steering scenarios [15, 16, 17], where incompatibility is necessary and sufficient for steerability [18, 19].

Reaching the extreme opposite side of correlation scenarios with communication, we can also assume the devices to be fully characterized. This is exactly the case for the well-known quantum teleportation [20] and dense coding schemes [21]. In the paradigmatic *dense coding* protocol, one party communicates a d-dimensional quantum system to another. One can show that, by exploiting pre-existing entanglement, the receiving end may perfectly recover two classical $d$its of information, thus doubling the capacity of a noiseless classical communication channel. This celebrated protocol is also remarkable for its theoretical simplicity, and its applicability paves the way towards the development of quantum technologies. However, these protocols have only been discussed in the device *dependent* case, where we have full knowledge on which states may be prepared and what measurements can be applied. In chap. 5, we propose a (semi-)device independent formulation of this protocol. Not surprisingly, this can be done using the framework of prepare-and-measure scenarios but — now surprisingly —, this has not been much discussed before. Our formulation of dense coding as a prepare-and-measure instance can be used to certify lower bounds on the Schmidt number of the pre-existing entangled resource, witness the entanglement of any isotropic state, and self-test maximally entangled resources, among other results therein shown. Moreover, as a first step towards the study of arbitrary quantum correlated prepare-and-measure scenarios, which are largely missing in the literature, we also study an specific case where more than one measurement choice is allowed by the measuring device.

Part I, where we review basic aspects of quantum theory, convexity and mathematical optimization, should be particularly useful to newcomers. These tools are indispensable to Part II, in which we discuss general instances of prepare and measure scenarios (chap. 3), classicality certification methods (chap. 4), and the formulation of device-independent dense coding (chap. 5). Appendix A provide computational details for all our classicality results, and appendix B proves all results presented in chap. 5.

Part I

Tools

Chapter 1

Quantum theory

States, transformations and measurements are the basic building blocks in the description of any physical system. In this chapter, I review the mathematical structures that quantum theory assigns to each of these building blocks, emphasizing the aspects that most drastically differ from their classical counterparts. Pedagogical introductions from a similar viewpoint are found in [22, 23, 24, 25, 26].

1.1 States

A structure at the core of the most unusual of quantum behaviors is *entanglement* — a property that may or not be present in composite quantum systems. The unusual correlations that entanglement may originate were pointed out as early as 1935 by Einstein, Podolsky and Rosen [4], and also discussed by Schrödinger, who coined the term (originally, "Verschränkung") [27]. This discussion was later rescued by John Bell in 1964 [7], and since then been extensively tested and developed [28], ultimately becoming a central feature of many quantum informational protocols, such as dense coding [21], quantum key distribution [8] and teleportation [20]. Entanglement theory is a vast and endlessly interesting field of study in itself, some aspects of which I now review with special focus on bipartite systems, and making no attempt of comprehensiveness.

* * *

A *quantum state* is described by a *density operator*, commonly denoted by ρ. Any density operator is a linear, unit-trace and positive-semidefinite operator in a Hilbert space $\mathcal{H}$. Conversely, any operator satisfying these properties represents a valid quantum state. Hence,

$$\mathcal{D}(\mathcal{H}^d) = \{\rho \in \mathcal{L}(\mathcal{H}^d) \mid \mathrm{tr}\rho = 1, \rho \succeq 0\}, \tag{1.1}$$

where $\mathcal{L}(\mathcal{H}^d)$ is the set of linear operators in $\mathcal{H}^d$, is the space of density operators in dimension d. Only finite-dimensional Hilbert spaces will be considered.

As $\rho \succeq 0 \Longrightarrow \rho = \rho^\dagger$, we can use the spectral decomposition to write $\rho = \sum_m m\Pi_m$, where each Π_m is a projection onto the eigenspace of the eigenvector of ρ associated with eigenvalue m. Such eigenvectors are orthogonal. They need not be normalized, but we can always and will take them as being. Orthogonality implies that $\Pi_m\Pi_n = \delta_{mn}\Pi_m$ and $1 \leq \mathrm{rank}(\rho) \leq d$, and normalization that $\mathrm{tr}\,(\Pi_m) = 1$. Furthermore, all $m \geq 0$, and $\mathrm{tr}\,(\rho) = 1 \Longrightarrow \sum_m m = 1$.

You may recall the more usual definition that a quantum state is described by a unit vector in a Hilbert space and, conversely, that such unit vectors describe quantum states. This is only true

for a subset of states called *pure quantum states*. Following along the tradition, we will denote pure quantum states as $|\psi\rangle$, where ψ is some label that describes the state. Similar notation is used for the dual vector $(|\psi\rangle)^\dagger \equiv \langle\psi|$, which is useful to write the inner product between any two vectors in the same space as $\langle\psi|\phi\rangle$, and the outer product as $|\psi\rangle\langle\phi|$. An useful geometric intuition on these products is to interpret the inner product as the overlap between $|\psi\rangle$ and $|\phi\rangle$, and an outer product $|\psi\rangle\langle\psi|$ as a projection onto $|\psi\rangle$.

Recalling that all eigenvectors of a density operator ρ are normalized, we may now interpret the $\Pi_m \equiv |m\rangle\langle m|$ as projections onto the pure states labeled by $|m\rangle$, and the spectral decomposition $\rho = \sum_m m\Pi_m$ as a probability distribution, weighted by the eigenvalues m, over those. Any pure state $|\psi\rangle$ can equivalently be described as $\rho = |\psi\rangle\langle\psi|$, and whenever $\operatorname{rank}(\rho) = 1$, we may infer that ρ stands for a pure state. Equivalently, whenever ρ is such a one-dimensional projection, the *purity* $\operatorname{tr}\left(\rho^2\right) = 1$, while in general $1/d \leq \operatorname{tr}\left(\rho^2\right) \leq 1$. This is one of the reasons why density matrices are more general than pure states. All other density operators (i.e., those of non-unit rank) are said to describe *mixed quantum states*. Although the spectral decomposition of ρ suggests that a mixed state can be interpreted as a probability distribution over pure states, the understanding of a mixed state as lack of knowledge on the exact state of the system should not be taken literally. One of several reasons for this assertion is that there may be many pure state ensembles generating the same density operator [29].

Given a basis $\{|e_i\rangle\}_{i=1}^d$ for $\mathcal{H}^d$, any pure state $|\psi\rangle$ in $\mathcal{H}^d$ can be written as $|\psi\rangle = \sum_{i=1}^d c_i |e_i\rangle$, where the $c_i \in \mathbb{C}$ and $\sum_i |c_i|^2 = 1$ due to $\langle\psi|\psi\rangle = 1$. We will frequently be interested in $\mathcal{H}^2$, in which it's common to work with the orthonormal *computational basis* $\{|0\rangle, |1\rangle\}$. The vector representations associated with the computational basis elements are $|0\rangle \equiv (1\ 0)^\intercal$ and $|1\rangle \equiv (0\ 1)^\intercal$. Any $|\psi\rangle \in \mathcal{H}^2$ can thus be identified with $|\psi\rangle = c_1|0\rangle + c_2|1\rangle = (c_1\ c_2)^\intercal$. An extension to a generalized d-dimensional computational basis $\{|i\rangle\}_{i=0}^{d-1}$ is similarly done. Due to its analogy with two-level classical systems (bits), a $|\psi\rangle \in \mathcal{H}^2$ is termed a quantum bit (*qubit*) and, similarly, any $|\psi\rangle \in \mathcal{H}^d$ is a qu*d*it.

Entanglement — and its opposite concept, *separability* —, are properties related to composite quantum systems. If we choose 2 for the number of subsystems, the underlying Hilbert space $\mathcal{H}$ of a state ρ can be correspondingly factored as $\mathcal{H} \equiv \mathcal{H}_A \otimes \mathcal{H}_B$, for choices of $\mathcal{H}_A$ and $\mathcal{H}_B$ respecting $\dim \mathcal{H}_A \dim \mathcal{H}_B = \dim \mathcal{H}$. Using the tensor product for composition is the quantum analogue of using the Cartesian product to build composite phase spaces in classical mechanics.

Letting $\{|\psi_i\rangle\}_{i=1}^{d_A}$ and $\{|\varphi_\alpha\rangle\}_{\alpha=1}^{d_B}$ be orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, we can easily build an orthonormal basis for $\mathcal{H}$ as $\{|\psi_i\rangle \otimes |\varphi_\alpha\rangle\}_{i=1,\alpha=1}^{d_A,d_B}$. This means that any vector $|\psi\rangle \in \mathcal{H}$ has a decomposition

$$|\psi\rangle \sum_{i=1}^{d_A} \sum_{\alpha=1}^{d_B} c_{i\alpha} |\psi_i, \varphi_\alpha\rangle . \tag{1.2}$$

Analogously, with $\{|\psi_i\rangle\langle\psi_j|\}_{i,j=1}^{d_A}$ as a basis for $\mathcal{L}(\mathcal{H}_A)$ and $\{|\varphi_\alpha\rangle\langle\varphi_\beta|\}_{\alpha,\beta=1}^{d_B}$ one for $\mathcal{L}(\mathcal{H}_B)$, any operator $O \in \mathcal{L}(\mathcal{H})$ may be decomposed as

$$O = \sum_{ij\alpha\beta} O_{ij\alpha\beta} |\psi_i\rangle\langle\psi_j| \otimes |\varphi_\alpha\rangle\langle\varphi_\beta| \tag{1.3}$$

Suppose A is an operator acting only on the part $O^A \in \mathcal{L}(\mathcal{H}_A)$ of O. The corresponding operator in $\mathcal{H}$ is just $A \otimes \mathbf{1}_B$. To find a description for O^A, we notice that the expectation value $\operatorname{tr}(A \otimes \mathbf{1}_B O)$ should be equal to $\operatorname{tr}\left(A O^A\right)$, and to comply with it we define the *partial trace* over

B, $\mathrm{tr}_B : \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B) \mapsto \mathcal{L}(\mathcal{H}_A)$, as

$$\mathrm{tr}_B(O) \equiv \sum_{ij\alpha\beta} O_{ij\alpha\beta}\, \mathrm{tr}\left(|\psi_i\rangle\langle\psi_j|\right) \otimes |\varphi_\alpha\rangle\langle\varphi_\beta| = \sum_{i\alpha\beta} O_{ii\alpha\beta}\, |\varphi_\alpha\rangle\langle\varphi_\beta|$$

and call $O^A = \mathrm{tr}_B(O)$ the reduced operator. This definition can be trivially adapted to tracing out $\mathcal{H}_A$ instead, and to dealing with more than two subsystems. Moreover, it can be shown that this is the unique operation satisfying the expectation value equality condition, and that it is also completely positive and trace preserving (the significance of these latter conditions will be discussed later) [][**To-Do:**]. This operation is especially useful when applied to density operators, in which case we call $\mathrm{tr}_B(\rho) = \rho^A$ the *reduced state* (of ρ in subsystem A).

Given a factorization of $\mathcal{H}$, a state ρ acting on $\mathcal{H}$ is said to be *separable* if and only if it can be written as

$$\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B \tag{1.4}$$

where the p_i are probabilities, and $\rho_i^A \in \mathcal{H}_A$; correspondingly for ρ_i^B. A state that is not separable is entangled. For pure states, this reduces to $|\psi\rangle_{AB} = |\psi\rangle_A \otimes |\psi\rangle_B$, and when this form is not possible, $|\psi\rangle_{AB}$ is entangled.

Asking whether a state $\rho \in \mathcal{H}$ is entangled is only meaningful when the factorization structure is specified. As a matter of fact, any bipartite pure entangled state $|\psi\rangle \in \mathcal{H}_A^2 \otimes \mathcal{H}_B^2$ can be made separable by a fitting choice of factorization [30]. This observation implies that entanglement is a property of a state *with respect to* a choice of subsystems, and not of the state in itself. Naturally, one may also discuss entanglement in larger number of subsystems [28], but the complexity scales significantly fast, and the discussion would be of little usefulness to our objective.

A rationale for the definition of separability comes from a preparation procedure [31]. Consider two separate laboratories, each equipped with a device that prepares quantum states, and sharing a source of (classical) randomness. Given a random number i, generated by the source with probability p_i, the laboratories locally prepare states ρ_i^A and ρ_i^B. Now suppose the first laboratory measures $\mathcal{M}_A$, and the second $\mathcal{M}_B$, each on their respective preparation. For a pair of measurement effects $E_{m_A} \in \mathcal{M}_A$ and $E_{m_B} \in \mathcal{M}_B$, we then have

$$p(m_A, m_B) = \sum_i p_i \mathrm{tr}\left(E_{m_A}\rho_i^A\right) \mathrm{tr}\left(E_{m_B}\rho_i^B\right) = \mathrm{tr}\left(E_{m_A} \otimes E_{m_B}\rho\right).$$

In the last equality, ρ matches the definition of a separable state.

With this discussion, it is also clearer that separable states are *not* uncorrelated. However, they may only exhibit correlations as strong as the ones possible in classical systems, and for this reason are said to be *classically correlated*. Entangled states, conversely, manifest correlations that are not classically reproducible, which makes it a intrinsically nonclassical property.

Properly justified, the definition of entanglement is quite amicable. It is not, however, computationally friendly, and determining whether a given state ρ can be decomposed as in eq. (1.4) or not can be a daunting task. Even in bipartite structures, the problem is fully solved only under special circumstances, such as for pure states, dimensionally limited Hilbert spaces, or for some special families of quantum states, including Werner states and isotropic states. These will become important in due time, so we discuss them now.

For bipartite pure states of any dimension, the problem can be fully solved through the so-called Schmidt decomposition. The Schmidt decomposition theorem states that any $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ can

be decomposed as

$$|\psi\rangle = \sum_{i=1}^{d} \eta_i \, |i_A\rangle \otimes |i_B\rangle \,, \tag{1.5}$$

where $\{|i_A\rangle\}_{i=1}^{d_A}$ and $\{|i_B\rangle\}_{i=1}^{d_B}$ are orthonormal bases for $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively, $\eta_i \geq 0$, and $d = \min\{d_A, d_B\}$. We call η_i *Schmidt coefficients*, denote by $r_S(\psi)$ the number of non-zero coefficients (*Schmidt rank*), and say $\{\eta_i(\psi)\}_{i=1}^{r_S(\psi)}$ is the *Schmidt spectrum*.

Contrasted to eq. (1.2), this is a remarkable simplification. For one, this representation requires a single sum, but also, d is the *minimum* local dimension, irrespective of how (finitely) large the other dimension may be. It also resembles the definition of separability versus entanglement for pure states. Actually, it can be shown that a pure bipartite state $|\psi\rangle$ is entangled if and only if its Schmidt rank $r_S(\psi)$ is larger than one; equivalently, if the Schmidt decomposition has more than one term, or if any $\eta_i = 1$, because $\sum_i \eta_i^2 = 1$.

This observation urges us to ask: are some states *more entangled* than others, and can a Schmidt "something" be used to measure this? Attempting to adequately discuss *entanglement measures* would be going too far. However, as some introductory basic concepts will turn useful, we discuss them with no intention on reproducing the thoroughness that can be found in [32, 33, 28, 34].

The first important thing is that there is a whole zoo of entanglement measures, such as concurrence, negativity, entanglement of formation, etc. The second is that they do not always agree with each other on the ordering they impose on the set of entangled states. Thus, depending on the intended use, there may be some measure more adequate than another. Nevertheless, there are several desirable properties for an entanglement measure $E : \mathcal{D}(d) \mapsto \mathbb{R}$ to satisfy, such as $E(\rho) = 0$ if ρ is separable, and that it should not increase under local operations with classical communication (LOCC); a condition reminiscent of, but actually weaker than the operational definition of separability given above.

Making the Schmidt spectrum respect the first condition is easy (just subtract 1). Using it to impose a partial ordering on the set of pure states requires some extra caution, but it can be done through majorization. We first order the *Schmidt spectrum* $\{\eta_i(\psi)\}_{i=1}^{r_S(\psi)}$ of some state $|\psi\rangle$ in non-increasing order, then define

$$\psi \prec \varphi \quad \Longleftrightarrow \quad \sum_{i=1}^{r} \eta_i^2(\psi) \leq \sum_{i=1}^{r} \eta_i^2(\varphi), \ \forall r.$$

If $\psi \prec \varphi$, we say that ψ is majorized by φ, or that φ majorizes ψ. In relation to entanglement, ψ would then be *more* entangled than $|\varphi\rangle$, in the sense that we may convert $|\psi\rangle$ to $|\varphi\rangle$ solely by means of LOCC [35, 33]. With this in mind, states for which $\eta_i = 1/\sqrt{d}$ are said to be *maximally entangled*.

Although the Schmidt decomposition only works for pure states, the idea of the Schmidt rank can be nicely generalized to an entanglement measure over mixed states. The so called *Schmidt number* [36] is given by

$$r_S(\rho) = \min_{\{|\psi_i\rangle\}_i} \{\max_i [r_S(\psi_i)]\}, \tag{1.6}$$

where I reuse the notation r_S from the Schmidt rank because the two notions are equivalent for pure states.

Arguably opaque, this definition is better understood through a procedure. Starting from ρ, we find an ensemble of pure states $\rho = \sum_i p_i \, |\psi_i\rangle\langle\psi_i|$ for it, list the $r_S(\psi_i)$ for each state of the ensemble, and take the maximum. This corresponds to the inner maximization. However, the decomposition we choose for ρ is not, in general, unique. So we do this procedure for all possible

sets of pure states $\{\psi_i\}_i$ that may be used to build ρ, and take the minimum element of the resultant set, which is what the outer minimization means. Denoting as S_k the set of density operator with Schmidt number less than or equal to k, it will be of most importance for us that $S_{k-1} \subset S_k$, that S_1 is the set of separable states, that each S_k is a convex set, and that its extremal points are the pure states.

Going back to the problem of determining whether a given ρ is entangled, whenever we limit the dimensions as $\mathcal{H} = \mathcal{H}^2 \otimes \mathcal{H}^2$ or $\mathcal{H} = \mathcal{H}^2 \otimes \mathcal{H}^3$, the *positive partial transpose* (PPT, also "Peres-Horodecki") criterion provides a necessary and sufficient condition. In all other cases, the condition is still sufficient, though not necessary. To understand the sufficiency affirmation, we recall that the *partial transpose* is a transposition operation acting only on some subsystems. Reusing the decomposition in eq. (1.3), the partial transpose over B is defined as

$$O^{\intercal_B} = (\mathbf{1}_A \otimes T)\, O = \sum_{ij\alpha\beta} O_{ij\alpha\beta}\, |\psi_i\rangle\langle\psi_j| \otimes |\varphi_\beta\rangle\langle\varphi_\alpha| = \sum_{ij\alpha\beta} O_{ij\beta\alpha}\, |\psi_i\rangle\langle\psi_j| \otimes |\varphi_\alpha\rangle\langle\varphi_\beta|\,,$$

where T stands for the transposition map. Now suppose that we take a separable state ρ and transpose, for instance, its second subsystem (the argument is equivalent for transposing the other). Then $\rho^{\intercal_B} = \sum_i p_i \rho_i^A \otimes \left(\rho_i^B\right)^\intercal$. Building on the fact that ρ_i^B was a valid density operator, and that the transpose preserves its trace and eigenvalues, it follows that $\left(\rho_i^B\right)^\intercal$ is also a density operator. With ρ_i^A left unchanged, this implies that $\rho^{\intercal_B} \succeq 0$. Consequently, all separable states have positive partial transpose, which also implies that if the partial transpose of some ρ has negative eigenvalues, it must be entangled.

Shortly after Peres made this argument [37], Horodecki et al. showed that for $d_A d_B \leq 6$ the PPT criterion is actually necessary *and* sufficient: no entangled states in these factorizations have positive partial transpose [38]. For larger dimensions, though, they also prove this is not always true; except under special circumstances.

One such special case is that of *Werner states*. They were central to the first proof that entanglement and Bell nonlocality[39, 11] are not equivalent concepts [31]. More specifically, it was shown that a large set of entangled Werner states are nevertheless local under projective measurements. Later, the bound for locality in $d = 2$ was improved several times [40, 41, 42], the result was extended to POVMs [43], and they were also made pivotal in the study of quantum steering [15, 17, 16] and as a test bed for the capabilities of many quantum informational protocols, such as (semi)device-independent entanglement witnesses. What makes them especially tractable is that they are highly symmetric, as Werner states are bipartite states in $\mathcal{H}^d \otimes \mathcal{H}^d$ for which $(U \otimes U)\, \rho\, (U^\dagger \otimes U^\dagger) = \rho$. It can be shown that this is a one-parameter family of states, and that they may be written as

$$W_d(\alpha) = \left(\frac{d-1+\alpha}{d-1}\right) \frac{\mathbf{1}}{d^2} - \left(\frac{\alpha}{d-1}\right) \frac{S}{d}. \tag{1.7}$$

Here, $S = \sum_{i,j=0}^{d-1} |ij\rangle\langle ji|$ is the swap operator. When written in this form, $\alpha = 0$ stands for the maximally mixed state, and $\alpha \leq 1$. Werner states are entangled if and only if $\alpha > \frac{1}{d+1}$ but, under projective measurements, they are unsteerable if and only if $\alpha \leq 1 - \frac{1}{d}$, as shown in [15].

A second family of states that will become useful in due time are the *isotropic states*. Bipartite and also highly symmetric, they are defined as states in $\mathcal{H}^d \otimes \mathcal{H}^d$ for which $(U \otimes U^*)\, \rho\, (U^\dagger \otimes U^{*^\dagger}) = \rho$. They were originally constructed to aid in proofs of entanglement distillability criteria [44]. Later, together with Werner states, they were used to show that entanglement, EPR steering and Bell nonlocality form a strict hierarchy [15, 45], and they have likewise been useful in a multitude of

benchmarks. They can be described through a single real, linear parameter α by

$$\chi(\alpha) = (1-\alpha)\,\frac{\mathbf{1}}{d^2} + \alpha\left|\Phi^+\middle\rangle\middle\langle\Phi^+\right|, \tag{1.8}$$

with $|\Phi^+\rangle = \frac{1}{\sqrt{d}}\sum_{i=0}^{d-1}|ii\rangle$, a maximally entangled state. For $d=2$, they are identical to Werner states up to local unitaries, but this is not true for larger dimensions. They are also nonseparable if and only if $\alpha > \frac{1}{d+1}$, and are unsteerable under projective measurements if and only if $\alpha \leq \frac{H_d-1}{d-1}$, where $H_d = \sum_{n=1}^{d} 1/n$ is a truncated harmonic series [15]. Letting α run in $[0,1]$, the isotropic state $\chi(0)$ is the maximally mixed state, and for $\chi(1)$ we have a maximally entangled one.

Maximally entangled states are resources for many informational protocols, and the performance of some protocols can be characterized through the *singlet fraction*, which measures the maximal overlap of a resource ρ with a maximally entangled state. Starting from $|\Phi^+\rangle$, all other maximally entangled states $|\Phi\rangle$ can be reached through local unitaries alone, $|\Phi\rangle = (U_A \otimes U_B)\,|\Phi^+\rangle$, thus the singlet fraction is determined by

$$\zeta(\rho) = \max_{\Phi} \langle\Phi|\rho|\Phi\rangle\,.$$

In particular, the singlet fraction of the isotropic states is

$$\zeta\left[\chi(\alpha)\right] = \alpha + \frac{1-\alpha}{d^2}. \tag{1.9}$$

1.2 Transformations

We now know how to describe static physical systems, which is something that physical systems rarely are. In introductory quantum theory, we learn that the evolution of closed quantum systems is governed by the Schrödinger equation. Its solution dictates that an initial state ρ that transforms to ρ' does so unitarily, following $\rho' = U\rho U^\dagger$. Here, U must be an unitary operator, which means that $UU^\dagger = U^\dagger U = \mathbf{1}$. Given U, we can always find a Hamiltonian H and a suitable interaction time to perform the evolution. Nonetheless, these are not the most general transformations that a quantum state can undergo.

Taking the operational approach, we will define the most general transformation as any one that takes a density operator into another, i.e., any $\mathcal{N} : \mathcal{D}(\mathcal{H}) \mapsto \mathcal{D}(\mathcal{H}')$, and look to what properties this implies. First of all, $\mathcal{N}$ must be linear, and the reasoning for that goes as usual: if we mix $\rho, \sigma \in \mathcal{H}$ as $w\rho + (1-w)\sigma$ then put it through $\mathcal{N}$, we surely expect the resulting state to be equivalent to independently passing ρ and σ through $\mathcal{N}$ and then mixing. Equation-wise, this means that $\mathcal{N}\left[w\rho + (1-w)\sigma\right] = w\mathcal{N}(\rho) + (1-w)\mathcal{N}(\sigma)$. Additionally, it is easy to argue that $\mathcal{N}$ must be such that, for any $\rho \in \mathcal{D}(\mathcal{H})$, it is trace-preserving, $\mathrm{tr}\left[\mathcal{N}(\rho)\right] = 1$, and positive, $\mathcal{N}(\rho) \succeq 0$.

Positivity, however, is not a strong enough condition. Suppose that we add an auxiliary system $\mathcal{H}_B$ of arbitrary dimension to $\mathcal{H}$, thus $\mathcal{H} \mapsto \mathcal{H} \otimes \mathcal{H}_B$, and that $\mathcal{N} \equiv \mathcal{N}_{\mathcal{H}\to\mathcal{H}'} \otimes \mathbf{1}_B$. In some situations like this, a positive $\mathcal{N}_{\mathcal{H}\to\mathcal{H}'}$ does not guarantee that $\mathcal{N}$ will map all input states to positive operators. This is precisely the case of the transposition map previously discussed, where the fact that the partial transposition may generate non-positive operators is used as an entanglement criterion. Amending this requires the stronger condition of *complete positivity* (CP), whereby any CP map $\mathcal{N}$ generates a valid density operator no matter what the dimension of $\mathcal{H}_B$ is. Together, complete positivity and trace preservation are the conditions that define a CPTP map, or *quantum channel* $\mathcal{N}$, which is the most general type of quantum evolution we will consider.

These requirements are easily agreeable, but they are not very practical. What we must now

do is find a computation-friendly representation for CPTP maps. This can be found in the Kraus representation theorem [46], stating that general map $\mathcal{N} : \mathcal{L}(\mathcal{H}) \mapsto \mathcal{L}(\mathcal{H}')$ is CPTP (i.e., a quantum channel) if and only if it has a decomposition

$$\mathcal{N}(O) = \sum_{i=1}^{D} K_i O K_i^\dagger,$$

with $O \in \mathcal{L}(\mathcal{H})$, all $K_i : \mathcal{L}(\mathcal{H}) \mapsto \mathcal{L}(\mathcal{H}')$, and $\sum_{i=1}^{d} = K_i^\dagger K_i = \mathbf{1}_{\mathcal{H}}$. The limit of the sum, D, will be at least 1 (in this case we recover an unitary evolution), and will never need to be larger than $\dim(\mathcal{H})\dim(\mathcal{H}')$.

The Kraus representation is just one of several other convenient representations for CPTP maps [47], of which we will need the one called *Choi-matrix representation*. It comes from an application of the *Choi-Jamiołkowski isomorphism* [48, 49]. This remarkable result states that any quantum channel $\mathcal{N} : \mathcal{L}(\mathcal{H}) \mapsto \mathcal{L}(\mathcal{H}')$ can be uniquely mapped to a bipartite state

$$\rho_{\mathcal{N}} = (\mathbf{1}_{\mathcal{H}} \otimes \mathcal{N})\, \rho_{\Phi^+} \in \mathcal{D}(\mathcal{H}) \otimes \mathcal{D}(\mathcal{H}), \tag{1.10}$$

where $\rho_{\Phi^+} = \sum_{i,j=1}^{d_{\mathcal{H}}} |ii\rangle\langle jj|$ is the density operator of the previously introduced maximally entangled state. Taking a closer look, this is actually saying that the space of CPTP maps to the space of bipartite quantum states, and the way to do it is to apply the $\mathcal{N}$ in question to half of that maximally entangled state, while doing nothing to the second part.

To finish the isomorphism, we must also know how to go back, which is done through

$$\mathcal{N}_\rho = \mathrm{tr}_{\mathcal{H}} \left[\rho_{\mathcal{N}} \left(\rho^{\mathsf{T}} \otimes \mathbf{1}_{\mathcal{H}'}\right)\right]. \tag{1.11}$$

This inverse mapping shows us that we can also take a bipartite state and turn it into a quantum channel. Together, eqs. (1.10) and (1.11) consist in the so-called *channel-state duality*.

The representation of quantum channels as bipartite states will become handy when dealing with optimization problems over channels, where a CPTP constraint can be cast as a positivity condition (see secs. 2.2.2 and 5.3).

1.3 Measurements

To finish our review of quantum theory, we must remember how to measure a state. While in classical mechanics we usually gloss over the concept of measurements, taking for granted that any physical property of a state is trivially accessible, in quantum theory we must not. On a par with entanglement, measurement incompatibility is a concept at the heart of the most interesting quantum phenomena, especially those with no classical counterpart. It is unavoidably linked to some of the most intriguing consequences of quantum theory. Decision problems on the Einstein-Podolsky-Rosen steering scenario, for one, can be one-to-one mapped to joint-measurability problems, in the sense that a measurement set is incompatible if and only if it can be used to demonstrate steering [18, 50, 19]. In the also widely studied Bell nonlocality scenario, generating nonlocal statistics require incompatible measurements, but not every set of incompatible measurements is sufficient to observe nonlocality [51, 52, 53]. In this section I review the main mathematical structures related to quantum measurements, together with some aspects of measurement incompatibility.

* * *

Any property of a quantum system ρ must be assessed by measuring it. Abstractly, a measurement procedure takes a quantum input (i.e., the state) and returns a classical output (the measurement result). Inside certain constraints, which measurement is chosen for a given application depends on the property to be measured (e.g., the $\mathbf{z}$ component of a spin 1/2 particle), and the number and values of the possible outcomes are associated to the choice of measurement (e.g., either $\pm\hbar/2$ for the spin 1/2 measurement). A quantum measurement procedure is inherently probabilistic: quantum theory goes only so far as telling us how to ascribe probabilities to each possible outcome. Continuing with the spin example, this means that the quantum formalism will only tell us the probability of getting either the $\pm\hbar/2$ result. When the measurement is actually performed, we may end up with any outcome predicted to have non-zero probability of happening. Depending on this outcome (or rather, on our knowledge of it), the very state ρ that was measured is changed in a non-reversible way. This non-reversibility implies that we cannot fully access ρ with a single copy of the system, a fact that is at the heart of several quantum informational protocols. Although the century-long debate on the nature of quantum measurements is certainly interesting, we will take the pragmatic route and focus on its operational definition.

A positive semidefinite (PSD) operator is a Hermitian operator $E : \mathcal{H} \mapsto \mathcal{H}$ such that $\langle\psi|E|\psi\rangle \geq 0$, $\forall\,|\psi\rangle \in \mathcal{H}$. In this case, we denote $E \succeq 0$. We require a quantum measurement to be described by a set $M = \{E_m\}_m$ of PSD operators obeying the completeness relation $\sum_m E_m = \mathbf{1}$. The conditions $E_m \succeq 0$ and $\sum_m E_m$ can be interpreted as enforcing that $p(m) \geq 0$, $\forall m$ and $\sum_m p(m) = 1$, for any possible ρ. Any such set M is called a *positive operator-valued measure* (or POVM; also called *unsharp measurement*), and each of its elements a *measurement effect*. The possible outcomes are labeled by m. Whenever a measurement M is performed on a state ρ, we get result m with probability $p(m) = \mathrm{tr}\,(E_m\rho)$. When $\rho = |\psi\rangle\langle\psi|$, this definition recovers the Born rule in its usual form, $p(m) = \langle\psi|E_m|\psi\rangle$.

A special case of POVMs arise when every $E_m = \Pi_m$, and $\Pi_m\Pi_n = \Pi_m\delta_{mn}$, where the Π_m are projection operators. It then follows that $1 \leq |M| \leq d$. This is the case of *projective measurements* (or PVMs; commonly also called *ideal*, or *sharp*, or *von Neumann* measurements). Quantum mechanics courses usually introduce projective measurements through the concept of *observables*, which are Hermitian operators. Recalling these can be decomposed as $A = \sum_m m\Pi_m$, we can see they define a PVM where the possible outcomes are the eigenvalues m of observable A, and the projection operators are a set of orthonormal eigenvectors.

A further restriction on measurements is sometimes put on the rank of each effect. A rank-1 projective measurement happens when M is projective and $|M| = d$ or, equivalently, when the associated observable A has no degenerate eigenvalues.

An useful intuition on POVMs is to interpret them as noisy projective measurements. Consider a sharp measurement $M' = \{|0\rangle\langle 0|\,, |1\rangle\langle 1|\}$. Performed on an arbitrary state ρ, we will get the result associated to the the i-th projection with probability $p(i) = \mathrm{tr}\,(|i\rangle\langle i|\,\rho)$, where $i \in \{0, 1\}$. Suppose, though, that our experimental apparatus is such that it states the wrong result with probability $(1 - w)$. Then $p(i) = w\,|i\rangle\langle i| + (1 - w)\,|i \oplus 1\rangle\langle i \oplus 1|$, and a measurement describing this situation is $M = \{w\,|0\rangle\langle 0| + (1 - w)\,|1\rangle\langle 1|\,, w\,|1\rangle\langle 1| + (1 - w)\,|0\rangle\langle 0|\}$. It is easy to see that M's effects are not orthogonal, hence this is not a projective measurement. But it is a valid unsharp measurement.

After a measurement is performed, we may be interested in what happened to ρ. As already stated, performing a measurement will generally incur in a mapping $\rho \mapsto \rho'$. Lüders rule states that, for a PVM $\{\Pi_m\}_m$ returning result m,

$$\rho' = \frac{\Pi_m\rho\Pi_m}{\mathrm{tr}\,(\rho\Pi_m)}. \tag{1.12}$$

Two interesting facts are that knowing the updated state requires knowledge of the measurement result, and that retaking the same measurement will produce the same outcome with definiteness. Without access to the measurement result, but knowing that $\{\Pi_i\}_i$ was performed, all we can say is that $\rho' = \sum_m \Pi_m \rho \Pi_m$. POVMs results are not always reproducible, and the update rule only exists when all effects can be written as $E_m = K_m^\dagger K_m$, where the K_m are Kraus operators. A post-measurement state is then given by

$$\rho' = \frac{K_m \rho K_m^\dagger}{\mathrm{tr}\,(\rho E_m)}.$$

One of the most intriguing aspects of the quantum measurement formalism is that it implies the existence of quantities that may not be simultaneously measured with arbitrary precision on a single copy of a quantum system. Measurements behaving in this way are called *incompatible measurements*. Throughout this work, we will understand "compatibility" as a synonym of "joint measurability". A set $\mathcal{M} = \{E_m^x\}_{m,x}$ of X quantum measurements indexed by x with $\mathcal{O}(x)$ outcomes each is said to be *jointly measurable* whenever a parent measurement J_ℓ exists. To be a parent measurement, J_ℓ must be a valid quantum measurement from which any $E_m^x \in \mathcal{M}$ may be recovered. Letting $\ell = \ell_1 \ell_2 \ldots \ell_X$, where each $\ell_i \in \{1, \ldots, \mathcal{O}(i)\}$, the latter condition requires that $\sum_\ell J_\ell \delta_{m,\ell_x} = E_m^x$. We interpret this as saying that the measurement statistics that the set $\mathcal{M}$ could generate can be reproduced by applying the single measurement J_ℓ then coarse-graining the results. In the plenty of situations where no parent measurement exist, $\mathcal{M}$ is called incompatible or non-jointly measurable. We will get back to this topic in sec. 2.2.2.1, where we discuss an approach to quantify how incompatible a measurement is.

Several other notions of measurement compatibility exist. Introductory quantum mechanics courses, for instance, usually discuss incompatibility as commutativity: two non-commuting observables A and B can only be simultaneously determined up to a degree of certainty, and a trade-off between the standard deviations on the obtained results is given by Robertson's uncertainty relation $\sigma_A \sigma_B \geq \frac{1}{2}|\langle [A, B] \rangle|$. Notably, in the particular case of sharp measurements, commutativity is equivalent to joint-measurability. Regarding POVMs, though, this is not the case, as with several other inequivalent definitions of incompatibility, such as non-disturbance and coexistence [54].

From now on, let us call $\mathcal{P}(d, n)$ and $\mathbb{P}(d, n)$ the set of POVMs and projective measurements, respectively, with n effects acting on $\mathcal{H}^d$. It is clear that $\mathbb{P}(d, n) \subset \mathcal{P}(d, n)$ for any d and n. Now suppose we want to realize an experiment in which we must perform some $M \in \mathcal{P}(d, n)$ but we only have access to a subset $\mathcal{M} \subset \mathcal{P}(d, n)$ where $M \notin \mathcal{M}$. Could we, somehow, reproduce the results we would obtain with M by only using $\mathcal{M}$ and classical processing? In many cases, the answer is *yes* [55, 56, 57].

As an example, in chap. 4 we will be mostly interested in when the some set of POVMs can be simulated using only projective measurements. A good starting point is to notice that the trivial measurement $M = \{\mathbf{1}_d\}_{i=1}^n \subset \mathcal{P}(d, n)$ can be simulated solely by means of classical post-processing. We can simply sample a result from the uniform distribution on $\{1, \ldots, n\}$. Defining the *depolarizing map* as

$$\Phi_t(O) \mapsto tO + (1 - t)\frac{\mathrm{tr}\,(O)}{d}\mathbf{1}, \tag{1.13}$$

and its action on a measurement as $\Phi_t(M) \equiv \{\Phi_t(E_m)\}_m$, we can see that any $M \in \mathcal{P}(d, n)$, when fully depolarized, is simulatable with classical randomness, hence with projective measurements. The question that now arises is whether we need to go all the way through, or if there is some $t > 0$ that suffices. Astonishingly, a non-trivial, general lower bound of $t = 1/d$ exists, and it turns the whole set $\mathcal{P}(d, n)$, for any number n of effects, simulatable with projective measurements [58].

This bound can be improved for specific cases using optimization techniques, and it is known that, for qubits, $t = \sqrt{2/3} - \epsilon$, for some small ϵ, suffices [55].

1.4 Gell-Mann operators and Bloch vectors

Before ending this chapter, we must discuss an interesting representation for quantum states (and some measurement operators) in terms of vectors instead of operators. This will be central to our discussions in chap. 4.

In sec. 1.1, we argued that any quantum state is a linear, positive semi-definite, unit-trace d-dimensional density operator ρ, and that any such ρ is a quantum state. Thus we defined the set of density operators as

$$\mathcal{D}(d) = \{\rho \in \mathcal{L}(\mathcal{H}^d) \mid \mathrm{tr}\rho = 1, \rho \succeq 0\},$$

where $\mathcal{L}(\mathcal{H}^d)$ is the set of linear operators in $\mathcal{H}^d$. This definition implies that $\rho^\dagger = \rho$ and $\mathrm{tr}(\rho^2) \leq 1$, with equality only holding for pure states.

When describing a quantum state, it is often convenient to choose some basis for $\mathcal{H}^d$ and define ρ through a vector. A widely used choice of basis in $\mathcal{H}^2$ are the Pauli matrices $\{\sigma_x, \sigma_y, \sigma_z\}$ (complemented with the identity operator). The Pauli matrices are Hermitian, traceless operators that, together with the identity, form a basis in which any 2×2 Hermitian matrix can be written. Letting $v = (v_x, v_y, v_z) \in \mathbb{R}^3$ and $\sigma = (\sigma_x, \sigma_y, \sigma_z)$, it's easy to see that

$$\rho = \frac{\mathbf{1}_2}{2} + v \cdot \sigma$$

is any unit-trace Hermitian operator. Restricting to $\rho \succeq 0$ further requires that $|v| \leq 1$. Thus, any element of the ball $\mathcal{B}(2) = \{x \in \mathbb{R}^3 \mid |x| \leq 1\}$ is a 2-dimensional quantum state and, conversely, any 2-dimensional quantum state can be associated to a 3-dimensional real vector v with $|v| \leq 1$. Also importantly, $|v| = 1$ if and only if the state is pure, as can be verified by constraining $\mathrm{tr}(\rho^2) = 1$. We'll call $\mathcal{B}(2)$ the *Bloch ball*, its surface the *Bloch sphere*, and any $v \in \mathcal{B}(2)$ a *Bloch vector*. This geometric interpretation of qubit states will prove to be incredibly convenient.

Generalizing these concepts to $\mathcal{H}^d$ can be done in a similar fashion, but will lead us to an important caveat. Our starting point will be picking a basis. While there are a bunch of useful choices [59], the most adequate for our future endeavors will be the generalized Gell-Mann matrices. The (standard) Gell-Mann matrices naturally extend the Pauli matrices from SU(2) to SU(3), and they are likewise traceless and Hermitian. As these are the most important properties we will want to preserve in our applications, it is sensible to choose the standard SU(d) generators when moving on to $\mathcal{H}^d$. Apart from the identity operator, we will need another $d^2 - 1$ operators to span $\mathcal{H}^d$. The *generalized* Gell-Mann matrices match our intents, and can be conveniently written as $\sigma = \{\sigma^{(s)}_{jk}, \sigma^{(a)}_{jk}, \sigma^{(d)}_l\}$, where

$$\begin{aligned}
\sigma^{(s)}_{jk} &= |j\rangle\langle k| + |k\rangle\langle j|, && 1 \leq j < k \leq d, \\
\sigma^{(a)}_{jk} &= -i|j\rangle\langle k| + i|k\rangle\langle j|, && 1 \leq j < k \leq d, \\
\sigma^{(d)}_l &= \sqrt{\frac{2}{l(l+1)}} \left(\sum_{j=1}^{l} |j\rangle\langle j| - l|l+1\rangle\langle l+1| \right), && 1 \leq l \leq d-1.
\end{aligned}$$

Representing an arbitrary element of $\mathcal{H}^d$ asks for d^2 coefficients. However, because we fix $\mathrm{tr}(\rho) = 1$, there's a redundancy that will leave us with only $d^2 - 1$ free parameters. Preserving the

previous notation we then propose that

$$\rho = \frac{\mathbf{1}_d}{d} + v \cdot \sigma = \frac{\mathbf{1}_d}{d} + \sum_{i=1}^{d^2-1} v_i \sigma_i$$

where with the summation we made it explicit that now $v \in \mathbb{R}^{d^2-1}$, and by σ_i we mean the elements of σ defined above. This is, again, clearly a unit-trace and Hermitian operator. The condition $\mathrm{tr}(\rho^2) \leq 1$ further imposes that $r_d = |v| \leq \sqrt{\frac{d-1}{2d}}$, called the *Bloch radius*. It is possible, and useful, to normalize any-dimensional Bloch vectors to unity and write

$$\rho = \frac{\mathbf{1}_d}{d} + r_d^{-1} v \cdot \sigma \tag{1.14}$$

instead. This it the parameterization we will use throughout chap. 4.

Our next step would be to characterize $\mathcal{B}(d)$ and consequently the (generalized) Bloch vectors. Unfortunately, finding a structure on v that guarantees that $\rho \succeq 0$ is not as straightforward as in the 2-dimensional case. Naively trying to associate any vector in a d-dimensional ball with a density operator won't take us far, as we'll soon find out that many choices lead to operators having negative eigenvalues. Albeit the complete characterization of Bloch vectors for arbitrary dimensions has been done [60], it does not lead to a natural parameterization such as in the 2-dimensional case. We will keep calling any $\mathcal{B}(d)$ a Bloch "ball", but it is important to keep in mind that are not balls at all, for there will in general be holes where there are no allowed Bloch vectors.

Even being more complex than for qubits, this geometric interpretation is still fruitful. For one, its inverse map has quite an intuitive interpretation inside of quantum theory. We have so far been interested in defining which Bloch vectors lead to a density operator, but haven't mentioned the converse problem — the one of determining v from ρ — at all. The answer, which can be straightforwardly verified, is that $v_i = \mathrm{tr}(\sigma_i \rho)$. Hence, each component of v is the expectation value of an observable σ_i that can, at least in principle, be experimentally obtained. Luckily, this is the mapping we will usually worry about.

Another useful consequence of this geometric description is that the effect of operations on a state can be interpreted visually. An specific example that will later become important is that of depolarizing channels. These channels are worst-case scenario noise models describing the situation where information on a state ρ is, with some probability p, completely lost. Thus $\rho \mapsto (1-p)\rho + p\frac{\mathbf{1}_d}{d}$, where $\mathbf{1}_d/d$ is the maximally mixed state (the noise). Letting v_ρ be the Bloch vector associated to ρ, and observing that $v = 0$ defines the maximally mixed state, we are led to the conclusion that a depolarizing channel shrinks ρ's Bloch vector towards the origin of the Bloch sphere.

Considering that measurement operators share similarities with density matrices, we may attempt to extend this representation. Our first observation is that not all measurement effects can be written as in eq. 1.4, because they are not required to have unit-trace. However, as any measurement effect is a positive-semidefinite operator, those of which do have unit-trace can also be described through Bloch vectors. As already pointed out, projective measurements are an important class of measurements where every effect is a projection, which is furthermore orthogonal to all the others. For any projection Π, it is true that $\mathrm{tr}(\Pi) = \mathrm{rank}(\Pi)$. Therefore, all rank-1 projective measurements can be described by a set of Bloch vectors, each associated to one of its effects. As all projections with larger rank can be simulated by rank-1 projections and coarse-graining, we conclude that all projective measurements can be interpreted through Bloch vectors together with post-processing. Because, by definition, $\Pi^2 = \Pi$, then $\mathrm{tr}(\Pi^2) = \mathrm{tr}(\Pi) = 1$, which means that a rank-1 projection operator's Bloch vector is actually on the Bloch sphere, analogously to

pure states. Finally, as a measurement's effects must sum to the identity, a nice interpretation of projective measurements on qubits arises: given one of the effect's Bloch vector, the other must be its antipodal.

* * *

Quantum information is the study of the encoding, processing and decoding of information in quantum systems, which are done by preparing a state ρ, transforming, and measuring it. In this chapter we learned that quantum theory tells us we should associate these processes to density operators, CPTP maps, and complete collections of positive-semidefinite operators, respectively. In many cases, we will be interested in quantities that are very nicely described by these structures, but that are nevertheless not easily computable. Luckily, several well-known and efficient optimization techniques are well suited to dealing with quantum theory. These techniques will be central to chap. 4 and sec. 5.3, so we now review them.

Chapter 2

Convexity and optimization

Catenaries, Fermat's principle of least time and thermodynamic equilibrium are well known examples of extremum principles applied to the description of nature. On the opposite side, multiple essential modern technologies, such as network routing, logistics and integrated circuit design rely on finding good solutions to optimization problems. No matter whether you are minimizing a Gibbs free energy under constant T and P, or searching for the best arrangement of electronic components on a chip to reduce its footprint while satisfying heat dissipation and fabrication constraints, this is what an optimization problem looks like

$$p^* = \max\big\{f_0(x) \mid f_i(x) \leq b_i,\, \forall i \in \{1,\dots,m\}\big\}. \tag{2.1}$$

Solving it means finding some $x^* \in \mathbb{R}^n$ such that the value of the *objective function* $f_0(x^*) \leq f_0(x)$, $\forall x \in \mathcal{F}$, where

$$\mathcal{F} = \big\{x \mid f_i(x) \leq b_i,\, \forall i \in \{1,\dots,m\}\big\}$$

is the *feasible set*. The objective function represents the quantity to be minimized (the free energy; the footprint), while $\mathcal{F}$ imposes the constraints (constant temperature and pressure; rate of heat dissipation and limitations of the fabrication procedure).

Finding global optima for nonlinear programs is difficult, and no general methods to do so are known. That is why one usually consider subsets of eq. 2.1 where some special stucture is imposed on the f_i. Our goal for this chapter is to learn how to recognize the special structures of linear and semidefinite programs, which requires some groundwork on convexity. Furthermore, polytopes — a special type of convex sets —, are ubiquitous in the geometry of correlation scenarios, and will make an appearance in later chapters. This will be our starting point.

2.1 Convexity

This subsection closely follows the expositions in [61, 62, 63, 64], where most of the alluded proofs can be found. Unless otherwise specified, all sets are subsets of $\mathbb{R}^d$.

Let $x_1 \neq x_2$ be two elements in $\mathbb{R}^d$. We define

$$\alpha x_1 + (1-\alpha)x_2, \quad \alpha \in \mathbb{R}$$

as the *line* passing through x_1 and x_2. A set A containing all lines between pairs of its elements is an *affine set*. The real line and the Cartesian plane are affine sets, but a sphere and a cube are not.

Finite induction on the definition shows that, for $x_1, \ldots, x_n \in A$ and $\sum_i \alpha_i = 1$, all points $x = \sum_i \alpha_i x_i \in A$. Such a sum is an *affine combination* of the x_i.

When A_i are affine sets, $A = \bigcap_i A_i$ is also affine. Intersections can only decrease cardinality. Hence, if we take some (not necessarily affine) set S and intersect all possible $A_i \supseteq S$, we will get the smallest affine set containing S. This is called an *affine hull*, and denoted $\operatorname{aff}(S)$. Equivalently, it can be shown that

$$\operatorname{aff}(S) = \Big\{ \sum_i \alpha_i x_i \mid \sum_i \alpha_i = 1,\ x_i \in S \Big\}.$$

The affine hull of two points in $\mathbb{R}^n$ is the line through them, and for three noncollinear points we end up with a plane.

Vector spaces and affine sets are close siblings: any vector space is an affine set, but the converse is not true, for the latter may not contain the 0 vector. Furthermore, each affine set A is parallel to a unique vector space V. Taking any $x_0 \in A$, we can translate $A - x_0 = V$. The *dimension* of an affine space A is the dimension of its parallel vector space V. In $\mathbb{R}^n$, dimensions $0, 1, 2$ and $n-1$ corresponds to points, lines, planes and hyperplanes.

Any hyperplane H can be represented as the set

$$H = \{u \mid \langle u | b \rangle = \beta\},$$

where $b \in V$, $\beta \in \mathbb{R}$ are constants, and $\langle \cdot | \cdot \rangle$ is an inner product on V. Switching from equality to $<, >, \leq$ or $\geq$, we get either open or closed *halfspaces*. Halfspaces contain halflines, so they are not affine sets. Rather, they are convex sets.

Convex sets are somewhat similar to affine sets. But, instead of lines, a convex set C must contain all *line segments*

$$\alpha x_1 + (1 - \alpha) x_2, \quad \alpha \in [0, 1]$$

passing through any $x_1, x_2 \in C$. Any affine set is trivially convex, but a sphere and a cube also are.

Convex combinations are affine combinations with the extra condition that the $\alpha_i \geq 0$. Likewise, it can be shown that a set C is convex if and only if it contains all convex combinations of its elements. They are also closed under intersections, and the convex hull of a (not necessarily convex) set S is the smallest convex set containing S,

$$\operatorname{conv}(S) \equiv \bigcap \big\{ C \supseteq S \mid C \text{ is convex} \big\}.$$

Equivalently, it is also the set of all convex combinations of S's elements,

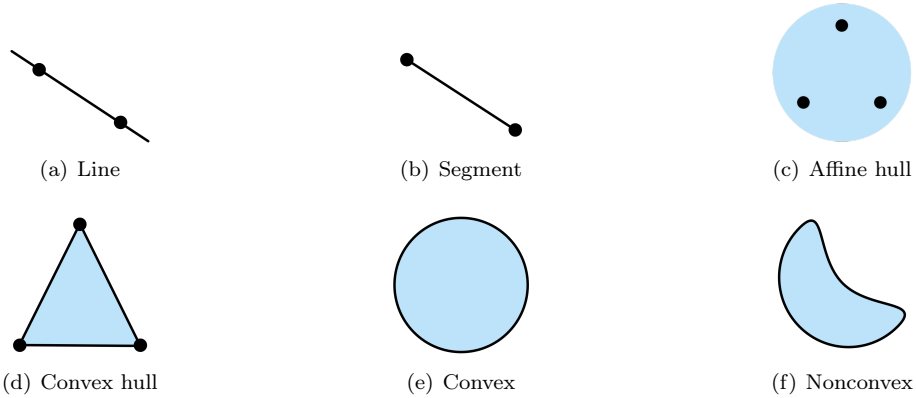
$$\operatorname{conv}(S) = \Big\{ \sum_i \alpha_i x_i \mid \alpha_i \geq 0, \sum_i \alpha_i = 1,\ x_i \in S \Big\}.$$

Convex sets can be further divided into extremal and nonextremal points. An $x \in C$ is an *extreme point* of C when it is not in the relative interior of any segment of C; respectively, when $x = \alpha y + (1 - \alpha) z \Rightarrow x = y = z$ for any $y, z \in C$ and $\alpha \in (0, 1)$.

Polyhedra are intersections of *finitely* many closed halfspaces. Any polyhedron $L \subset \mathbb{R}^d$ can thus be written as

$$L(A, b) = \big\{ x \in \mathbb{R}^d \mid Ax \leq b \big\},$$

where $A \in \mathbb{R}^{d \times m}$ and $b \in \mathbb{R}^m$ specify a set of linear inequalities. A single halfspace, which is obviously unbounded, fits the definition. If a polyhedron L is furthermore bounded (i.e., has no



(a) Line (b) Segment (c) Affine hull

(d) Convex hull (e) Convex (f) Nonconvex

Figure 2.1: Pictorial representation of the definitions introduced in sec. 2.1. Fig. 2.1(a) shows a line, which is an affine combination of two points. In 2.1(b), a line segment (or a convex hull of two points) is shown. Two affine hull of three noncollinear points is the plane (2.1(c)), while the convex hull of a finite number of points defines a polytope (2.1(d)). Any polytope is a polyhedron, but the converse is not true. Fig. 2.1(e) shows a convex set which is not a polytope. Is has an infinite number of extremal points. Nonconvex sets do not contain every line segment between its elements (2.1(f)).

rays), it is a convex *polytope*, denoted by P. A convex hull of finitely many points is thus a polytope. Polyhedrons and polytopes inherit their dimensions from their affine hull's. From the definition follows that one way to test if an $x \in P$ (or L) is to check whether it does not violate any of the inequalities defining the halfspaces. This is closely linked to nonclassicality witnesses, to be further discussed in chap. 3.

As much as cubes have vertices, edges and faces, polytopes have k-faces. Any *face* F of a polytope P is a set

$$F = P \cap \{x \in \mathbb{R}^d \mid c^\intercal x = b\},$$

where $c \in \mathbb{R}^d$ is a column-vector, $b \in \mathbb{R}$, and any $x \in P$ is such that $c^\intercal x \leq b$. Visually, a face is any intersection of P with a closed halfspace that touches P but is not inside of it. The dimension k of the affine hull of F goes into the name k-face. Faces of dimension 0 and 1 are vertices and edges, while faces of dimension $\dim(P) - 1$ are especially named *facets*. For a cube ($\mathbb{R}^3$), the facets would match the commonly called faces.

Our previous definitions of polyhedra and polytopes are sometimes termed $\mathcal{H}$-polyhedra and $\mathcal{H}$-polytopes. An alternative that will later become important is that of $\mathcal{V}$-polytopes, which are defined as the the convex hull of a *finite* set of points. After the convex hull is performed, they become the extremal points of the convex set. The $\mathcal{V}$-description of a unit square is $\mathcal{V} = \{(0,0),(1,0),(0,1),(1,1)\}$, and after taking $\mathrm{conv}\,(\mathcal{V})$ we get all points satisfying the $\mathcal{H}$-description $Ax \leq b$ defined by

$$A = \begin{pmatrix} -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}.$$

$\mathcal{H}$ and $\mathcal{V}$ representations can be proven mathematically equivalent (theorem 1.1 in [63]) but, computationally, the choice of description may significantly matter. As will be made explicit in chaps. 3 and 4, classicality in prepare-and-measure scenarios can be interpreted as the belonging of a behavior in a polytope. Conversely, nonclassicality may be witnessed through the violation of some inequality defining its facets. The $\mathcal{H}$-description of these polytopes is thence preferable for

being more ergonomic. Finding these descriptions is important not only for prepare-and-measure scenarios, but are rather important in several other correlation scenarios. In Bell nonlocality scenarios, for instance, they are the celebrated Bell inequalities. Despite theoretical and practical importance, no general methods to directly build $\mathcal{H}$-descriptions of these polytopes are known. On the other hand, it is conceptually easy to enumerate all extremal points of classicality polytopes (the $\mathcal{V}$-description). Casting one to the other (i.e., performing the vertex or facet enumeration) can be done through specialized, user-friendly softwares such as `PANDA`, `lrs` and `cdd` [65, 66, 67] but, unfortunately, this a computationally expensive problem that can only be done for the very simplest cases of interest. The conceptual significance of $\mathcal{H}$-descriptions will be further explored in chap. 3, and an example of description conversion will be discussed in chap. 4. Results therein were obtained with `PANDA`, and descriptions of the algorithms and user guides can be found in the aforementioned references.

A second important fact involving polyhedra and polytopes is that linear functions can be efficiently optimized inside them through so-called linear programming. In the remainder of this chapter, we go back to the problem posed in the introduction, and briefly discuss two efficiently computable instances of optimization problems; both of which are widely used in quantum information.

2.2 Optimization

Solving an arbitrary instance of eq. (2.1) is a hard problem, and general methods only exist for special instances of optimization programs. Convex optimization (`CONV`) — which happens when all the f_i are convex functions — is one of the largest classes of programs that can be efficiently solved to global optimality. Even though `LIN` $\subset$ `SDP` $\subset$ `CONE` $\subset$ `CONV`, specialized algorithms for subsets of convex optimization, such as conic (`CONE`), semidefinite (`SDP`) and linear (`LIN`) programming, renders even larger instances of those practical. Whilst we will not discuss them, the main reference for convex programming is the textbook [64], and a good reference for cone programming, tuned to quantum information, is [68].

Because of their use in the industry, ready-to-use solvers for optimization problems are broadly accessible. I will mention them in due time, but algorithms will not be introduced in depth. For thorough discussions, see, e.g., [69, 70] for linear programming, [71, 72] for SDPs and [64] for convex optimization.

2.2.1 Linear programming

Linear programming happens when all the f_i are linear functions. Recalling eq. (2.1), you will see that the feasible region $\mathcal{F}$ is thence an intersection of halfspaces — a polyhedron. Apart from being widely used in the industry to optimize supply chains, workforce allocation and delivery routes, it is also useful to search for probability distributions satifying a set of constraints. The significance of this problem will become clearer in chap. 4.

Being linear, the objective function $f_0 : \mathbb{R}^n \mapsto \mathbb{R}$ can conveniently be written as $c^\intercal x$, where $c = (c_1, \ldots, c_n)$ is a constant vector in $\mathbb{R}^n$ defining the quantity to be maximized. Minimizing $c^\intercal x$ is the same as maximizing $-c^\intercal x$, so the discussion holds both ways. For an *unconstrained* problem, there is not much more to say: just follow along the direction of the gradient. In this case, though, unless f_0 is constant, the program would end up being *unbounded*. Things get more interesting when x is constrained to be in a subset $\mathcal{F} \subset \mathbb{R}^n$. We can do that by saying $0 \leq x_i \leq 1$, $\forall i$, for instance, and then we would be optimizing $c^\intercal x$ over some hypercube. More generally, in linear

programming we allow the constraints to be of the form $a_i^\intercal x \leq b_i$, where $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$. There is no need to consider equality conditions separately, as we can just use two inequalities with inverted directions. So we do not have to write all constraints separately, a shorthand notation is to build an $m \times n$ matrix $A = [a_1 \ \dots \ a_m]^\intercal$ and a vector $b = [b_1 \ \dots \ b_m]^\intercal$ of bounds so that

$$\text{given} \quad c, A, b \tag{2.2a}$$

$$\max_x \quad c^\intercal x \tag{2.2b}$$

$$\text{s.t.} \quad Ax \leq b \tag{2.2c}$$

is a general form of a linear program (LP). When $\mathcal{F} = \emptyset$, the program is *unfeasible*, meaning there are no vector x that satisfies the constraints. Requiring that $x_i \geq 1$ and $x_i \leq -1$ would certainly put you in that situation. If $\mathcal{F}$ is not empty, then the program may either have a solution or be unbounded. To understand why the latter may be the case, notice that $\mathcal{F}$ is a polyhedron. As such, it may itself be unbounded. Consider, for instance, $x \in \mathbb{R}^2$ with $\mathcal{F} = \{0 \leq x_1 \leq 1, x_2 \geq 0\}$. If $c = [c_1 \ c_2]$, with $c_2 > 0$, the optimal value p^* would go to infinity. On the other hand, $c_2 \leq 0$ would be alright even though $\mathcal{F}$ is unbounded. When $\mathcal{F}$ is a (nonempty) polytope, there is always some solution p^* to the program. A solution is either unique (only a single x^* leads to p^*) or there infinitely many solutions yielding the same p^*. To understand why this is so, consider the *fundamental theorem of linear programming*, which states that every feasible, bounded linear program has an optimal solution on a vertex of $\mathcal{F}$. If optimal solutions happen at two different vertices, then any x in the line segment between them results in p^*. Otherwise, there would either be a larger feasible value in the segment, or the feasible region would not be convex.

Linear programs are efficiently solvable in practice. The simplex method, proposed by George Dantzig in 1947, builds upon the fact that optimal solutions occur in vertices. Starting from any basic feasible solution (a corner of $\mathcal{F}$), it smartly hops to neighboring vertices along edges that increase f_0. If you end up at a basic feasible solution connected to no edges that increase the objective value, or if you visit an unbounded edge, you are done. A rigorous presentation and complexity analysis of it can be found in [69], chap. 2. Interestingly, the simplex method is *not* of polynomial time complexity (i.e., it is not efficient *in theory*). Rather, there are families of linear programs for which it performs poorly. Nevertheless, it is a reliable and efficient method in practice, and is widely used (in several variations) up to this day. There are provably polynomial time algorithms in theory, of which the ellipsoid method — invented in 1970 for nonlinear programming, then adapted and proved polynomial for linear programming in 1979 —, was the first. Even so, it is not efficient in practice. The more recent interior-point methods are provably polynomial in theory, and fast in practice. Pedagogical discussions of these methods, together with historical remarks, are nicely presented in chap. 7 of [70].

Due to its wide applicability in the industry, there are many open source and proprietary linear programming solvers available, such as GLPK, Gurobi, and Mosek [73, 74, 75]. Most provide user-friendly interfaces to widely used programming languages, such as C, C++, Python and MATLAB. To further aid in using them, there are also modeling languages that can be used to specify programs in a high-level format, and that internally converts and dispatches the problem to specific solvers. YALMIP [76] and CVX [77] are widely used inside MATLAB, while PICOS [78] and CVXPY [79] are common choices when working in Python.

2.2.2 Semidefinite programming

Semidefinite programming largely generalizes linear programming, and has found numerous applications in statistics, economics, control theory, pattern recognition and machine learning, to mention a few (see sec. 2 of [71] and chap. 1 of [64] for a survey). It is also widely used as a tool in approximation algorithms to graph theoretical problems [72], and polynomial and non-commutative polynomial optimization [80, 81]; the latter being widely used in quantum information.

While linear programming is the optimization of a linear function over a polyhedron, semidefinite programming is the optimization of a linear function over a spectrahedron. Before we further explain it, recall that

$$S^m = \{X \in \mathbb{R}^{m\times m} \mid X = X^\intercal\}$$

is the set of $m \times m$ symmetric matrices, and

$$S^m_+ = \{X \in S^m \mid X \succeq 0\}$$

the one of positive $m \times m$ semidefinite matrices. All eigenvalues of a symmetric ($X = X^\intercal$) matrix are real, and the positive semidefiniteness condition ($X \succeq 0$) further requires they must be nonnegative.

Building upon the general linear program (2.2), we write a semidefinite program (SDP) as

$$\begin{aligned}
&\text{given} && c, F_0, \ldots, F_n && \text{(2.3a)}\\
&\max_x && c^\intercal x && \text{(2.3b)}\\
&\text{s.t.} && F_0 + \sum_{i=1}^n x_i F_i \succeq 0. && \text{(2.3c)}
\end{aligned}$$

Here, $F(x) \equiv F_0 + \sum_{i=1}^n x_i F_i \succeq 0$ is what we call a *linear matrix inequality* (LMI), and it enforces the r.h.s. to be in S^m_+. The $n+1$ matrices $F_0, \ldots, F_n$ are in S^m, and we maintain $x \in \mathbb{R}^n$. A semidefinite program is thus the optimization of a linear function $c^\intercal x$ under the constraint that $F(x)$ is positive semidefinite.

This may seem like a rather arbitrary definition. To debunk this impression, first notice that if $F(x) \succeq 0$ and $F(y) \succeq 0$, then

$$F[\alpha x + (1-\alpha)y] = \alpha F(x) + (1-\alpha)F(y) \succeq 0$$

for all $0 \leq \alpha \leq 1$. Thus, both the objective function and the constraint are convex, and semidefinite programming is a special case of convex optimization. More than that, S^n_+ is a convex cone (i.e., it is closed under conic combinations), meaning it is also a special case of conic optimization. Now it does not look that much arbitrary, but rather it seems too specialized. But it is not, for if we let $F_0 = \operatorname{diag}(b)$ and $F_i = \operatorname{diag}(a_i)$, we recover the form of a linear program just like eq. (2.2). Putting it all together, `LIN` $\subset$ `SDP` $\subset$ `CONE` $\subset$ `CONV`, as stated in the beginning of this section.

A general $F(x)$ may have a block-diagonal section of the form $\operatorname{diag}(Ax+b)$, and a more general structure elsewhere. Thus, a linear matrix inequality represents an affine section of S^n_+, also called an *spectrahedron*. Unlike polyhedra, they may have curved boundaries. A cylinder, for instance, can be parametrized as an $F(x)$ with 4×4 matrices [82]. A last interesting link to linear programming is that $F(x) \succeq 0$ if and only if $z^\intercal F(x) z \geq 0$ for all $z \in \mathbb{R}^m$. An SDP is thus a linear program with infinitely many linear constraints on x.

The usefulness of semidefinite programming in quantum information is hinted by the fact that

quantum states and measurement effects are positive semidefinite matrices. As long as the objective function is linear, we can optimize over them, and they have proven to be useful in quantum state discrimination, quantum steering and hierarchies for nonlocal correlations, among other applications. One of them will be shown in chap. 5. A drawback in dealing with density operators or measurement effects is that, while semidefinite programming is, as in here, usually discussed over the real field, quantum objects make use of complex numbers. One can nevertheless embed them in real variables. This can be a cumbersome task to do manually but, luckily, the modeling interfaces mentioned in the last section can do this under the hood before calling the solver.

Semidefinite programming can also be efficiently solved both in theory and in practice. Interior-point methods are usually robust and efficient, and are implemented in some user-friendly solvers such as SDPA [83] and Mosek [75]. Performance comparisons between these and several other SDP solver are available at [84].

2.2.2.1 Joint measurability robustness

The concept of joint measurability was introduced in sec. 1.3. Following along the same notation, $\mathcal{M} = \{E_b^y\}_{b,y}$ is a set of Y quantum measurements, indexed by y, each having $\mathcal{O}(y)$ outcomes. A parent measurement is called J_ℓ, and only exists if $\sum_\ell J_\ell \delta_{b,\ell_y} = E_b^y$, where $\ell = \ell_1 \ell_2 \dots \ell_Y$, and each $\ell_i \in \{1, \dots, \mathcal{O}(i)\}$. From this requirement, it follows that J_ℓ can extract all information needed to reproduce $\mathcal{M}$. Whenever such a J_ℓ does not exist, $\mathcal{M}$ is incompatible, or not jointly measurable.

More than determining whether an $\mathcal{M}$ is joint measurable, it is also interesting to ask to what extent a measurement is incompatible. Such *incompatibility robustness* must be defined relative to *something*. This something is what we call a noise model. Formally, a *noise model* $\mathbf{N} : \mathcal{P}(d, \mathbf{n}) \mapsto \mathbb{S}(\mathcal{P}(d, \mathbf{n}))$ maps from the set of $Y = |\mathbf{n}|$ d-dimensional POVMs, each with $\mathbf{n} = (n_1, \dots, n_Y)$ outcomes, to the set $\mathbb{S}(\mathcal{P}(d, \mathbf{n}))$ of subsets of $\mathcal{P}(d, \mathbf{n})$. The simplest possible choice for $\mathbf{N}$ is the white noise model

$$\mathbf{N} = \left\{ \left\{ \frac{\mathbf{1}}{n_1} \right\}_{i=1}^{n_1}, \dots, \left\{ \frac{\mathbf{1}}{n_Y} \right\}_{i=1}^{n_Y} \right\},$$

but several other common choices, such as depolarization maps or probabilistic noise also fit the definition. In general, it may also depend on $\mathcal{M}$, but we require that it contains at least one set of jointly measurable measurements. Adequateness is a matter of application, and different choices may lead to different results.

Having the noise model fixed, we define the incompatibility robustness of $\mathcal{M}$, with respect to $\mathbf{N}$, as

$$\chi^*_{\mathcal{M}} = \sup_{\substack{\chi \in [0,1] \\ \{N_{b|y}\} \in \mathbf{N}(\mathcal{M})}} \Big\{ \chi \mid \chi\{E_{b|y}\} + (1-\chi)\{N_{b|y}\} \in \mathbf{JM} \Big\}. \tag{2.4}$$

JM denotes the set of all jointly measurable measurements, and $\chi^*_{\mathcal{M}} = 1$ readily implies $\mathcal{M}$ is jointly measurable. Any lower value means it is not, for to become it must be mixed with noise. More than that, if $\chi^*_{\mathcal{M}} < \chi^*_{\mathcal{M}'}$ then $\mathcal{M}$ is *more* incompatible than $\mathcal{M}'$.

Incompatibility robustness was investigated and applied in several works involving quantum measurements and more general quantum devices, such as [85, 86, 19]. Our presentation follows the very thorough investigation in [86], where analytical and numerical results for several noise models are presented, and further references can be found. The fact that eq. (2.4) can be solved through semidefinite programming is also known. To see why this is so, first notice that, if we take a closed noise set, the supremum in (2.4) becomes a maximum [**Review!** If it's not a maximum then can't solve as SDP? Not quite sure if I understand this]. Furthermore, $f_0 = \chi$, so we are indeed optimizing a linear function. The measurement effects, $\{E_{b|y}\}$ and noise effects $\{N_{b|y}\}$ are given,

and J_ℓ, an optimization variable, is a collection of PSD operators (i.e., the parent measurement's effects). They must be constrained to sum to the identity, and the parent's marginals must recover $\mathcal{M}$. These are semidefinite constraints, as more clearly seen by writing the program

$$\begin{aligned}
&\text{given} && \{E_{b|y}\}, \{N_{b|y}\} && \text{(2.5a)}\\
&\max_{\chi, J_\ell} && \chi && \text{(2.5b)}\\
&\text{s.t.} && \sum_\ell J_\ell \delta_{b,\ell_y} = \chi E_{b|y} + (1-\chi) N_{b|y}, \quad \forall b, y && \text{(2.5c)}\\
& && J_\ell \succeq 0 && \text{(2.5d)}\\
& && \chi \leq 1. && \text{(2.5e)}
\end{aligned}$$

Notice that the constraints $\eta \geq 0$ and $\sum_\ell J_\ell = \mathbf{1}$ are not missing, but rather are automatically enforced by the first constraint and the fact that the noise set must have at least one jointly measurable element, such that 0 will always be a lower bound. We will come back to this program in sec. 4.2.1, where it will be crucial to our proof that measurement incompatibility is not sufficient for nonclassicality in prepare-and-measure scenarios. Before getting to that, we must first learn what prepare-and-measure scenarios actually are.

Part II

Prepare and measure

Chapter 3

The prepare-and-measure scenario

The prepare-and-measure (PM) scenario is one of the simplest and most fundamental examples of correlation scenarios. In it, a preparation apparatus produces and then sends a physical system, over a communication channel, to a measurement device which reads out information from the received state. Wherefore, it is an adequate setting in which to investigate two of the most fundamental building blocks of physical theories: states and measurements.

Differently from the more widely studied Bell nonlocality and EPR steering scenarios, a quantum prepare-and-measure experiment may behave nonclassically even in the absence of entanglement. Quantum behaviors in PM scenarios must then rely on other strictly quantum features, such as measurement incompatibility [87] and non-orthogonality of states [88], but the exact relations are still unknown.

Other than quantum communication, a second resource that also leads to drastically different behaviors is whether the preparation and measurement devices are independent or not. There are, in general, three possible cases; namely, full independence, shared randomness and entanglement assistance. Together with either classical or quantum communication, this will lead to six inequivalent prepare-and-measure configurations.

Prepare-and-measure scenarios are also the simplest correlation scenarios that presume communication, and, as such, should become an indispensable ingredient in quantum networks [89, 90]. As with other correlation scenarios, quantum behaviors in the PM scenario can be exploited to build informational protocols that show advantage over their classical counterparts (sec. 3.2). On a more fundamental aspect, they are at the core of proposed informational principles to quantum theory [91, 92], and of quantum ontologies [14].

In chapters 4 and 5, novel results regarding some of these settings will be presented. To build towards that end, we now discuss these many different instances of preparation and measurement devices, and show how this scenario can be seen as a physical implementation of two paramountly important communication protocols.

3.1 Prepare-and-measure behaviors

The simplest prepare-and-measure setup consists of two black-boxes. One is the preparation device P, handled by Alice, and the other the measurement device M, handled by Bob. Nothing about the inner workings of these devices is assumed *a priori* (see), except that P prepares and communicates a physical system to M, which extracts information from the received preparation by measuring it.

Alice is allowed to interact with her device through a classical input $x \in \mathcal{X} \equiv \{1, \ldots, X\} \equiv [X]$. Her choice may weight on the probability $p(m \mid x)$ with which a state labeled by $m \in \mathcal{S}$ is prepared.

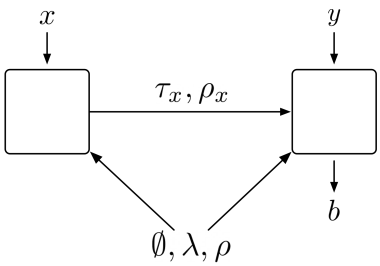


Figure 3.1: Prepare-and-measure scenarios. The measurement box, upon receiving an input x, prepares either a classical, τ_x, or quantum ρ_x, state to the measurement device. In turn, given a choice y of measurement, an output b is observed. From their causal past, the devices may be independent ($\emptyset$), share classical randomness (λ), or even quantum correlations (ρ).

Here, $\mathcal{S}$ represents the set of possible preparations. This choice can be more intuitively understood as the message Alice aims to encode in her preparation.

Similarly, Bob can choose a configuration $y \in \mathcal{Y} \equiv [Y]$ for the measurement device, which will thence output $b \in \mathcal{B} \equiv [B]$ with probability $p(b \mid m, y)$. This reflects the fact that the outcome of the experiment may be influenced by both the received message and Bob's choice of measurement (fig. 3.1). More intuitively, Y limits how many choices of decoding procedures can be used by Bob to try and recover Alice's message x.

A last, indispensable condition is that neither Alice nor Bob know what happens in each other's labs. In other words, she must have no knowledge of y, as this is a choice made in her causal future, and he cannot know what was her choice x of preparation, as this trivializes the scenario in a sense that will soon be made clearer. Taken together, these constraints amount to saying that all communication between them is mediated by the message m. Given that our intention is to study how this communication influence the observed statistics, this is clearly a natural assumption.

An *round* is a single run of the protocol described above. After many such rounds, Alice and Bob are allowed to share their knowledge with each other, and together they can build the *behavior* $\mathbf{p} = \{p(b \mid x, y)\}_{b,x,y}$ of their devices. This behavior characterizes the prepare-and-measure *experiment*. Naturally, each $p(b \mid x, y) \geq 0$ and $\sum_b p(b \mid x, y) = 1$, $\forall x, y$ or, equivalently, $\mathbf{p}$ is a collection of conditional probability distributions, one for any fixed choice of a pair of settings (x, y).

We now arrive at the central question of (semi)device-independent quantum information: if all we have is the behavior $\mathbf{p}$, with no (or restricted) access to the actual workings of the devices, can we still certify some property about the states, measurements, or other quantities of interest? For instance, could we, by only observing $\mathbf{p}$, affirm that P prepares quantum — as opposed to classical —, states? Or that M applies nonprojective measurements? In many cases, the answer, surprisingly, is *yes*.

3.1.1 Classical preparations

Quantum preparations may behave quite distinctly from classical preparations, and knowing how to tell them apart is of the essence for developing quantum communication protocols. An important open question regards the properties that allow some sets of preparations and measurements to behave nonclassically. In chapter 4, I will discuss the problem of classical simulatability of quantum behaviors in reasonable generality, and prove that measurement incompatibility is not a sufficient condition for nonclassicality in the PM scenario. To make that discussion precise, we must begin by defining what it is that we will call classically-simulatable behaviors or, for shortness, *classical behaviors*.

Starting from the paradigmatic prepare-and-measure scenario, let us further impose that $\mathcal{S}$, the set of possible preparations, contains only *classical states* (dits). Naturally, as this is a communication scenario, the dimension $|\mathcal{S}|$ of the classical system used for encoding the states must be bounded, otherwise communication becomes trivial (i.e., Alice can perfectly encode her message x and Bob can perfectly recover it) and all behaviors are possible.

Our first aim is to investigate the set of behaviors that can be achieved when communicating d-dimensional classical systems, or rather, when $\mathcal{S} = \{0, \ldots, d\}$. They will, in general, also depend on X (the size of Alice's input alphabet) and Y (the number of choices for Bob's measurement). Letting this set of behaviors be $\mathcal{C}_{d,X,Y}$, our previous discussion implies that

$$\mathbf{p} \in \mathcal{C}_{d,X,Y} \quad \iff \quad p\left(b \mid x, y\right) = \sum_{m \in \mathcal{S}} p\left(m \mid x\right) p\left(b \mid m, y\right) \quad \forall b, x, y. \tag{3.1}$$

Essentially, then, $\mathcal{C}_{d,X,Y}$ contains all behaviors that may occur when (i) the devices are *uncorrelated*, (ii) the preparations are classical and (iii) with dimension at most d, (iv) Alice has X preparation choices, and (v) Bob picks one out of Y measurement settings.

To simplify the notation, whenever X and Y are arbitrary or clear by context, the subscripts will be omitted.

Briefly detouring, it is interesting to notice that model (3.1) is very much in the spirit of ontological models [14, 93, 94]. To see that, consider $\mathcal{S}$ as a finite, dimension-bounded ontic state space, and x as the preparation procedure. Then, $p\left(m \mid x\right)$ models our epistemic state. Similarly, if we take y as a choice of measurement procedure, we may interpret $p\left(b \mid m, y\right)$ as the indicator function. In this way, any theory that only produces behaviors $\mathbf{p} \in \mathcal{C}_{d,X,Y}$ admits a dimension-bounded ontological model.

Conditions (iii)–(v) above are pretty much natural for any communication scenario, but the case is different for (i) and (ii). While in many situations, such as when we have some trust on our devices, (i) is justifiable, it is not always safe to assume that the devices are uncorrelated. The worst-case scenario for classical variables is when P and M can share an unbounded amount of pre-established classical correlations. Such correlations must reside in the causal past of the experiment (fig. 3.2), but even so, they can be used to achieve better performance in several communication protocols implemented in the PM scenario [95], and lead to quite different geometrical structures [96, 97]. Without knowledge on what information they share, the best we can do is to call it λ and say that π is some probability distribution over this random variable. As both devices can fully access $\lambda \in \Lambda$,

$$\mathbf{p} \in \mathcal{C}^{\lambda}_{d,X,Y} \quad \iff \quad p\left(b \mid x, y\right) = \int_{\Lambda} \sum_{m \in \mathcal{S}} \pi(\lambda) p\left(m \mid x, \lambda\right) p\left(b \mid m, y, \lambda\right) d\lambda \quad \forall b, x, y. \tag{3.2}$$

Here, $\mathcal{C}^{\lambda}_{d,X,Y}$ is the set of behaviors we get from $\mathcal{C}_{d,X,Y}$ by allowing for shared randomness. As $\pi(\lambda) \geq 0$ and $\sum_{\lambda} \pi(\lambda) = 1$, eq. (3.2) is actually telling us that $\mathcal{C}^{\lambda}_{d,X,Y} = \mathrm{conv}\left(\mathcal{C}_{d,X,Y}\right)$. More than that, a slight variation on Fine's theorem [98] (or sec. 2.3 of [99] for a more pedagogical discussion) can show that the set $\mathcal{C}_{d,X,Y}$ has a finite amount of extremal points, called deterministic strategies [100, 95]. They are the points given by eq. (3.1) when the response functions are deterministic, i.e., when $p\left(m \mid x\right) = \delta_{m,f(x)}$ and $p\left(b \mid m, y\right) = \delta_{b,g(m,y)}$, for some functions $f : \mathcal{X} \to [d]$ and $g : [d] \times \mathcal{Y} \to \mathcal{B}$ that are made precise in the aforementioned references. As $\mathcal{C}^{\lambda}_{d,X,Y}$ can now be seen as a convex hull of finitely many points, this proves that $\mathcal{C}^{\lambda}_{d,X,Y}$ is a polytope. Recalling the discussion in sec. 2.1, we emphasize that $\mathcal{C}^{\lambda}_{d,X,Y}$ can thus be described by an intersection of half-spaces, which are given by the linear inequalities defining its facets. This description will turn out to be especially useful during chap. 4, in which we will get back to this topic and work out an example that should clarify this discussion.

For an example of the usefulness in understanding these sets, notice that for some fixed X and Y, and some $d' > d$, we have the proper inclusion $\mathcal{C}_d \subset \mathcal{C}_{d'}$, which implies the same for the SR case. Ultimately, this means that larger dimensional communication can carry strictly more information.

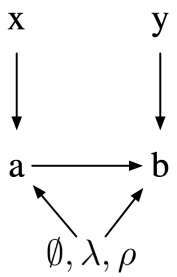


Figure 3.2: Causal structure of a prepare-and-measure experiment. Choice x of preparation can only directly influence the message m, which can be encoded in a classical or quantum system. Communicating m effectively screens-off the choice of x, unless the measurement device can perfectly recover it. Observable b is influenced by the received message m and the choice y of measurement, which can be seen as a decoding function. Both m and b can be further classically or quantumly correlated by some variable in the causal past of the experiment.

Now suppose that W_d is a linear functional defining some facet of $\mathcal{C}_d^\lambda$ which is not a facet of $\mathcal{C}_{d'}^\lambda$ (at least one such W_d must exist, since $\mathcal{C}_d^\lambda \subset \mathcal{C}_{d'}^\lambda$) and that, for any behavior $\mathbf{p} \in \mathcal{C}_d^\lambda$, we have a bound $W_d \cdot \mathbf{p} \leq C_d$. Because W_d does not define a facet of $\mathcal{C}_{d'}^\lambda$, there is some $\mathbf{p}' \in \mathcal{C}_{d'}^\lambda$ such that $W_d \cdot \mathbf{p}' > C_d$. Hence, if we are given preparation and measurement boxes which are guaranteed to prepare only classical systems, and we observe some behavior that, like $\mathbf{p}'$, violates the bound on W_d, we can certify that our uncharacterized devices is preparing states of dimension at least $d+1$.

In disguise, I have exemplified what is called a (dimension) witness: any functional and bound that, (i) for some set of behaviors is never violated, but that (ii) can be violated by at least one behavior of some other set, is said to *witness* some property. In our case, we are witnessing dimension in a semi device-independent fashion, as we have assumed the preparations are classical. Lastly, notice that the facets defining any $\mathcal{C}_d^\lambda$ polytope are, by definition, dimension witnesses. Furthermore, they are *tight* witnesses, something which not all witnesses must be.

For completeness, I note it is also possible to define a $\mathcal{C}_d^\rho$ set where the preparations remain classical, but the devices can be correlated through a shared quantum state ρ [101, 90]. When ρ is an entangled state, it can lead to interesting behaviors associated with advantages in communication protocols [102].

3.1.2 Quantum preparations

To continue generalizing our discussion, let us also remove the assumption of classical preparations.

In that case, $\mathcal{S}$ will be a finite subset of $\mathcal{D}(\mathcal{H})$, and y a choice of quantum measurement. Employing Born's rule, we rewrite the behaviors as $\mathbf{p}_Q = \{\mathrm{tr}\left(\rho_x E_{b|y}\right)\}_{b,x,y}$. Here, all $\rho_x \in \mathcal{S}$, and $\{E_{b|y}\}_b$ is a POVM for each y. For the dimension bound, let $\dim(\mathcal{S}) = \dim \sum_{\rho_x \in \mathcal{S}} \mathrm{supp}\,(\rho_x)$ be the smallest Hilbert space dimension needed to represent all density operators in $\mathcal{S}$. Then, in direct analogy to the classical behaviors sets, for any fixed X and Y, we define $\mathcal{Q}_d$ as the set of behaviors for quantum communication with uncorrelated devices.

Allowing for shared randomness introduces a slight change in the elements of the behaviors. Using the same notation as before, they turn into

$$p\,(b \mid x, y) = \int_\Lambda \pi(\lambda)\mathrm{tr}\left(\rho_x^\lambda E_{b|y}^\lambda\right),$$

making it true that $\mathcal{Q}_d^\lambda = \mathrm{conv}\,(\mathcal{Q}_d)$ [**Review!** Mas $\mathcal{Q}_d$ já é convexo, não? Então $\mathcal{Q}_d^\lambda = \mathcal{Q}_d$?]

Lastly, and most importantly for chap. 5, Alice and Bob may share a quantum system $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$ and use it as a resource to improve their quantum communication. Most generally, Alice can use her share of ρ as an aid in her encoding of x. The way to do it is by applying a local CPTP map Λ_x. As we must bound the communication to d-dimensional systems, $\Lambda_x : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_C)$, where $\mathcal{H}_C \simeq \mathbb{C}^d$. This system in $\mathcal{H}_C$ is then transmitted to Bob, who will afterwards hold $\rho_x = (\Lambda_x \otimes \mathbf{1}_B)\,\rho$. His measurements' effects act on $\mathcal{H}_C \otimes \mathcal{H}_B$. Behaviors compatible with experiments

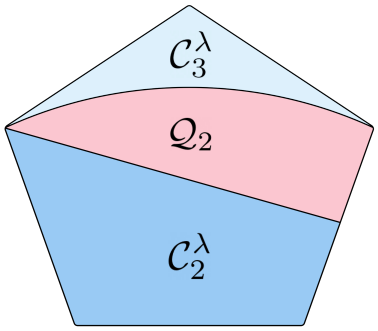


Figure 3.3: Pictorial representation of prepare-and-measure behaviors. $\mathcal{C}_2^\lambda$, arising from 2-dimensional classical communication with shared randomness, is a subset of any $\mathcal{C}_d^\lambda$ with $d > 2$ (eq. 3.2). Qubit preparations, $\mathcal{Q}_2$, can behave in ways that violate the facets of $\mathcal{C}_2^\lambda$ [100].

implementing this procedure are in the set $\mathcal{Q}_d^\rho$.

Sometimes, it is a reasonable assumption to make $\mathcal{H}_A \simeq \mathcal{H}_C$ (see sec. 3.2.2 and chap. 5). Bear in mind, though, that it can lead to some loss of generality: recently, Tavakoli et al. [101] considered a qubit communication protocol where 4-dimensional entanglement shows advantage over having only a 2-dimensional entangled resource. To the best of my knoledge, those, together with Moreno et al.'s ([2] and chap. 5), were the first results considering this most general quantum communication scenario.

As a side note, although I have presented entanglement-assisted quantum communication scenarios as the most general kind of PM scenario, there are still further possible generalizations if we allow Alice's and/or Bob's inputs x and y to also be quantum variables [103]. These are interesting and little explored scenarios, but dealing with them is out of our scope.

Clearly, any of the sets $\mathcal{Q}$, $\mathcal{Q}^\lambda$ and $\mathcal{Q}^\rho$ is contained in their respective larger-dimensional counterparts. Thence, the same observations made before make it possible to also derive *quantum* dimension witnesses [100]. Whereas witnessing dimension for classical or quantum preparations is already of practical and fundamental interest, we ideally want to also distinguish between classical and strictly quantum behaviors. To that end, it is useful to recognize that a set of pair-wise commuting density operators can only generate classically reproducible statistics. More precisely, if $\mathcal{S}_C$ is a set of d-dimensional density operators such that $[\rho_i, \rho_j] = 0, \forall \rho_i, \rho_j \in \mathcal{S}_C$, then they can be simultaneously diagonalized w.r.t. some orthonormal basis $\{|e_i\rangle\}_{i=1}^d$. There can be at most $X = d$ d-dimensional preparations with this property. Letting $\rho_x = \sum_i c_{i,x} |e_i\rangle\langle e_i|$, we get that $\mathrm{tr}\left(E_{b|y}\rho_x\right) = \sum_i c_{i,x} \langle e_i|E_{b|y}|e_i\rangle$. Operationally, this means that communicating the label x (a classical variable with dimension $X \leq d$) is enough to reproduce all possible behaviors: upon getting x and Bob's choice y, the measurement device samples $|e_i\rangle$ with probability $c_{i,x}$, measures it considering y, and after collecting asymptotically many results the parties recover the correct $p(b \mid x, y)$.

As not all quantum states commute, it turns out that $\mathcal{C}_d \subset \mathcal{Q}_d$, and similarly for the sets allowing for shared randomness and entanglement assistance. Collecting the aforestated set relationships, we find that

$$\mathcal{C}_d \subset \mathcal{Q}_d \subset \mathcal{Q}_d^\lambda \subset \mathcal{Q}_d^\rho, \tag{3.3}$$

$$\mathcal{C}_d \subset \mathcal{C}_d^\lambda \subset \mathcal{Q}_d^\lambda \subset \mathcal{Q}_d^\rho, \tag{3.4}$$

$$\mathcal{C}_d \subset \mathcal{C}_d^\lambda \subset \mathcal{C}_d^\rho \subset \mathcal{Q}_d^\rho, \tag{3.5}$$

showing it is thus also possible to build nonclassicality witnesses for the prepare-and-measure scenario. Fig. 3.3 illustrates some of the set relationships discussed in this section.

* * *

In recent years, a multitude of dimension and nonclassicality witnesses for prepare-and-measure scenarios have been proposed [100, 88, 104, 97, 95, 105, 106] and some of them experimentally tested [107, 108, 109]. More stringent conditions can also be used to perform self-testing of states and measurements, which can be used to certify the use of mutually unbiased bases, nonprojective measurements, or targeted sets of states [110, 111, 112, 113, 114, 115, 2]. All the while, known computational methods can aid in bounding the set of finite dimensional quantum correlations in prepare-and-measure scenarios [116, 117]. The prepare-and-measure scenario and its sequential, or prepare-transform-and-measure variations [118, 119] are also expected to serve as building blocks for quantum networks [89, 120]. Complementing their versatility, they can also be seen as a physical implementation of many paradigmatic communication protocols such as random access coding and dense coding, which I now briefly introduce.

3.2 Prepare-and-measure communication protocols

The preeminent practical interest in prepare-and-measure scenarios is due to the fact that quantum communication can beat classical protocols in many tasks that can me modeled as preparing and measuring states. These include cryptographic key distribution [121], secret sharing [122] and communication complexity scenarios [123]. Simplifiedly, the latter deals with the question of how much is the least amount of communication needed to compute or approximate some function $f(x, y)$ where the inputs x and y are distributed. Prepare-and-measure scenarios clearly resemble this structure. All we have to do is to stipulate a figure of merit that, generally depending on f, measures how well some protocol performs. If we additionally manage to show some separation between the best performance achievable, for instance, in $\mathcal{C}_d^\lambda$ against $\mathcal{Q}_d^\lambda$, we can prove quantum advantage in the task. A special and widely studied choice of $f(x, y) = x_y$ leads us to the so-called *random access coding* (RAC) protocol.

3.2.1 Random access coding

An $(n, d) \mapsto m$ random access code (RAC) is a communication task in which a party — commonly Alice, — is given a n-length $d$itstring $\mathbf{x} = x_1 x_2 \dots x_n$, with each $x_i \in \{0, \dots, d-1\}$, and required to encode it in another ditstring $\mathbf{a}$ of length m, where $m < n$.

The m dits representing $\mathbf{a} = \mathcal{E}(\mathbf{x})$, where $\mathcal{E} : \{0, \dots, d-1\}^n \mapsto \{0, \dots, d-1\}^m$ is an encoder function, are then sent to a second party, whom we'll call Bob. Bob is queried with an $y \in \{1, \dots, n\}$ and must correspondingly guess what was the value of x_y. His guess may be modeled through n decoding functions $\mathcal{D}_y : \{0, \dots, d-1\}^m \mapsto \{0, \dots, d-1\}$ that are chosen depending on the query y. We name *encoding-decoding strategy* the ordered set $(\mathcal{E}, \mathcal{D}_1, \dots, \mathcal{D}_n)$.

When Alice and Bob use the same strategy in all rounds of the protocol, the observed statistics are always deterministic, i.e, the probability of Bob answering b is given by $p(b \mid \mathbf{x}, y) = \delta[b, (\mathcal{D}_y \circ \mathcal{E})(\mathbf{x})]$, and in each round of a RAC, Bob's probability of guessing the right value is given by $p(b = x_y \mid \mathbf{x}, y)$.

She and he must cooperate to guess as best as they can. Their performance is typically measured through the *worst-case success probability* p_{worst}, defined as the minimum guess probability $p(b = x_y \mid \mathbf{x}, y)$ occurring for their particular encoding-decoding strategy. When the best possible strategy for an $(n, d) \mapsto m$ scenario is such that $p_{\text{worst}} \leq 0.5$, the RAC is said to not exist, as in that case a better or equivalent performance could be achieved through independently guessing.

Shared randomness is known to improve performance in this task. This is done by allowing the encoding and decoding functions to be correlated by some pre-established random variable λ. For

a concrete example, consider the simplest $(2,2) \mapsto 1$ scenario. When no SR is allowed, $p_{\text{worst}} \leq 0.5$ [124]. With SR, however, Alice and Bob can cooperate in the following way. Before each round, a random variable $\lambda \in \{0,1\}$ instructs Alice to send $\mathbf{a} = x_\lambda$. Bob knows λ, so if he's queried with $y = \lambda$, he can answer correctly with certainty; otherwise he can just flip a coin. Of course, we still have $p_{\text{worst}} \leq 0.5$ in this situation. However, it is known from [125] that when SR is allowed, the best possible strategy is such that [**Review!** Confirmar se é isso mesmo]

$$p_{\text{worst}} = p_{\text{avg}} \equiv \frac{1}{nd^n} \sum_{\mathbf{x},y} p\left(b = x_y \mid \mathbf{x}, y\right) \tag{3.6}$$

for an uniform distribution on y, proving that a $(2,2) \mapsto 1$ SR-RAC with $p_{\text{worst}} = 0.75$ exists.

Something else that can further improve performance is allowing quantum communication. In this case, we interpret the encoders $\mathcal{E}(\mathbf{x}) = \rho_x \in \mathcal{D}(d)$ as preparation procedures and the decoders $\mathcal{D}_y = \{E_{b|y}\}_b$ as quantum measurements.

Quantum random access codes (QRACs) first appeared in [126] and were later rediscovered and linked to quantum automata in [124]. With this popularization, multiple new results and experimental demonstrations regarding the existence and advantage of QRACs over their classical counterparts rapidly ensued [127, 128, 129, 125, 130, 121].

Mapping QRACs to prepare-and-measure scenarios with quantum communication amounts to choosing $|\mathcal{X}| = d^n$ preparations $\rho_x \in \mathcal{D}(d)$, and $|\mathcal{Y}| = n$ choices of POVMs with $|\mathcal{B}| = d$ outcomes each [95]. The set of behaviors is then $\mathcal{Q}_{d,d^n,n}$, and optimizing the protocol means looking for a $\mathbf{p} \in \mathcal{Q}_{d,d^n,n}$ that maximizes either p_{worst} or p_{avg}, depending on which is the chosen figure of merit. In the next chapter we will make use of this mapping to demonstrate an interesting quantum advantage activation phenomenon in QRACs.

Additionally to RACs, SR-RACs and QRACs, one could also investigate SR-QRACs [125], EA-RACs [102] and EA-QRACs, where "EA" stands for "entanglement-assisted". Each of these cases (which I'll herein collectively refer to as simply "RACs") could analogously be mapped to a PM scenario, and the optimal solutions would be searched for inside the PM behavior sets $\mathcal{C}$, $\mathcal{C}^\lambda$, $\mathcal{Q}$, $\mathcal{Q}^\lambda$, $\mathcal{C}^\rho$ and $\mathcal{Q}^\rho$, respectively, all with subscript d, d^n, n.

To end this section I note that while RACs can be cast as an instance of prepare-and-measure, the inverse is not true. Investigating what different kinds of *information retrieval tasks* [131] arise from such other instances could be an interesting research problem.

3.2.2 Dense coding

Holevo's bound guarantees that n qubits can perfectly encode no more than n bits of information [132]. Calling to mind that n qubits require $2^n - 1$ complex coefficients to be fully described, this result comes to be tremendously surprising. Looking more closely, the setting where this conclusion arises from is a prepare-and-measure scenario with quantum communication but independent devices. Another seminal result proves that when the devices are not independent, but rather, share a maximally entangled state as a resource, it becomes possible to communicate two bits by sending a single qubit.

Dense coding was first proposed by Bennett and Wiesner [21] similarly to the following argument. Let Alice and Bob share a two-qubit maximally entangled state $|\Phi^+\rangle = \frac{1}{\sqrt{2}} \sum_{i=0}^{1} |ii\rangle$ (the same argument is valid for any unitary transformation of it). Her task is to communicate one out of four messages to Bob. We label those messages with choices from the ordered set $\mathcal{X} = (0,1,2,3)$, which could be perfectly encoded with two bits. They previously agree on some special set of transformations, hereby $\Lambda = (\mathbf{1}, \sigma_x, \sigma_y, \sigma_z)$ — where σ_i are the Pauli matrices —, to represent the

respective encodings. Alice's encoding of x is a simple matter of applying Λ_x to her share of $|\Phi^+\rangle$. After receiving her qubit, Bob will have the state $(\Lambda_x \otimes \mathbf{1})\,|\Phi^+\rangle\langle\Phi^+|$. The four possibilities are mutually orthogonal, thus can be perfectly distinguished by some suitable measurement (e.g., a standard Bell-basis measurement). He will hence be able to recover x from a single communicated qubit.

Purposefully, I have framed the protocol in the previously introduced notation for a prepare-and-measure scenario with quantum communication and entanglement assistance. My intention was to suggest that dense coding can be implemented as a special instance of prepare-and-measure scenarios, as it indeed can. More than that, several generalizations of this protocol are possible. The most straightforward ones are to allow for higher dimensional communication and entanglement [21] or mixed-state entanglement [133]. More recently, discussions have been opened in regard to unbounded entanglement [101] or even dense coding protocols with errors [134, 2]. Chapter 5 will detailedly discuss how we can use insights from prepare-and-measure scenarios to study the latter case. [**Comment:** I plan to do the formal definition on chap. 5 because even if I did it here I'd probably have to repeat it there. That's why I didnt make it already but I can put it here as well if it's better...]

Chapter 4

Classicality in the prepare-and-measure scenario

A key objective in the study of quantum correlations is understanding what makes some behaviors be surely quantum, while others may be not. Progress on this issue improves our understanding of what is "quantum" in quantum theory, which in turn will lead to improved or even innovative technological applications.

In this chapter, I introduce novel methods for deciding whether some set of quantum preparations or quantum measurements behaves classically. These will, in turn, provide new insights about a nonclassicality activation phenomenon and the relation between measurement incompatibility and quantum behaviors in prepare-and-measure scenarios.

These results were published in [1], and computational details are provided in appendix A.

* * *

To make the problem we want to solve precise, let us first recall that the most general prepare-and-measure behaviors involving strictly classical variables are those in $\mathcal{C}^{\lambda}_{d,B,X,Y}$ (eq. (3.2)). Contrastingly, the $\mathcal{Q}_{d,B,X,Y}$ contain the *least* general quantum behaviors, which are those needing only quantum preparations with independent devices (sec. 3.1.2). The relation $\mathcal{C}^{\lambda}_{d,B,X,Y} \subset \mathcal{Q}_{d,B,X,Y}$ reflects the fact that some behaviors arising from quantum preparations are nevertheless in $\mathcal{C}^{\lambda}_{d,B,X,Y}$, while others are definitely not. Those which are, can be simulated by classical preparations with shared randomness. Hereafter, we will call any behavior in the $\mathcal{C}^{\lambda}_{d,B,X,Y}$ *classically reproducible/simulatable* — or just *classical*, for short.

It is important to keep in mind that several distinct notions of classicality exist in the literature. As is the case of entanglement measures, some applications may call for a classicality model or another. Our choice of $\mathcal{C}^{\lambda}_{d,B,X,Y}$ for the classical behaviors set is general in the sense discussed in chap. 3. It is also in line with widely studied classicality models in different correlation scenarios, such as local hidden variables and local hidden states models [11, 16, 17].

Suppose a behavior $\mathbf{p}$ arose from a set $\mathcal{S}$ of quantum preparations under some collection of measurements, and furthermore that it is classically reproducible with $d$its. This means that $\mathbf{p} \in \mathcal{C}^{\lambda}_{d,B,X,Y}$. But it does *not* mean that $\mathcal{S}$ always behaves classically, for it may be the case that different measurements would reveal nonclassicalities in $\mathcal{S}$. Even if we certify $\mathcal{S}$ is classical for all possible sets of Y measurements, it could happen that increasing the number of measurement choices leads to nonclassicality [90]. To better understand what makes quantum preparations behave classically of not, we must then deal with any number of uncharacterized measurements. This is precisely the task we partially solve in sec. 4.1.

Before tackling that, it is both instructive and useful to further examine the $\mathbf{p} \in \mathcal{C}^{\lambda}_{d,B,X,Y}$ membership problem. As previously discussed, any $\mathcal{C}^{\lambda}_{d,B,X,Y}$ is a polytope, thus set membership can be certified through compliance with a finite set of linear inequalities representing its facets. If one is given the facets, this a computationally easy problem, but obtaining the facets is not. The whole complexity of the problem then lies in characterizing a the polytope for a given scenario. As first noticed in [100], this can be done by enumerating all of $\mathcal{C}^{\lambda}_{d,B,X,Y}$'s extremal points and finding their convex hull (see sec. 2.1). The extremal points — also called *deterministic behaviors/strategies* —, are found from eq. (3.1) when the response functions are deterministic, and their enumeration can be cast as the following procedure.

Let $\mathbf{p}_D = (p(b \mid x, y))_{b,x,y}$ be an ordered list representing a deterministic behavior. To simplify the notation, let us consider $b \in \{0, 1\}$ (the generalization is straightforward) and write $\mathbf{p}_D = (\mathbf{v}_x)_{x=1}^{X}$, where $\mathbf{v}_x = (p(b = 0 \mid x, 1), \ldots, p(b = 0 \mid x, Y))$ will be called a *substring*. Only the $b = 0$ probabilities must be considered, as $p(1 \mid x, y) = 1 - p(0 \mid x, y)$. A deterministic behavior is valued as 1 for one of the outcomes and, of course, 0 for all others. Whenever $d = X$, there will be 2^{XY} distinct $\mathbf{p}_D$ vectors. In this $B = 2$ case, those are exactly all possible XY-length $d$itstrings but, in the general case, there will be B^{XY} vectors corresponding to only those $(B-1)XY$-length $d$itstrings with a single 1 in each $B-1$-length consecutive substring. We can think of each of these vectors as a deterministic strategy where Alice's choice of x is unambiguously encoded in a classical state τ_x, from which Bob may then perfectly recover x through classical post-processing. The nontrivial case happens when $d < X$, i.e., when there are only d distinct classical preparations. As the preparation box has more buttons than distinct classical states, at least $\lceil X/d \rceil$ inputs will be mapped to the same message. Whenever $\tau_i = \tau_j$, obviously $\mathbf{v}_i = \mathbf{v}_j$, hence our vector $\mathbf{p}_D$ will have $d - X$ repeated substrings. With this in mind, enumerating all possible deterministic points amounts to generating every unique dY-length ditstrings, then repeating $d - X$ substrings $\mathbf{v}_x$ in all possible arrangements.

The number of deterministic points scales exponentially as $N_\lambda \propto (B-1)^{dY}$, which makes the change to the hyperplane description intractable for all but the smallest parameters. One solvable case that will be used in sec. 4.1.1 is for $\mathcal{C}^{\lambda}_{2,2,4,2}$, where the polytope is defined by the two following classes of nontrivial inequalities [90]

$$S = E_{11} - E_{12} - E_{21} + E_{22} - E_{31} - E_{32} + E_{41} + E_{42} \leq 4 \tag{4.1a}$$

$$S' = E_{11} + E_{12} + E_{21} - E_{22} - E_{41} \leq 3 \tag{4.1b}$$

where the $E_{xy} = p\,(b = 0 \mid x, y) - p\,(b = 1 \mid x, y)$ are the so-called *correlators*.

Framing it with quantum preparations, this scenario corresponds to a set $\mathcal{S}$ with $X = 4$ possible $d = 2$-dimensional quantum systems (qubits) that will be prepared by Alice's device then measured by Bob's. He will have $Y = 2$ choices of POVMs, where each POVM is composed of $B = 2$ effects. An interesting fact in this scenario (as well as many others) is that some quantum preparations and measurements, even with independent devices, can violate inequalities S or S'. When this happens, the behavior is not classically reproducible. In many cases, nonclassicalities provide quantum advantage in informational protocols (sec. 3.2).

4.1 Classicality of preparations

But we are interested in a much more general question than that of certifying whether or not a *behavior* is classically reproducible, namely, that of certifying if the set $\mathcal{S}$ of quantum preparations is itself classical. For that to be true, there must be no measurements that upon acting on $\mathcal{S}$ end

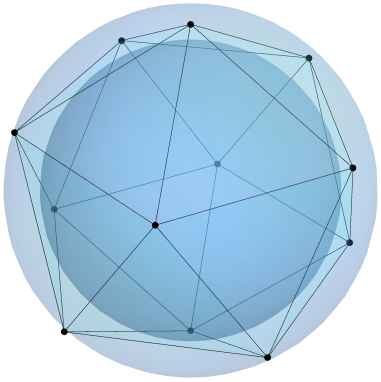


Figure 4.1: Representation of the method for $d = 2$. Each vertex on the Bloch sphere represents a measurement operator. To every $\Pi_{0|y}$ we associate a corresponding antipodal $\Pi_{1|y}$. Measurements inside the set enclosed by $\mathbf{conv}(\mathcal{M})$ can be simulated by mixing these extremal ones. In particular, any measurement $\{\Pi^{\eta}_{b|u}\}_b$ in a ball with radius η inscribed in the polytope is simulable in such manner.

up in a nonclassical behavior. Brute-forcing our way into the answer is off the table, for this would amount to testing the behaviors of $\mathcal{S}$ for *all* (infinitely many) possible measurements.

This is a similar problem to that of finding local hidden variable models for entangled states in Bell scenarios [40, 42], for which a general computational method guaranteeing a sufficient condition has recently been proposed [135, 136]. We now adapt it to the prepare-and-measure scenario.

Suppose that, for the quantum states in $\mathcal{S}$ and a finite set $\mathcal{M} = \{E_{b|y}\}_{b,y}$ of Y measurements, the classical model (3.2) exists, i.e., that

$$\mathrm{tr}\left(\rho_x E_{b|y}\right) = \int_{\Lambda} \sum_{m=1}^{d} \pi(\lambda) p\left(m \mid x, \lambda\right) p\left(b \mid m, y, \lambda\right) d\lambda \quad \forall b, x, y. \tag{4.2}$$

From trace's linearity, for any convex sum of effects,

$$\mathrm{tr}\left[\rho_x \left(w E_{b|y} + (1-w) E_{b'|y'}\right)\right] = w\mathrm{tr}\left(\rho_x E_{b|y}\right) + (1-w)\mathrm{tr}\left(\rho_x E_{b'|y'}\right).$$

Whence, $\mathcal{S}$'s behaviors are classical under all measurements in $\mathrm{conv}\left(\mathcal{M}\right)$, and in particular for all measurements whose effects are inside the largest ball we can inscribe in $\mathrm{conv}\left(\mathcal{M}\right)$ (fig. 4.1). For normalized Bloch vectors (c.f. eq. (1.14)), $0 \leq \eta \leq 1$ will be its radius. This ball is nothing but a depolarized Bloch ball $\eta\, \mathcal{BL}\left(d\right)$. This suggests that by probing the preparations with but a finite number of measurements and (possibly) finding a classical model, we can nevertheless certify classicality for the infinitely many measurements whose effects are in $\eta\, \mathcal{BL}\left(d\right)$. They can be written as

$$\Phi_{\eta}(E_{b|y}) = \eta E_{b|y} + (1-\eta)\frac{\mathrm{tr}\left(E_{b|y}\right)}{d}\mathbf{1} \equiv E^{\eta}_{b|y}, \tag{4.3}$$

but they are still not all possible measurements. As an example, unless $\eta = 1$, many rank-1 projective measurements (those on the Bloch sphere) will be left out.

The work-around is to consider the dual map $\Phi^{\dagger}_{\eta}$ applied to the preparations. This will result in each

$$\rho_x \mapsto \Phi^{\dagger}_{\eta}(\rho_x) = \frac{1}{\eta}\left[\rho_x - (1-\eta)\frac{\mathbf{1}}{d}\right] \equiv O_x. \tag{4.4}$$

Analyzing the behavior of the O_x under the $E^{\eta}_{b|y}$ tells us that

$$\begin{aligned}\operatorname{tr}\left(O_x E^{\eta}_{b|y}\right) &= \operatorname{tr}\left\{\frac{1}{\eta}\left[\rho_x - (1-\eta)\frac{\mathbf{1}}{d}\right]\left[\eta E_{b|y} + (1-\eta)\frac{\operatorname{tr}\left(E_{b|y}\right)}{d}\mathbf{1}\right]\right\} \\ &= \operatorname{tr}\left(\rho_x E_{b|y}\right) - \left[\frac{1-\eta}{d} - \frac{1-\eta}{d\eta} + \frac{(1-\eta)^2}{d\eta}\right]\operatorname{tr}\left(E_{b|y}\right).\end{aligned}$$

If all effects $E_{b|y}$ are rank-1 projectors, then $\operatorname{tr}\left(E_{b|y}\right) = 1$ (sec. 1.4), making it true that

$$\operatorname{tr}\left(O_x E^{\eta}_{b|y}\right) = \operatorname{tr}\left(\rho_x E_{b|y}\right). \tag{4.5}$$

The procedure to test $\mathcal{S}$ for classicality follows from precisely this trace equality relation. We start out by defining an operator O_x for each state $\rho_x \in \mathcal{S}$. Ensuingly, we probe the $\mathcal{O} = \{O_x\}_x$ with some set $\mathcal{M} = \{E_{b|y}\}_{b,y}$ of rank-1 projective measurements. If the model on the r.h.s. of eq. (4.2) exists for $\mathcal{O}$ and $\mathcal{M}$, the preceding discussion reveals it also exists for every measurement in $\operatorname{conv}(\mathcal{M})$. It will then be true that $\mathbf{p} = \left\{\operatorname{tr}\left(O_x E^{\eta}_{b|y}\right)\right\}_{b,x,y}$ is a classical behavior for *all* measurements whose effects can be written as $E^{\eta}_{b|y}$. In turn, eq. (4.5) implies that the ρ_x admit a classical model for all $E_{b|y}$, which are all rank-1 projections. Measurement results for projections of larger rank can be inferred from rank-1 projections through coarse graining (classical post-processing), hence the result is valid for all projective measurements. This construction is stated as theorem **??** in appendix A.

To cast this procedure as an optimization problem, recall that the deterministic strategies λ can be enumerated as discussed in the previous section. Assuming that is done, the integral in (4.2) turns into a finite sum. Besides, the distance from the nearest hyperplane in the convex hull of $\mathcal{M}$ to the origin is η. Further observing that the maximally mixed state $\mathbf{1}/d$ is trivially classical, we write

$$\begin{aligned}
&\text{given} && \mathcal{S},\ \mathcal{M},\ \eta, \{\lambda\} && \text{(4.6a)}\\
&\max_{\pi(\lambda)} && \alpha && \text{(4.6b)}\\
&\text{s.t.} && \alpha\rho_x + (1-\alpha)\frac{\mathbf{1}}{d} = \eta O_x + (1-\eta)\frac{\mathbf{1}}{d}, \quad \forall x && \text{(4.6c)}\\
& && \operatorname{tr}(E_{b|y}O_x) = \sum_{m,\lambda}\pi(\lambda)p(m|x,\lambda)p(b|m,y,\lambda), \quad \forall b,x,y && \text{(4.6d)}\\
& && 0 \leq \alpha \leq 1 && \text{(4.6e)}\\
& && \pi(\lambda) \geq 0 && \text{(4.6f)}\\
& && \sum_{\lambda}\pi(\lambda) = 1. && \text{(4.6g)}
\end{aligned}$$

Allowing the ρ_x to be mixed with the identity in the l.h.s. of eq. (4.6c) guarantees the program will always have a feasible solution. Constraints (4.6d)–(4.6g) force the local model to exist, the mixtures to be valid density operators, and π to be a probability distribution over the deterministic strategies, respectively. Any solution where $\alpha = 1$ certifies that preparations in $\mathcal{S}$ admit a classical model for all projective measurements. On the other hand, any $\alpha < 1$ certifies only that the preparations $\alpha\mathcal{S} = \{\alpha\rho_x\}_x$ are classical. This criterion turns from sufficient to both necessary and sufficient only when $\eta \to 1$, a regime approachable by increasing the number Y of probe measurements.

Being a linear program (sec. 2.2.1), eq. 4.6 can be efficiently solved up to numerical precision.

Regardless of linear programming complexity, an instance of program (4.6) has N_λ variables. As already noted, $N_\lambda \propto (B-1)^{dY}$ thus the program size scales exponentially with the number of measurements. It will always be of interest to maximize η, and consequently, to maximize Y. With $X = 4$ preparations, a common desktop computer can usually only handle less than $Y = 12$ projective qubit measurements. This limitation can be circumvented through the iterative optimization procedure described in sec. A, which was applied in all upcoming results.

It is also possible to extend our criterion to non-projective measurements. Dichotomic projective measurements are the extremal two-outcome POVMs. Consequently, out method actually guarantees classicality for all $B = 2$-outcome measurements. For all other cases, we can extend it by simulatating POVMs with projective measurements (c.f. sec. 1.3). To see how, recall that any set of generalized measurements $\mathcal{M} = \{E_{b|y}\}_{b,y}$ becomes projective-simulatable after a certain amount t of depolarization Observing that

$$\mathrm{tr}\left[E_{b|y}\Phi_t(\rho_x)\right] = \mathrm{tr}\left[\Phi_t(E_{b|y})\rho_x\right] \tag{4.7}$$

leads us to conclude that testing $\mathcal{S} = \{\rho_x\}_x$ for classicality under all POVMs is the same as certifying that preparations

$$\rho'_x \equiv \Phi_t^\dagger(\rho_x) = \frac{1}{t}\left(\rho_x - \frac{1-t}{d}\mathbf{1}\right) \tag{4.8}$$

are classical for all projective measurements, where t is some amount of depolarization such that $\Phi_t(\mathcal{M})$ are projective-simulatable.

Program (4.6) can be easily modified to this case by either adding an extra constraint or through explicit rewriting. For the former, it reads

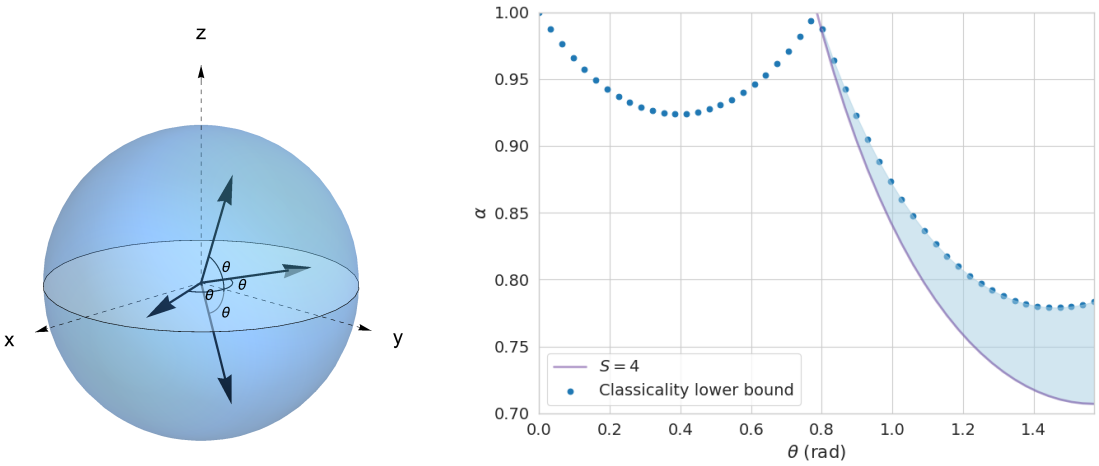
$$\begin{aligned}
&\text{given} && \mathcal{S},\ \mathcal{M},\ \eta, \{\lambda\}, t && \text{(4.9a)}\\
&\max_{\pi(\lambda)} && \alpha && \text{(4.9b)}\\
&\text{s.t.} && \rho'_x = \frac{1}{t}\left(\rho_x - \frac{1-t}{d}\mathbf{1}\right), \quad \forall x && \text{(4.9c)}\\
& && \alpha\rho'_x + (1-\alpha)\frac{\mathbf{1}}{d} = \eta O_x + (1-\eta)\frac{\mathbf{1}}{d}, \quad \forall x && \text{(4.9d)}\\
& && \mathrm{tr}(E_{b|y}O_x) = \sum_{m,\lambda}\pi(\lambda)p(m|x,\lambda)p(b|m,y,\lambda), \quad \forall b,x,y && \text{(4.9e)}\\
& && 0 \leq \alpha \leq 1 && \text{(4.9f)}\\
& && \pi(\lambda) \geq 0 && \text{(4.9g)}\\
& && \sum_\lambda \pi(\lambda) = 1, && \text{(4.9h)}
\end{aligned}$$

which is also a linear program.

4.1.1 Nonclassicality activation and quantum advantage in a RAC

Nonclassicality activation phenomena have been long-known in Bell scenarios, where an entangled state that only behaves locally may have its nonlocality activated, for instance, by local filtering, broadcasting, or by exploring multiple copies of the state [137, 138, 139, 140, 141, 142, 143]. Very recently, two interesting, similar phenomena were explored in PM scenarios [90] — one related to increasing the number of allowed measurements, and the other to the number of preparations. Our method can be applied to demonstrate a stronger form of the latter.

Let us start by defining $\mathcal{S}(\alpha,\theta) = \{\rho_{\mathbf{r_1}}, \rho_{\mathbf{r_2}}, \rho_{\mathbf{r_3}}, \rho_{\mathbf{r_4}}\}$ as the preparation set illustrated in fig.



(a) Preparations

(b) Activation

Figure 4.2: Nonclassicality activation in the prepare-and-measure scenario. On the left, preparations $\mathcal{S}(\alpha,\theta) = \{\rho_{\mathbf{r_1}}, \rho_{\mathbf{r_2}}, \rho_{\mathbf{r_3}}, \rho_{\mathbf{r_4}}\}$ are represented by their Bloch vectors, for $\alpha = 0.8$. At $\theta = 0$, all preparations are at $\alpha\mathbf{y}$. For $\theta = \pi/2$, $\mathcal{S} = \{-\alpha\mathbf{x}, \alpha\mathbf{x}, -\alpha\mathbf{z}, \alpha\mathbf{z}\}$, corresponding to the largest violation of inequality (4.1a). On the right, result of applying program 4.6 to $\mathcal{S}$. To run it, 12 probe measurement effects arranged as the vertices of a rhombicuboctahedron ($\eta \approx 0.86$). As $S \propto \alpha$, every preparation set above the $S = 4$ curve is non-classical. On the other hand, the classicality curve shows the maximum visibility such that any triad of states in the preparation set are classical. The shaded region thence represents sets of four preparations that are three-wise classical, but that that when taken together behave nonclassically. In this sense, nonclassicality is activated by measurement inclusion.

4.2(a). They are parametrized by α — a shrinking factor from the surface of the Bloch sphere —, and the angle θ. For the probe measurements, we will choose the ones parametrized by the Bloch vectors $\mathbf{q}_1 = (-\mathbf{x} + \mathbf{z})/\sqrt{2}$ and $\mathbf{q}_2 = (\mathbf{x} + \mathbf{z})/\sqrt{2}$. Substituting back into inequality S (eq. (4.1a)), we get that $S(\alpha,\theta) = 4\sqrt{2}\alpha \sin\theta$. This curve is shown in fig. 4.2(b), where any violation of $S \leq 4$ certifies nonclassicality.

Our second step is to take every subset of 3 preparations from $\mathcal{S}(\alpha,\theta)$ at some θ. For each possibility, we run program (4.6) and find the largest α^* such that *any* collection of 3 preparations is classical. Any $\alpha < \alpha^*$ has more of the identity state, thus preserves classicality. Sweeping θ, we obtain the dotted curve shown in fig. 4.2(b). The shaded region then corresponds to a situation where all triads of preparations are classical, but that taken together have their nonclassicality activated.

The significance of this results can be illustrated when considering a $(2,2) \mapsto 1$ RAC. Section 3.2.1) tells us it can me mapped to a $(d, B, X, Y) = (2, 2, 4, 2)$ prepare-and-measure instance, which is exactly the case for the class of inequalities represented by S. A more ergonomic representant for our current task is $s = E_{11} + E_{12} + E_{21} - E_{22} - E_{31} + E_{32} - E_{41} - E_{42} \leq 4$. Borrowing from [121], we note that, for this RAC, the average probability of success *is* the facet s. To see that,

let's relabel preparations $(1,2,3,4) \mapsto (00,01,10,11)$ and open up

$$\begin{aligned} p_{\text{suc}}^{\text{avg}} &= \frac{1}{8}\sum_{\mathbf{x},y} p(b = x_y|x_0x_1,y) \\ &= \frac{1}{8}\big[p(0|00,0) + p(0|00,1) + p(0|01,0) + p(1|01,1) + \\ &\qquad + p(1|10,0) + p(0|10,1) + p(1|11,0) + p(1|11,1)\big] \\ &= \frac{1}{8}\big[p(0|00,0) + p(0|00,1) + p(0|01,0) - p(0|01,1) + \\ &\qquad - p(0|10,0) + p(0|10,1) - p(0|11,0) - p(0|11,1) + 4\big] \end{aligned}$$

Switching s to full-form and applying normalization, we get

$$\begin{aligned} s = 2\big[&p(0|00,0) + p(0|00,1) + p(0|01,0) - p(0|01,1) + \\ &- p(0|10,0) + p(0|10,1) - p(0|11,0) - p(0|11,1)\big], \end{aligned}$$

and together,

$$p_{\text{avg}} = \frac{1}{8}\left(\frac{1}{2}s + 4\right) = \frac{s+8}{16} \leq \frac{3}{4}, \tag{4.10}$$

where the classical bound for s was substituted in the last term. This relation means that any violation of s in the $(2,2,4,2)$ PM scenario can be associated with a quantum advantage in the corresponding $(2,2) \mapsto 1$ QRAC. We have numerically found a lower bound for the maximum quantum violation of s to be approximately 5.65685. Substituting into eq. (4.10), $p_{\text{avg}} \approx 0.853553$, exactly matching the result found in sec. 3.3.1. of [125]. This also suggests that our bound for s is tight. More interestingly yet, this construction shows that the nonclassicality activation phenomenon reported above also transfers to this case, where it can be interpreted as an activation of quantum advantage in a relevant quantum communication protocol.

4.2 Classicality of measurements

In the absence of entanglement, one must look for other quantum features to explain nonclassical behaviors. Measurement incompatibility (sec. 1.3) is usually at the center of many interesting quantum phenomena, and from a more practical perspective they may be seen as resources in informational tasks [144]. It is then natural to wonder whether non-joint measurability is necessary, or even sufficient, for the arisal of classically irreproducible behaviors in the prepare-and-measure scenario. Reframing the question, we ask: given a set of incompatible measurements, is there *always* some set of preparations that leads to nonclassicality?

The question at hand has been positively answered for quantum steering [18, 50, 19] and negatively for Bell nonlocality [51, 52, 53], but only partial results exist for prepare-and-measure scenarios [87]. Interestingly, this problem is similar to what we have just dealt with, but while we were previously interested in certifying some set of preparations is classical for all measurements, we now move on to certify there exists no set of *preparations* such that a fixed set of measurements unveils nonclassicality. We will soon see the previous method is straightforwardly adaptable. Even so, regarding the existing literature, this is a more innovative approach. While Sec. 4.1 develops a procedure akin to known results for different correlation scenarios [135, 136], the method we now present has not, to the best of my knowledge, been considered elsewhere.

* * *

Starting with the measurement set $\mathcal{M} = \{E_{b|y}\}_{b,y}$ of our interest, we build the operators

$$O_{b|y} = \eta E_{b|y} + (1-\eta)\frac{\mathrm{tr}\left(E_{b|y}\right)}{d}\mathbf{1}, \tag{4.11}$$

where η will soon be defined. For some set $\mathcal{S} = \{\rho_x\}_x$ of *pure* probe preparations, we consider the behavior $\mathbf{p} = \{\mathrm{tr}\left(O_{b|y}\rho_x\right)\}$. If, for all b, x and y, classicality (in the sense of eq. (3.2)) holds, then it also does for every preparation in $\mathrm{conv}\,(\mathcal{S})$. In particular, this will be true for all quantum states in the largest sphere inscribed in the convex hull of $\mathcal{S}$. We write η for its radius, and ρ_x^η for any preparation in it. A key difference from sec. 4.1 is that the η defining the $O_{b|y}$ is now the radius with respect to $\mathrm{conv}\,(\mathcal{S})$, as opposed to $\mathrm{conv}\,(\mathcal{M})$.

Our working hypothesis is that

$$\mathrm{tr}\left(O_{b|y}\rho_x\right) = \int_\Lambda \sum_{m=1}^{d} \pi(\lambda) p\left(m \mid x, \lambda\right) p\left(b \mid m, y, \lambda\right) d\lambda \quad \forall b, x, y.$$

From the discussion in the previous section, this would imply the collection of $\mathrm{tr}\left(O_{b|y}\rho_x^\eta\right)$ is also classical for *every* preparation in the η-sphere. Similarly to the trace equality (4.5), $\mathrm{tr}\left(O_{b|y}\rho_x^\eta\right) = \mathrm{tr}\left(E_{b|y}\rho_x\right)$ holds whenever the $E_{b|y}$ are rank-1 projectors, as can be seen by direct calculation. The r.h.s. of this last equation then tells the measurements in $\mathcal{M}$ leads to classicality for every possible ρ_x. Recalling we started with pure preparations, this procedure reveals that such a set of rank-1 projective measurements is classical in regard to all pure states. Pure states are the extremal points in the set of quantum states, thence, any state is a convex combination of those, and the result is valid for all states. Moreover, coarse graining of rank-1 projective measurements simulates PVMs with all larger ranks, proving our procedure works for all projective measurements. This construction is stated as theorem **??** in appendix A.

A slight modification of program (4.6) implements this criterion as the linear program

$$\begin{aligned}
&\text{given} && \mathcal{S},\ \mathcal{M},\ \eta,\ \{\lambda\} && \text{(4.12a)}\\
&\max_{\pi(\lambda)} && \alpha && \text{(4.12b)}\\
&\text{s.t.} && \alpha E_{b|y} + (1-\alpha)\frac{\mathbf{1}_d}{d} = \eta O_{b|y} + (1-\eta)\,\frac{\mathbf{1}_d}{d}, \quad \forall x && \text{(4.12c)}\\
& && \mathrm{tr}(O_{b|y}\rho_x) = \sum_{a,\lambda} \pi(\lambda) p(a|x,\lambda) p(b|a,y,\lambda), \quad \forall b,x,y && \text{(4.12d)}\\
& && 0 \le \alpha \le 1 && \text{(4.12e)}\\
& && \pi(\lambda) \ge 0 && \text{(4.12f)}\\
& && \sum_\lambda \pi(\lambda) = 1. && \text{(4.12g)}
\end{aligned}$$

Running it amounts to choosing a suitable set of pure probes $\mathcal{S} = \{\rho_x\}_x$, finding its corresponding η, generating the deterministic strategies $\{\lambda\}$, and querying for the maximum α such that the projective measurements in $\mathcal{M}$ are classical. As before, $\alpha = 1$ certifies they are classical, while $\alpha < 1$ tells us only that the measurements $\mathcal{M}_\alpha = \{\Phi_\alpha\left(E_{b|y}\right)\}_{b,y}$ are. From self-duality of the depolarizing map Φ, another valid interpretation is that $\mathcal{M}$ is classical for all preparations at least as mixed as $\Phi_\alpha\left(\rho_x\right)$.

Once more in full analogy to the preparation classicality case, we may use concepts of projective simulatability to extend this criterion to POVMs. Let t be a depolarization parameter that turns a POVM set $\mathcal{M}$ projective-simulatable, and recal that $\mathrm{tr}\left[\phi_t(E_{b|y})\rho^t\right] = \mathrm{tr}\left[E_{b|y}\phi_t(\rho)\right]$. Testing $\mathcal{M}$

for classicality is hence equivalent to certifying $\Phi_t(\mathcal{M})$ is classical with respect to a set

$$\mathcal{S}^t = \{\rho_x^t\}_x = \left\{ \frac{1}{t}\left(\rho_x - \frac{1-t}{d}\mathbf{1}\right) \right\}_x$$

of probe preparations. Starting from a set $\mathcal{S}$ of pure probes, program (4.12) can be undemandingly modified to implement this criterion, resulting in

$$\begin{aligned}
\text{given} \quad & \mathcal{S},\ \mathcal{M},\ \eta, \{\lambda\}, t && \text{(4.13a)}\\
\max_{\pi(\lambda)} \quad & \alpha && \text{(4.13b)}\\
\text{s.t.} \quad & \alpha E_{b|y} + (1-\alpha)\frac{\mathbf{1}_d}{d} = \eta O_{b|y} + (1-\eta)\frac{\mathbf{1}_d}{d}, \quad \forall x && \text{(4.13c)}\\
& \rho'_x = \frac{1}{t}\left(\rho_x - \frac{1-t}{d}\mathbf{1}\right) && \text{(4.13d)}\\
& \mathrm{tr}(O_{b|y}\rho'_x) = \sum_{a,\lambda} \pi(\lambda) p(a|x,\lambda) p(b|a,y,\lambda), \quad \forall b,x,y && \text{(4.13e)}\\
& 0 \leq \alpha \leq 1 && \text{(4.13f)}\\
& \pi(\lambda) \geq 0 && \text{(4.13g)}\\
& \sum_\lambda \pi(\lambda) = 1. && \text{(4.13h)}
\end{aligned}$$

While, as discussed around program (4.12), we must start with a set $\mathcal{S}$ of pure probe preparations, there is no issue in working with operators $\rho^t \in \mathcal{S}^t$ inside program (4.13), as all quantum states are more mixed than those.

To demonstrate the usefulness of our method, we solve an important question on the relation between measurement incompatibility and classicality in the PM scenario.

4.2.1 Measurement incompatibility is insufficient for nonclassicality

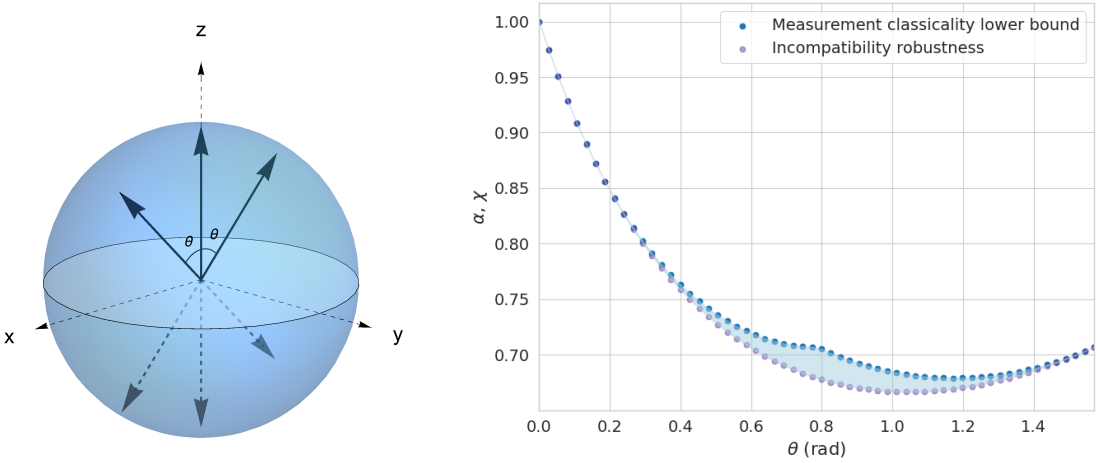
Non-joint measurability is an important concept of measurement incompatibility with a clear operational interpretation (sec. 1.3). Withal, incompatibility robustness is an useful measure of the extent to which a set of measurements is non-jointly measurable, and it can be efficiently computed through semidefinite programming (sec. 2.2.2.1). Employing our measurement classicality certification procedure, we now prove measurement incompatibility is insufficient for nonclassicality in prepare-and-measure scenarios.

Define the $Y = 3$ parametrized mirror-symmetric projective measurements defined as shown in fig. 4.3(a). For each value of θ, their incompatibility robustness is given by

$$\chi^*_{\mathcal{M}} = \sup_{\substack{\chi \in [0,1] \\ \{N_{b|y}\} \in \mathbf{N}(\mathcal{M})}} \Big\{ \chi \mid \chi\{E_{b|y}\} + (1-\chi)\{N_{b|y}\} \in \mathbf{JM} \Big\}, \tag{4.14}$$

where $\mathbf{N}(\mathcal{M})$ is a choice of noise model which is generally dependant on the desired application. Unbiased (or white) noise $N_{b|y} = \mathbf{1}/|\mathcal{B}|$ is a common choice when one wants to model experimental imperfections that affect all degrees of freedom undiscriminately. Considering this choice, the lower curve in fig. 4.3(b) stands for $\chi^*_{\mathcal{M}(\theta)}$, and it is a tight upper bound, up to numerical precision. Consequently, any value of χ above it defines a non-jointly measurable measurement set.

Program (4.12) was used to generate the upper curve in the same figure. Twenty-four probe preparations were disposed as the vertices of a rhombicuboctahedron, leading to $\eta \approx 0.86$, and the



(a) (b)

Figure 4.3: Incompatibility is insufficient for non-classicality in the prepare and measure scenario. On the left, parametrized measurements used in the demonstration. The projection on $\mathbf{z}$ and its antipodal effect are fixed, while the two other measurements vary with $0 \leq \theta \leq \pi/2$. In the upper bound, they degenerate in an $\mathbf{x}$ measurement. On the right, program (4.12) was applied. For each θ (see fig. 4.3(a)), any χ above the incompatibility robustness curve stands for an incompatible measurement set, and any α below the measurement classicality lower bound represents measurements that certifiedly do not generate non-classical statistics, regardless what preparations they act upon. Thenceforth, the shaded region contains incompatible albeit classical measurements.

deterministic strategies were iteratively explored following the algorithm described in appendix A. For each θ, the obtained α is a lower bound on the visibility of $\mathcal{M}$ such that a classical model exists for all preparations. Specifically, anything below the classicality curve defines a measurement set which cannot generate nonclassical behaviors. The non-empty region between these two curves certifies there are plenty of incompatible measurements which are classical. Accordingly, non-joint measurability is insufficient for nonclassicality in the prepare-and-measure scenario.

4.3 Open questions

Albeit the applications discussed in secs. 4.1.1 and 4.2.1 rely on the projective measurements case of our criterion, their respective instances for generalized measurements are easily implementable. They do, however, incur in more demanding computational requirements, as the measurement depolarization factor t can be seen as effectively lowering the insphere radius η. It could nevertheless be interesting to investigate applications of it, as for instance searching for a nonclassicality activation phenomenon valid for all POVMs.

A second kind of activation phenomenon, emerging by adding more measuremens instead of preparations, has also been observed when the number of preparations is limited [90]. Similarly to sec. 4.1.1, the measurement classicality criterion derived in sec. 4.2 could be used to attempt on finding a more general form of this result, where all preparations are considered.

Another interesting question is whether POVMs can be used to demonstrate nonclassicality when all PVMs fail to do so. One could approach this in the following manner. Start by selecting classical preparation sets from the results in sec. 4.1.1 — which are certifiedly classical for all projective measurements. Then characterize the facets of a $(d = 2, B > 2, X = 3, Y)$ prepare-and-measure polytope. Finally, look for a set of POVMs that, for some of those classical preparation

sets, violates any of the obtained facets.

In view of the method itself, choosing Φ as the noise model is a natural but not required choice. Different transformations, such as inscribing an ellipsis instead of a depolarized Bloch ball in $\mathrm{conv}\,(\mathcal{M})$ (and then applying the corresponding dual map to the preparations) could lead to better computational results, and, potentially, to new insights [145].

The relationship of prepare-and-measure scenarios with random access coding, which was only superficially explored in sec. 3.2.1 and sec. 4.1.1, also deserves more exploration. Although the mapping of a RAC to a PM scenario can be trivially done, this is not enough to prove a RACs p_{suc} is always a facet of the corresponding PM polytope. Believing that, as in the $(2,2) \mapsto 1$, this was always the case, we have further investigated this question. It turns out this is not true. In the general case, a RAC's p_{suc} must then be some lower dimensional face of prepare-and-measure polytopes. One possible direction is to investigate whether there are subclasses of RACs with this property, of which the $(2,2) \mapsto 1$ is but one example. Another is to explore the meaning of other classes of inequalities defining PM polytopes, and if they represent other instances of information retrieval tasks.

We believe all these suggestions are worthy of pursuit. Implementations of the programs used in this chapter could be helpful to some of these questions, and are available at a public code repository [3].

Chapter 5

Dense coding in the prepare-and-measure scenario

Dense coding is an astonishingly straightforward use of quantum systems to improve the transmission of classical information. In sharp contrast to Holevo's bound [132], it allows for the lossless encoding of two dits in one single qudit. Entanglement between the communicating parties is the price to pay for that.

In its original formulation [21], dense coding is a device dependent protocol: Alice's encoding operators and Bob's measurements must be fully characterized. Already hinted in sec. 3.2.2 is the possibility of interpreting a type of prepare-and-measure scenario — the one with quantum preparations, entanglement assistance, and a single measurement — as a physical implementation of dense coding. In return, the characterization requirements are eased, and, remarkably, interesting properties can be inferred from imposing bounds on the amount of communication and observing the behaviors.

Our task in this chapter is to define this device-independent formulation of dense coding and derive some first results. Among them, we will show how to build entanglement witnesses, self-test maximally entangled states, and optimize the preparations and measurements to better perform the protocol. Stepping in the direction of more general entanglement-assisted prepare-and-mesure scenarios, we also provide a witness in the case where more measurements are allowed.

These results were published in [2], and all their proofs are provided in appendix B.

5.1 Semi-device independent dense coding

In dense coding, two parties share an entangled pair and communicate via a quantum state. Quantum communication happens one-way, and the transmitted state is of the same local dimension (w.r.t. the encoding device) as the entangled pair. The task is to encode a classical message $x \in \mathcal{X} \equiv \{0, \ldots, N-1\}$, and it is known that two $d$its ($N = d^2$) can be perfectly recovered from one qudit of communication, for some choices of a measurement with N outcomes []**[To-Do:** where?].

Translating to the prepare-and-measure lingo, we let Alice and Bob share an entangled state $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ as a resource. Their local dimensions, $\dim(\mathcal{H}_A) = d_A$ and $\dim(\mathcal{H}_B) = d_B$, need not be the same, but Alice's preparation must be of dimension d_A. To achieve quantum advantage in the dense coding task, Alice and Bob must be able to exploit the correlations available in their entangled pair. Under the circumstances, her most general strategy is to apply a local

transformation $\Lambda_x : \mathcal{D}(\mathcal{H}_A) \mapsto \mathcal{D}(\mathcal{H}_A)$ directly on her share of ρ, then send it to Bob. After that,

$$\rho \mapsto (\Lambda_x \otimes \mathbf{1})\,\rho \equiv \rho_x$$

will be in his possession. It is crucial to enforce that

$$\mathrm{tr}_A(\rho_x) = \mathrm{tr}_A(\rho_{x'}) \quad \forall x, x',$$

a condition remindful of no-signalling, for it subsumes the fact the Alice's operations, being local, cannot affect Bob's marginal state.

On his end, a good selection of a quantum measurement $\mathcal{M} = \{E_b\}_{b=0}^{k-1}$ may provide advantage (over classical encodings) in the task of recovering her choice of x. Naturally, $\sum_b E_b = \mathbf{1}$, and each E_b is a positive semidefinite operator acting on $\mathcal{H}_A \otimes \mathcal{H}_B$, which means to say he measures on both his and her transformed share of ρ. Summing it up, dense coding behaves in $\mathcal{Q}^{\rho}_{d_A,d_A^2,d_A^2,1}$, but we will also discuss some further generalizations.

Many rounds of this protocols allow them to collectively infer the behavior

$$\mathbf{p} = \{p(b \mid x)\}_{b,x} = \{\mathrm{tr}(\rho_x E_b)\}_{b,x},$$

and from it we can assess their performance. To that end, we assume that her choices of x are equiprobable in $\mathcal{X}$ and define the probability of success as

$$p_{\mathrm{suc}} = \frac{1}{N}\sum_{x=0}^{N-1} p(b = x \mid x). \tag{5.1}$$

This is the very same average success probability discussed under the random access coding protocols of sec. 3.2.1, except that we now deal with a single measurement on Bob's side. Differently from the usual dense coding formulation — which requires knowledge on the exact preparations and measurements employed —, computing p_{suc} relies solely observational data. It is thus a device independent figure of merit, completing our device independent formulation of the dense coding protocol.

5.2 Witnessing and self-testing entanglement

Understanding the relationships between some figure of merit and the underlying resources is a commonplace question in device independent scenarios. Important ones for the prepare-and-measure formulation of dense coding are which bounds the amount of entanglement in ρ or its local dimensions imply on p_{suc}, and whether violating them witness or even self-tests some property of ρ.

Brunner et al. [88] considered similar questions in the quantum state discrimination scenario. Similarly to ours, their scenario allowed for a single measurement, and their figure of merit was equivalent to p_{suc}. However, the devices were independent, and their investigation focused on what could be inferred about quantum versus classical preparations. In chap. 3's terminology, they were mostly interested in analysing p_{suc} in $\mathcal{C}_{d<N,N,N,1}$ against $\mathcal{Q}_{d<N,N,N,1}$. Interestingly, they found out both classical and quantum bounds for p_{suc} are equal to d_A/N. Therefore, while that success probability can be used as a dimension witness, it cannot witness the type of preparations.

Our first result generalizes theirs in the following way

Result 5.1 (Schmidt number witness)**.** *Let ρ be a shared resource with Schmidt number s and local dimension d_A on Alice's side. If she chooses one out of N preparations, and Bob performs a single measurement with N outcomes on the joint state, then*

$$p_{suc} \leq \min\left(\frac{d_A s}{N}, 1\right). \tag{5.2}$$

When $N \geq d_A s$, the bound is tight.

Separable states ($s = 1$) thus recover the bound $p_{\text{suc}} \leq d_A/N$ from [88]. For this reason, whenever the transmitted state's dimension is knowingly d_A, a $p_{\text{suc}} > d_A/N$ witness entanglement. Because only local operations are used to prepare the ρ_x, and those cannot increase s, any $p_{\text{suc}} > d_A s/N$ unambiguously certifies that $\rho \notin S_s$, so providing a lower bound on its Schmidt number.

A particular instance of ineq. 5.2 can also be used for self-testing of maximally entangled states.

Result 5.2 (Self-testing maximally entangled states)**.** *In dense coding instances ($N = d_A^2$) of the prepare-and-measure scenario with $s = d_A$, saturation of ineq. 5.2 certifies, up to a local unitary transformation, that the shared state ρ is maximally entangled.*

Interestingly, this has implications for quantum key distribution protocols in the prepare-and-measure dense coding scenario. Suppose Alice does not actually share a maximally entangled pair with Bob, but rather with a third, malicious party that wants to eavesdrop on their communication. Let us suitably call her Eve. Apart from sharing a maximally entangled pair ρ_{AE} with Alice, she intercepts Alice's communication of her share of ρ_x. From result 5.2, she reads out x with $p_{\text{suc}} = 1$, thus perfectly learning Alice's message. Because, at the end of the day, it will be Alice and Bob who share data to infer their p_{suc}, Eve must share a second maximally entangled pair ρ_{EB} with Bob, which she exploits to reencode x and transmit to Bob. In this way, Alice and Bob's p_{suc} will saturate, tricking them into believing they share a maximally entangled pair. As Eve succeeds in eavesdropping without being detected, the prepare-and-measure dense coding protocol is not cryptographically secure.

Historically, dense coding was introduced as a perfect encoding protocol. Relaxing this condition, by letting $p_{\text{suc}} < 1$, can shine more light on the role of entanglement assistance. Sharing pure entangled states, as the following result shows, is always advantageous over having only classical correlations (in the form of separable states).

Result 5.3 (Pure states quantum advantage)**.** *Sharing a pure entangled state, which we write in the Schmidt decomposition $|\psi\rangle = \sum_{i=0}^{s-1} \eta_i |i\rangle \otimes |i\rangle$, the best probability of success in the encoding of $N = d_A^2$ dits is lower bounded as*

$$p_{suc} \geq \min\left(\frac{1+\Gamma}{d_A}, 1\right),$$

where $\Gamma \equiv \sum_{i \neq j} \eta_i \eta_j \geq 0$.

From result 5.1, $N = d_A^2$ with $\rho \in \text{SEP}$ implies that $p_{\text{suc}} \leq 1/d_A$, and as an entangled state is such that $\Gamma > 0$, all pure entangled states provide quantum advantage in the dense coding protocol.

As a corollary, we have an amusing alternative proof to the possibility of perfectly encoding two dits in a qudit: a maximally entangled state has all $\eta_i = 1/\sqrt{d_A}$ (c.f. sec. 1.1), by which $\Gamma = d_A - 1$, thus $p_{\text{suc}} = 1$.

Apropos of mixed states, a similar relation holds. Setting $d_A = d_B = d$ and making use of the

singlet fraction $\zeta(\rho)$, which, in a loose sense, measures how much of a maximally entangled state is in ρ (sec. 1.1), it is true that

Result 5.4 (Singlet fraction bound)**.** *The best probability of success achievable with a resource ρ is lower bounded as*

$$p_{suc} \geq \zeta(\rho). \tag{5.3}$$

Interestingly, this fact closely links our witness to *faithful entangled states*. These are defined to be those states whose entanglement can be certified by a fidelity-based witness, and it was recently proven that a state is faithful if and only if its singlet fraction is greater than $1/d$ [146]. Eq. 5.3 hence implies that any $p_{\text{suc}} > 1/d$ certifies ρ is faithful.

As another application of this same result, consider an isotropic state (sec. 1.1)

$$\chi(\alpha) = \alpha \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| + (1-\alpha)\frac{\mathbf{1}}{d^2},$$

which is separable only in the range $\alpha \leq \frac{1}{d+1}$. Its singlet fraction is

$$\zeta\left[\chi(\alpha)\right] = \alpha + \frac{1-\alpha}{d^2},$$

and monotonically decreases with decreasing α. As

$$\zeta\left[\chi\left(\frac{1}{d+1}\right)\right] = \frac{1}{d}$$

is the classical bound for p_{suc} (result 5.1), it follows that $p_{\text{suc}} \geq \zeta(\rho)$ can witness entanglement for all isotropic states.

As pointed in sec. 1.1, isotropic states are a common benchmark for quantum informational protocols. In particular, there are entanglement witnesses in the Bell nonlocality and quantum steering scenario that can certify a $\chi(\alpha)$ is entangled from some critical visibility α_{crit} upwards. Before contrasting them to our result, it is important to justify whether the comparison is fair. In the case of Bell nonlocality, it is not. Any prepare-and-measure scenario allow for communication, and all our witnesses rely on local dimension bounds, thus being only semi-device independent. Contrastingly, Bell nonlocality is built on a causal structure with no direct relations between the parties, and is fully device independent. Therefore, we expect any witness in this scenario to perform more poorly. The answer for quantum steering is not so clear-cut. While it also forbids communication, it does rely on local tomography for one of the parties. This presupposes full characterization of one measurement device, and hence also knowledge on the state dimension. Nevertheless, as made evident in fig. 5.1 it does perform rather badly in comparison to our singlet fraction witness. The tests used were the Collins-Gisin-Linden-Massar-Popescu (CGLMP) inequality [147] and Wiseman et al.'s truncated harmonic series relation [15] (also stated in sec. 1.1). Neither is proven optimal, and for specific dimensions better ones are known (e.g., [148]).

[**Review!** Na verdade, o Otfried recomendou um teste melhor no caso de steering. Vale a pena atualizar a figura e o texto depois, ou pelo menos comentar sobre?]

Taking first step towards the study of more general entanglement assisted prepare-and-measure scenarios, we extend another result from the state discrimination task studied in [88]. In it, the authors allowed there to be $Y = N(N-1)/2$ dichotomic measurement choices, where N is the number of preparations. Each measurement is labeled as the pair (x, x'), where $x > x'$ with $x, x' \in \{0, \ldots, N-1\}$, and the behaviors now have elements $p\left(b \mid x, y\right)$. Their result states that,

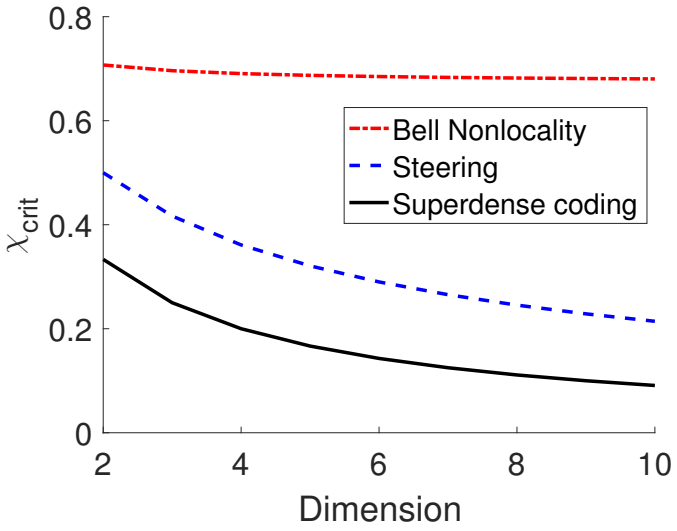


Figure 5.1: Comparing between witness 5.3 to the CGLMP Bell inequality witness [147] and Wiseman et al.'s steering witness [15] for isotropic states. With a dimension assumption, the prepare-and-measure dense coding scenario can witness all entangled isotropic states. See the main text for a discussion on the fairness of this comparison.[**To-Do:** I may or may not remake a prettier figure.]

for uncorrelated parties, the expression

$$V_N \equiv \sum_{x>x'} \left| p\left(b=1 \mid x,(x,x')\right) - p\left(b=1 \mid x',(x,x')\right)\right|^2 \leq \frac{N^2}{2}\left(1-\frac{1}{\min(d_A,N)}\right)$$

is a quantum dimension witness for any communication dimension $d_A < N$, and can also distinguish between quantum and classical systems (in the sense of pair-wise mutually commuting preparations, explained in sec. 3.1.2) for any N which is not a multiple of d. By allowing for an entangled resource of Schmidt number s between the parties, we found the following generalization.

Result 5.5 (Multiple measurements witness)**.** *Given a shared bipartite resource with Schmidt number s, a prepare-and-measure scenario with N d_A-dimensional preparations labeled by $x \in \{0,\ldots,N-1\}$, and $N(N-1)/2$ dichotomic measurements labeled as (x,x') for $x > x'$ is such that*

$$\begin{aligned} V_N \equiv \sum_{x>x'} &\left| p\left(b=1 \mid x,(x,x')\right) - p\left(b=1 \mid x',(x,x')\right)\right|^2 \\ &\leq \frac{N^2}{2}\left(1-\frac{1}{\min(d_A s,N)}\right). \end{aligned}$$

If $s = d_A$, and $N < d_A^2$ or N is an integer multiple of d_A^2, the inequality is tight.

For fixed N and d_A, V_N can witness whether $\rho \in S_s$ or not.

5.3 Optimizing the dense coding probability of success

Up until now, we have focused on what can be inferred about ρ from the observable statistics. In practice, ρ is sometimes treated as a resource to be consumed in the protocol, and it may be of interest to obtain the preparations ρ_x and measurement $\mathcal{M} = \{E_b\}$ that make the best use of it

to achieve the largest p_{suc}. More formally, we are interested in solving

$$\begin{aligned}
&\text{given} && \rho && (5.4a)\\
&\max_{\mathcal{M},\Lambda_x} && \frac{1}{N}\sum_{0}^{N-1} \operatorname{tr}\left[E_x\left(\Lambda_x \otimes \mathbf{1}\right)\rho\right] && (5.4b)\\
&\text{s.t.} && \Lambda_x \in \text{CPTP}, \quad \forall x && (5.4c)\\
& && E_x \succeq 0, \quad \forall x && (5.4d)\\
& && \sum E_x = \mathbf{1}. && (5.4e)
\end{aligned}$$

For an arbitrary ρ, this is a daunting task to approach analytically. Even numerically, the objective function is nonlinear, and optimizing over CPTP constraints is not directly a recognizable constraint. This lends little hope to the problem of efficiently finding global maxima. It is nevertheless possible to formulate an alternated semidefinite optimization procedure (also called *see-saw* optimization) that solves to local extrema.

Using channel-state duality (sec. 1.1), the CPTP map Λ_x can be cast as a bipartite state, for which the constraints of positive semidefinitess and unit-trace are amenable in semidefinite programming. Eq. (5.4) is then equivalent to [**Review!** Should I explain the transformation better?]

$$\begin{aligned}
&\text{given} && \rho && (5.5a)\\
&\max_{\mathcal{M},L_x} && \frac{1}{N}\sum_{0}^{N-1} \operatorname{tr}\left[\left(L_x \otimes \mathbf{1}_B\right)\left(\rho^{\intercal_A} \otimes \mathbf{1}_{A'}\right)\left(\mathbf{1}_A \otimes E_x\right)\right] && (5.5b)\\
&\text{s.t.} && L_x \succeq 0, \quad \forall x && (5.5c)\\
& && \operatorname{tr}_{A'}\left(L_x\right) = \mathbf{1}_A, \quad \forall x && (5.5d)\\
& && E_x \succeq 0, \quad \forall x && (5.5e)\\
& && \sum E_x = \mathbf{1}, && (5.5f)
\end{aligned}$$

where $^{\intercal_A}$ is a partial transposition on the A system and the $L_x \in \mathcal{L}(\mathcal{H}_A \otimes \mathcal{H}_B)$ correspond to the preparations Λ_x. The constraints are all linear matrix inequalities, but the objective function is still nonlinear.

To circumvent that, we sample an initial measurement $\mathcal{M}^0$, fix it for the time being, and run program 5.5 only on the L_x variables. The objective function is then clearly linear, and the last two constraints shall be removed.

With the optimal (w.r.t. $\mathcal{M}_0$) L_x from this first iteration, which we call L_x^1, fixed, optimizing over the measurements is also a semidefinite program. In this second step, the first two constraints may be left out, and we call its result $\mathcal{M}^1$. Together, $\mathcal{M}^1$ and L_x^1 are the result of the first iteration, and provide a lower bound on the best p_{suc}. After N iterations, convergence can be tested through some heuristic criterion, such as checking whether $p_{\text{suc}}\left(\mathcal{M}^N, L_x^N\right) - p_{\text{suc}}\left(\mathcal{M}^{N-1}, L_x^{N-1}\right) \approx 0$. Different samples of $\mathcal{M}_0$ can lead to distinct results. As this is a computationally inexpensive procedure, one can run it for several initial samples and take the best solution. There will be no formal guarantee of achieving a global maximum, but trustable numerical evidence can be reached. And, of course, one might as well start by sampling a L_x^0 and inverting the order of the programs.

To see it in action, consider the Werner states (sec. 1.1) parametrized as

$$\rho_W(\alpha) = \frac{\mathbf{1} - \alpha S}{d^2 - \alpha d}, \tag{5.6}$$

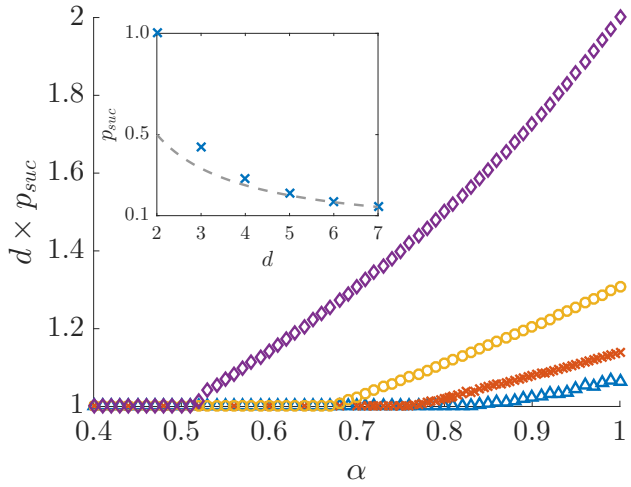


Figure 5.2: Optimal p_{suc}'s obtained for Werner states (5.6) with the alternated optimization procedure over program (5.5). Rhombuses, circles, crosses and triangles correspond to $d = 2, 3, 4$ and 5, respectively. Values are reescaled so that all classical bounds correspond to 1. The inset shows p_{suc} for $alpha = 1$ (crosses) against the classical bound of $1/d$ in dimensions 2 up to 7.[**To-Do:** May remake this figure too.]

where d is the local dimension, $-1 \leq \alpha \leq 1$, and $S = \sum_{i,j=0}^{d-1} |ij\rangle\langle ji|$ is the swap operator. For $N = d^2$ preparations and a measurement with N outcomes, we optimize the dense coding probability of success for $d \in \{2, \ldots, 5\}$ and a linear range over α. Results are shown in fig. 5.2. All values of $\alpha \lesssim (d-1)/d$ saturate the classical (i.e., $s = 1$) bound of $1/d$ from result 5.1, and every larger α violates it. In the inset, a comparison of p_{suc} with the classical bound $1/d$, for $\alpha = 1$, suggests that Werner states of larger dimension provide smaller quantum advantages in dense coding.

5.4 Open questions

Regarding device independent dense coding, an interesting possibility is to generalize its definition by allowing Alice's local dimension d_A to be larger than the actual communicated qudit, i.e., to let $\Lambda_x : \mathcal{D}(d_A) \mapsto \mathcal{D}(d)$, with $d < d_A$. First discussions on that situation were recently started in [134, 101]. As noted in [101], in that case a road worth taking is to investigate if there are witnesses that do not depend on d_A, but only on the communicated state dimension. Application-wise, while the see-saw optimization procedure is reasonably efficient, it would be useful to better understand what types of guarantees regarding global optimality are possible to provide. Secondly, one could think of a conceptual explanation as to why higher dimensional Werner states are increasingly worse in outperforming the classical bound on p_{suc}, and whether this trend is also present in other classes of entangled states.

General entanglement assisted prepare-and-measure scenarios — and especially the one with quantum preparations — are little explored in the literature. Most results hitherto presented are only valid for the $Y = 1$ case, which is of special interest but begs for generalizations. Result 5.5 is a small step in that direction. Simultaneously to our results, Tavakoli et al. provided important results on the same theme [101], but much is still to be done.

Below result 5.2, we discussed how an eavesdropper could intercept Alice's communication while still tricking she and Bob into believing they share a maximally entangled resource. One interest in studying $Y > 1$ witnesses is that they may provide cryptographically secure tests. To see how,

consider a BB84-like protocol [8] where Alice, before sending her qubit, randomly chooses whether to apply a σ_x gate to it. Up to an unmeasurable global phase, $(|\Phi^+\rangle, |\Phi^-\rangle, |\Psi^+\rangle, |\Psi^-\rangle) \xmapsto{\sigma_x}$ $(|\Psi^+\rangle, |\Psi^-\rangle, |\Phi^+\rangle, |\Phi^-\rangle)$. An eavesdropper sharing the entangled resource with Alice and measuring in the Bell basis would correctly get $x = 0$ in the first situation, but wrongly guess $x = 2$ in the latter. She could try switching to a σ_x-transformed Bell measurement, but without information on Alice's decision, this would be unhelpful. Worse than that, after mistaking x, she would send the wrong repreparation to Bob. During the whole protocol, Bob must also be blind to Alice's choice. Let us further suppose that he can choose between a standard Bell measurement or a σ_x-transformed one and that, to make ends meet, Alice sends $M + \delta$ qubits, where M is the intended key size. Afterwards, Alice publicly announces the choices of x and σ_x she made on the extra δ messages. In all rounds when Bob made the correct decision on his measurements, his result should be perfectly correlated with Alice's encoding. If he observes some of them aren't, he will know Eve eavesdropped. While this example is device dependent, it is of interest to find some witness for $Y = 2$ that can device-independently guarantee security in a similar protocol, or prove there is none.

[**Review!** Também tem o ponto do email do Jed, mas eu não entendi ainda o que ele disse lá...]

Appendices

Appendix A

Computational details for chap. 4

In chapter 4 we built linear programs that can be used to certify classicality in the prepare-and-measure scenario. When η — the insphere radius (see fig. 4.1 — approaches 1, this method provides a necessary and sufficient condition for classicality. To make $\eta \to 1$, it is necessary to increase the number of probe measurements (or preparations). Decomposing the deterministic strategies λ into their extremal points results in an exponentially large set, with $N_\lambda \propto B^{dY}$ extremal points. Even for reasonably sized instances, it becomes prohibitively expensive to even enumerate the strategies, let alone to solve the problem. Consequently, regardless of the efficiency of linear programming algorithms (sec. 2.2.1), it the size of our program instance that scales exponentially. This large amount of extremal strategies is the computational bottleneck of our approach. However, we can circumvent this issue using the approach we now describe (a similar procedure, for an analogous method applied to other correlation scenarios, is outlined in [145]).

Our approach is to avoid working on — and even enumerating — all N_λ extremal points at once. Instead, we iteratively explore the deterministic strategy space. To understand how, recall that the maximization programs discussed in chap. 4 search for the optimal weights $\pi(\lambda)$ such that a convex combination of extremal points of the local polytope describe the behavior of our system. The optimization variable, α, controls the amount of noise we have to put into the preparations (or measurements) so that a classical model exists. A solution will always exists, for the worst case $\alpha = 0$ corresponds to maximally mixed preparations (or measurements), which trivially have a classical model. As a result, any solved instance returns us the optimal value α, and the weights $\pi(\lambda)$ used to attain it.

A linear program searches for optimal solutions inside polyhedra. In our instances, the extremal points of the feasible region are the deterministic strategies themselves, and further constraints impose we are working on its convex hull. Carathéodory's Theorem [61] states that at most $d+1$ extremal points are necessary to optimally describe any point of a d-dimensional convex set. This implies that, optimally, most of the $\pi(\lambda)$ found will be zero. We cannot know, beforehand, which points make for an optimal description, so we start by taking $N'_\lambda \gg d+1$ — but much smaller than N_λ — points, and optimizing over them. To set up the next iteration, we take the resulting $\pi(\lambda) = 0$ from this firt step and replace them with previously unexplored deterministic strategies. Then we run the program with these as input. As, at each round, we are keeping all optimal λ from the previous round, the optimal value α will be non-decreasing between iterations. Furthermore, for so large Y that it would be prohibitive to enumerate and keep track of all previously visited strategies, we observe that simply randomly sampling λ's on each run makes our procedure rapidly converge to an α^*, which then assumes a constant value for all subsequent iterations. This result provides a lower bound on the maximum visibility imposed on the preparations such that their

behavior is classical. Naturally, this procedure works for any of the linear programs we have constructed. More precisely, it goes like this

input : N'_λ, $\mathcal{S}$, $\mathcal{M}$, η, optional: t
output: α^*

Let $\{\alpha\}$ and $\{\lambda\}$ be empty lists and $N_0 \longleftarrow N'_\lambda$
while `Stop?`$(\{\alpha\})$ **:**
  $\{\lambda\}$.`append`(`SampleStrategies`(d, B, X, Y, N_0))
  $\alpha^*, \pi(\lambda) \longleftarrow$ `FindClassicalityModel`$(\mathcal{S}, \mathcal{M}, \eta, \{\lambda\}, \text{optional: } t)$
  $\{\alpha\}$.`append`(α^*)
  $\{\lambda\}$.`remove`$(\pi(\lambda) == 0)$
  $N_0 \longleftarrow N'_\lambda -$ `length`$(\{\alpha\})$
end
return α^*

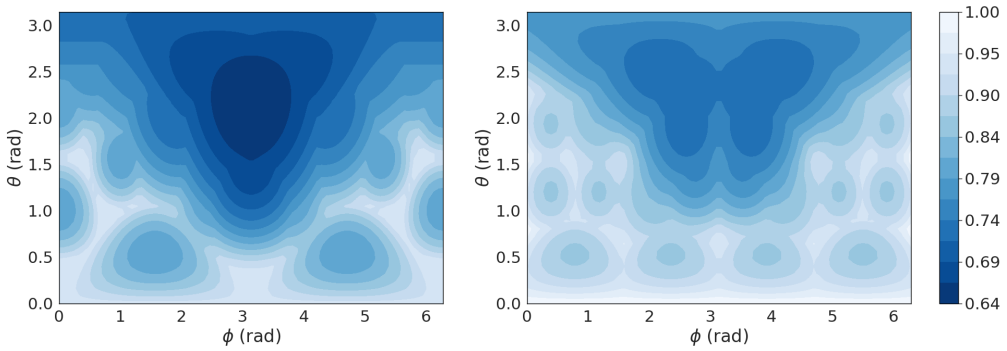
Algorithm 1: Iterating on the deterministic strategies space

where `FindClassicalityModel` is one of programs (4.12), (4.9), (4.12) or (4.9), the parameters d, B, X, Y in `SampleStrategies` can be extracted from the inputs, and `Stop?` implements some convergence criterion.

To illustrate the procedure, consider the problem of certifying that the preparation set $\mathcal{S}(\theta, \phi) = \{\rho_{\mathbf{x}}, \rho_{\mathbf{z}}, \rho_{\mathbf{r}(\theta,\phi)}\}$ is classical for projective measurements (i.e., `FindClassicalityModel` is program (4.6)). Here, $\rho_{\mathbf{v}}$ denotes a qubit state with Bloch vector $\mathbf{v}$, and the unit vector $\mathbf{r}(\theta, \phi)$ is given by its spherical polar and azimuthal angles, respectively.

Arranging our measurements operators as those associated to the vertices of an icosahedron ($Y = 6$ with $\eta \approx 0.79$), we get a fairly small problem that can be solved either directly or iteratively. Fig. A.1(a) shows the result through the iterative procedure, which exactly matches the direct approach. With twice the amount of measurements arranged as the vertices of a rhombicuboctahedron ($\eta \approx 0.86$), it is no longer possible to compute A.1(b) in a direct manner. As a matter of fact, even enumerating all deterministic strategies turns costly. Employing the procedure here described allowed us to compute fig. A.1(b). The advantage of having more measurements then becomes evident, with the rhombicuboctahedron resulting in increased visibilities. We pay for that with more computation time: while the model for each preparation set in fig. A.1(a) takes less than a minute to compute in a standard desktop computer, it takes around 7 minutes for fig. A.1(b) when using fairly robust parameters (likewise for figs. A.1(b), 4.2(b) and 4.3(b)). Implementations for these and all forthcoming applications can be found at a public repository [3].

While our results are fully general, moving on to higher dimensions presents us with two considerable drawbacks. Similarly to the example discussed above, qudits will naturally lead to more extremal points in our classicality polytope, and ultimately to more computation time. Yet another considerable downside is that generating a large depolarized Bloch ball in larger dimensions will be costlier, while also the intuition brought by using polyhedra (with antipodal vertices) as a measurements polytope is lost. Howbeit, picking random 3-qutrits preparations and 12 random projective measurements (average $\eta \approx 0.26$), we found the mean computation time to be around 37 minutes for each preparation set. Increasing the number of measurements and preparations is also possible by paying with more waiting until convergence is reached. In all cases, as each iteration returns a non-decreasing visibility α, one will always obtain lower bounds even if convergence is not quickly attained. These observations lead us to argue that, with a clever selection of probe



(a) Icosahedron (b) Rhombicuboctahedron

Figure A.1: Application of program (4.12) to $\mathcal{S}(\theta, \phi) = \{\rho_{\mathbf{x}}, \rho_{\mathbf{z}}, \rho_{\mathbf{r}(\theta,\phi)}\}$. Levels are the maximum visibility α such that preparation set $\alpha\mathcal{S}(\theta, \phi)$ has a classical model. (a) For the $Y = 6$ icosahedron measurements, $\eta \approx 0.79$, and program (4.12) can be directly applied. (b) A rhombicuboctahedron corresponds to $Y = 12$ projective measurements and $\eta \approx 0.86$. As the number of deterministic strategies scales exponentially, the computation of (b) is only possible by iteratively optimizing over subsets of deterministic strategies, as states in algorithm 1.

measurements, our method could still be useful for future applications even in larger dimensions.

Appendix B

Proofs for chap. 5

Proofs for all results in chap. 5 are hereby provided. To aid in their understanding, we begin by reviewing some useful results.

B.1 Mathematical tools

Generalized Gell-Mann matrices (sec. 1.4) are not the only useful choice of basis for $\mathcal{H}^d$. Another possibility, the *Weyl operator basis* [59] is composed of *discrete Weyl operators* which will be essential for the upcoming proofs. They are defined as

$$W^d_{x_1x_2} = \sum_{m=0}^{d-1} e^{2\pi i x_1 m/d} \left| m \oplus_d x_2 \right\rangle\!\left\langle m \right|, \tag{B.1}$$

where $\oplus_d$ is shorthand for a modulo d sum, and $x_1, x_2 \in \{0, \ldots, d-1\}$. The result of applying $W_{x_1x_2}$ to some basis vector $|j\rangle$ can be interpreted as a displacement to level $|j \oplus_d x_2\rangle$ together with the inclusion of a phase $e^{2\pi i x_1 j/d}$. If the specific values of x_1 and x_2 are unimportant, we will label W simply with x, and we may also omit d. When x_1 and x_2 running on different ranges (as will be used in the proof of result 5.1), we will call W *generalized* Weyl operators. But, for now, let us stick to the usual ones.

Weyl operators have been useful at least since 1993, when they were used by Bennett et al. to devise a d-dimensional quantum states teleportation protocol [20]. Much of their worthiness come from properties such as unitarity ($W_x W_x^\dagger = W_x^\dagger W_x = \mathbf{1}$, as can be verified by direct calculation) and orthonormality: $\mathrm{tr}\left(W_x^\dagger W_y\right) = d\delta_{x,y}$. Proofs can be found in [59], appendix A.3, and in sec. 4.1.2 of [149], presented among other interesting properties.

Every vector $|\psi\rangle \in \mathcal{H}^d \otimes \mathcal{H}^d$ can be obtained from a maximally entangled state, such as $|\Phi^+\rangle = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |j\rangle \otimes |j\rangle$, by means of a *unique* transformation $L_{|\psi\rangle} \otimes \mathbf{1}$, acting trivially on the second subsystem. Importantly, $|\psi\rangle$ is maximally entangled if and only if $L_{|\psi\rangle}$ is unitary [150]. From the uniqueness of the $L_{|\psi\rangle} \otimes \mathbf{1}$, then, there are as many orthogonal maximally entangled states as the dimension of the unitary group, i.e., d^2. Having the d^2 orthonormal unitaries $W_{x_1x_2}$ at hand, it is then true that

$$\left\{ (W_{x_1x_2} \otimes \mathbf{1}) \left| \Phi^+ \right\rangle \right\} \equiv \left\{ \left| \Phi_{x_1x_2} \right\rangle \right\}, \quad \forall x_1, x_2 \in \{0, \ldots, d-1\} \tag{B.2}$$

are maximally entangled and orthogonal preparations. In the prepare-and-measure dense coding scenario, Alice and Bob share a resource ρ that she locally transforms to prepare $\rho_x = (\Lambda_x \otimes \mathbf{1})\rho$.

This is precisely the structure above. Furthermore, orthogonal states are perfectly distinguishable, and we thence should expect them to be good candidates for optimal performance in prepare-and-measure protocols.

With poorly chosen measurements, good preparations are worthless. The d^2-outcome measurement

$$\mathcal{M}^W = \{|\Phi_{x_1x_2}\rangle\langle\Phi_{x_1x_2}|\}_{x_1,x_2=0}^{d-1} \equiv \{M_x^W\}, \tag{B.3}$$

whose effects project precisely in these maximally entangled vectors, will turn useful. To be sure it is well-defined, notice that each of its effects are obviously PSD and, as they are normalized and form a basis, $\sum_{x_1,x_2=0}^{d-1} |\Phi_{x_1x_2}\rangle\langle\Phi_{x_1x_2}| = \mathbf{1}$.

* * *

We end this section with three further results.

Lemma B.1 (Outcome probability upper bound)**.** *When ρ is a density operator and M is a measurement effect, $tr(\rho M) \leq tr(M)$.*

Proof. M is PSD, thus all its eigenvalues are nonnegative and it has a spectral decomposition. Let $M = \sum_i m_i |m_i\rangle\langle m_i|$ be it, and order the eigenvalues as $m_1 \geq \ldots \geq m_d$. Notice, also, that $\rho = \sum_i p_i |p_i\rangle\langle p_i|$, where the p_i form a probability distribution. Now write

$$\mathrm{tr}(\rho M) = \sum_i p_i \sum_j m_j \langle p_i|m_j\rangle^2 .$$

If $\rho = |m_1\rangle\langle m_1|$, then, $\mathrm{tr}(\rho M) = m_1 \leq \mathrm{tr}(M)$. It is now left to prove this choice leads to a maximum of $\mathrm{tr}(\rho M)$. That is indeed the case because, as m_1 is a largest eigenvalue, any convex combination of the m_i is at most as large as m_1. □

Lemma B.2. *If d^2 quantum states $\{\rho_x\}_{x=1}^{d^2}$, where each $\rho_x \in \mathcal{H}^d \otimes \mathcal{H}^d$, do not overlap (i.e., $tr(\rho_x\rho_y) = 0$, $\forall x \neq y$), they must be pure.*

Proof. Each of the ρ_x has a spectral decomposition $\rho_x = \sum_{i=1}^{d^2} p_i \Pi_i$ where each projector is normalized (they correspond to pure states), and $\Pi_i\Pi_j = \delta_{ij}$. Using it,

$$\mathrm{tr}\left(\rho_x\rho_y\right) = \sum_{i,j=1}^{d^2} p_i^{(x)} p_j^{(y)} \mathrm{tr}\left(\Pi_i^{(x)}\Pi_j^{(y)}\right).$$

Suppose that $\mathrm{rank}(\rho_x) = d^2$. Then no $p_i^{(x)} = 0$, and because $\{\Pi_i\}_i$ is a basis for $\mathcal{H}^d \otimes \mathcal{H}^d$, at least some of the traces in the r.h.s. sum will be positive. This happens even if $\mathrm{rank}(\rho_y) = 1$. So trim it down to $\mathrm{rank}(\rho_x) = d^2 - 1$. Now we may as well find *some* ρ_y of unit-rank that do not overlap with ρ_x. But if we find some *other* preparation $\rho_{y'}$ such that $\mathrm{tr}\left(\rho_x\rho_{y'}\right) = 0$, we must have that $\mathrm{tr}\left(\rho_y\rho_{y'}\right) \neq 0$, showing that $\mathrm{rank}(\rho_x) < d^2 - 1$. Proceeding with the analysis, we will conclude that $\mathrm{rank}(\rho_x)$, $\forall x$, must be one. □

Entanglement and its quantification are usually discussed with respect to the local operations and classical communications (LOCC) paradigm, briefly commented in chap. 1.1. Local operations and shared randomness (LOSR) is a subset of LOCC operations where the parties can share a coin, but have no access to fully fledged communication. Both choices lead to the same definition of separability (but to different understandings in other regards [**?**]), and are considered free operations

in entanglement theory. One example of LOSR is the *twirling operation*, where random, local $U\otimes U^*$ operations are applied to a bipartite shared state, resulting in

$$\rho \longmapsto \int dU\, U\otimes U^* \rho\, U^\dagger \otimes U^{*^\dagger}.$$

In this context, the following is an important result.

Lemma B.3 (Conversion to isotropic states through LOSR)**.** *All states ρ with maximal singlet fraction $\zeta(\rho)$ can be converted to an isotropic state $\chi\left[\frac{\zeta(\rho)d^2-1}{d^2-1}\right]$ (eq. (1.8)), with the same singlet fraction, through LOSR.*

Proof. It can be done with a twirling operation. See [44], sec. V. □

B.2 Proofs

For the following proofs, recall that

$$p_{\text{suc}} = \frac{1}{N}\sum_{x=0}^{N-1} \text{tr}\left(\rho_x M_x\right)$$

is the device independent dense coding average success probability, as discussed around eq. (5.1).

Result 5.1 (Schmidt number witness)**.** *Let ρ be a shared resource with Schmidt number s and local dimension d_A on Alice's side. If she chooses one out of N preparations, and Bob performs a single measurement with N outcomes on the joint state, then*

$$p_{suc} \leq \min\left(\frac{d_A s}{N}, 1\right). \tag{5.2}$$

When $N \geq d_A s$, the bound is tight.

Proof. Let the preparations $\{\rho_x\} \in S_s$, where S_s is the set of density operators with Schmidt number no larger than s. As discussed in sec. 1.1, these are convex subsets of $\mathcal{D}\left(\mathcal{H}_A\otimes\mathcal{H}_B\right)$. For any measurement $\mathcal{M}=\{M_x\}$, the average success probability p_{suc} is a convex function in any S_s. Its maximum value must hence occur in extremal points, which are pure states (all mixed states are convex combinations of those). Our interest is in its maximum, thus we can limit our discussion to preparations $\{|\psi_x\rangle\langle\psi_x|\} \in S_s$. Using their Schmidt decomposition,

$$|\psi_x\rangle\langle\psi_x| = \sum_{i,j=0}^{s-1} \eta_i^x \eta_j^x \left|\psi_i^x\right\rangle\left\langle\psi_j^x\right| \otimes |\varphi_i\rangle\langle\varphi_j|.$$

Here, each $|\psi_i^x\rangle \in \mathcal{H}_A$ and the $|\varphi_i^x\rangle \in \mathcal{H}_B$. Only the ψ vectors carry the x index because Alice acts with $(\Lambda_x\otimes\mathbf{1})\rho$ only on *her* share of the resource $\rho\in\mathcal{H}_A\otimes\mathcal{H}_B$ to prepare the ρ_x. In the dense coding protocol (sec. 5.1), we consider $\dim(\mathcal{H}_A)=d_A$, but Bob's local dimension, d_B, may be different. In any case, $s\leq\min(d_A,d_B)$. Because only s vectors $|\varphi_i\rangle$ are needed in the Schmidt decomposition above, let us define $\text{span}\left(\{|\varphi_i\rangle\}_{i=0}^{s-1}\right)\equiv\mathcal{H}_{\text{aux}}$. The $|\psi_x\rangle$ vectors thus belong to an

effective Hilbert space $\mathcal{H}_{\text{eff}} = \mathcal{H}_A \otimes \mathcal{H}_{\text{aux}}$, where $\dim(\mathcal{H}_{\text{eff}}) = d_A s$. That being the case,

$$\begin{aligned} p_{\text{suc}} &= \frac{1}{N}\sum_{x=0}^{N-1} \text{tr}\left(|\psi_x\rangle\langle\psi_x| M_x\right) = \frac{1}{N}\sum_{x=0}^{N-1} \text{tr}_{\text{eff}}\left(|\psi_x\rangle\langle\psi_x| M'_x\right) \\ &\leq \frac{1}{N}\text{tr}_{\text{eff}}\left(\sum_{x=0}^{N-1} M'_x\right) = \frac{1}{N}\text{tr}_{\text{eff}}\left(\mathbf{1}_{\text{eff}}\right) = \frac{d_A s}{N}, \end{aligned}$$

where the M'_x act on $\mathcal{H}_{\text{eff}}$ and lemma B.1 was used for the inequality. This proves the bound.

To further see that this is tight for $N \geq d_A s$, let $N = d_A c$, with $c \geq s$, and recall the shared resource $\rho \in S_s$. Suppose Alice uses it, together with generalized Weyl operators $W^c_{x_1 x_2}$, where $x_1 \in \{0, \ldots, c-1\}$ and $x_2 \in \{0, \ldots, d_A - 1\}$, to prepare the $N = d_A c$ states

$$\left|\Phi^c_{x_1 x_2}\right\rangle = \frac{1}{\sqrt{s}}\sum_{j=0}^{s-1}(W^c_{x_1 x_2} \otimes \mathbf{1})\,|j\rangle \otimes |j\rangle\,. \tag{B.4}$$

These are in the spirit of the states presented in eq. B.2 but, because $s \leq \min(d_A, d_B)$, we may have to build more than s^2 preparations, which will then turn out to be nonorthogonal.

Accordingly, let us also modify the M^W_x measurement effects (eq. B.3) to

$$M^c_{x_1 x_2} = \frac{s}{c}\left|\Phi^c_{x_1 x_2}\right\rangle\!\left\langle\Phi^c_{x_1 x_2}\right| + \frac{1}{N}\sum_{j=s}^{d_B-1} \mathbf{1}_A \otimes |j\rangle\langle j|\,. \tag{B.5}$$

Each of these is a conic combination of two positive semidefinite operators, hence also PSD. To see that they are also complete, notice that summing on the first term leads to

$$\begin{aligned} &\frac{s}{c}\sum_{x_1=0}^{c-1}\sum_{x_2=0}^{d_A-1}\left|\Phi^c_{x_1 x_2}\right\rangle\!\left\langle\Phi^c_{x_1 x_2}\right| \\ &\quad= \frac{s}{c}\sum_{x_1=0}^{c-1}\sum_{x_2=0}^{d_A-1} W^c_{x_1 x_2}\left|\Phi^+\right\rangle\!\left\langle\Phi^+\right|\left(W^c_{x_1 x_2}\right)^\dagger \\ &\quad= \frac{1}{c}\sum_{x_1=0}^{c-1}\sum_{x_2=0}^{d_A-1}\sum_{j,k=0}^{s-1}\sum_{l=0}^{d_A-1} e^{2\pi i x_1 l/c} e^{-2\pi i x_1 l/c}\,|l \oplus_{d_A} x_2\rangle\langle l|\,j\rangle\langle k\,|l\rangle\langle l \oplus_{d_A} x_2| \otimes |j\rangle\langle k| \\ &\quad= \sum_{x_2=0}^{d_A-1}\sum_{j=0}^{s-1}|j \oplus_{d_A} x_2\rangle\langle j \oplus_{d_A} x_2| \otimes |j\rangle\langle j| \\ &\quad= \sum_{j=0}^{s-1}\mathbf{1}_A \otimes |j\rangle\langle j|\,, \end{aligned}$$

making it true that

$$\sum_{x_1=0}^{c-1}\sum_{x_2=0}^{d_A-1} M^c_{x_1 x_2} = \sum_{j=0}^{d_B-1}\mathbf{1}_A \otimes |j\rangle\langle j| = \mathbf{1}_A \otimes \mathbf{1}_B.$$

Eqs. (B.4) and (B.5) are thus a valid quantum implementation such that

$$p_{\text{suc}} = \frac{1}{N}\sum_{x_1=0}^{c-1}\sum_{x_2=0}^{d_A-1}\text{tr}\left(M^c_{x_1 x_2}\left|\Phi^c_{x_1 x_2}\right\rangle\!\left\langle\Phi^c_{x_1 x_2}\right|\right) = \frac{s}{c} = \frac{d_A s}{N}$$

□

Result 5.2 (Self-testing maximally entangled states)**.** *In dense coding instances ($N = d_A^2$) of the*

prepare-and-measure scenario with $s = d_A$, saturation of ineq. 5.2 certifies, up to a local unitary transformation, that the shared state ρ is maximally entangled.

Proof. p_{suc} can only be saturated if the ρ_x are perfectly distinguishable. From lemma B.2, $N = d_A^2$ preparations are pair-wise distinguishable only if they are pure states. Consequently, there must be d_A^2 orthonormal preparations of the form

$$|\psi_x\rangle = \sum_{i=0}^{d_A-1} \eta_i (U_x \otimes \mathbf{1}) |i_A\rangle \otimes |i_B\rangle ,$$

where U are unitaries (c.f. eq. (B.2)). Let us analyze the reduced operator from $\sum_x |\psi_x\rangle\langle\psi_x|$ in two ways. The first tells us that

$$\mathrm{tr}_A \left(\sum_{x=0}^{d_A^2-1} |\psi_x\rangle\langle\psi_x| \right) = \mathrm{tr}_A \left(\mathbf{1}_{d_A^2} \right) = d_A \mathbf{1}_{d_A},$$

and the second that

$$\begin{aligned} \mathrm{tr}_A \left(\sum_{i,j=0}^{d_A-1} \eta_i \eta_j \sum_{x=0}^{d_A^2-1} U_x |i_A\rangle\langle j_A| U_x^\dagger \otimes |i_B\rangle\langle j_B| \right) &= \sum_{i,j=0}^{d_A-1} \eta_i \eta_j \sum_{x=0}^{d_A^2-1} \mathrm{tr} \left(U_x |i_A\rangle\langle j_A| U_x^\dagger \right) \otimes |i_B\rangle\langle j_B| \\ &= d_A^2 \sum_{i=0}^{d_A-1} \eta_i^2 |i_B\rangle\langle i_B| \end{aligned}$$

Together,

$$\mathbf{1}_{d_A} = d_A \sum_{i=0}^{d_A-1} \eta_i^2 |i_B\rangle\langle i_B| \quad \Longrightarrow \quad \eta_i = \frac{1}{\sqrt{d_A}}$$

□

Result 5.3 (Pure states quantum advantage)**.** *Sharing a pure entangled state, which we write in the Schmidt decomposition $|\psi\rangle = \sum_{i=0}^{s-1} \eta_i |i\rangle \otimes |i\rangle$, the best probability of success in the encoding of $N = d_A^2$ dits is lower bounded as*

$$p_{suc} \geq \min \left(\frac{1+\Gamma}{d_A}, 1 \right),$$

where $\Gamma \equiv \sum_{i \neq j} \eta_i \eta_j \geq 0$.

Proof. Prepare

$$|\psi\rangle \longmapsto |\psi_x\rangle = \sum_{i=0}^{s-1} \eta_i (W_x^{d_A} \otimes \mathbf{1}) |i\rangle \otimes |i\rangle$$

and measure

$$\mathcal{M} = \left\{ \frac{1}{d_A} \sum_{m,n=0}^{d_A-1} W_x^{d_A} |m\rangle\langle n| \left(W_x^{d_A}\right)^\dagger \otimes |m\rangle\langle n| \right\}_{x=0}^{d_A^2-1} \equiv \{M_x\}.$$

For some fixed Schmidt rank $s \leq \min(d_A, d_B)$, we can work on an effective space $\mathcal{H}^{d_A} \otimes \mathcal{H}^{d_A}$. In this case, $\mathcal{M}$ is a basis of d_A^2 orthonormal maximally entangled states (c.f. eq. (B.3)), therefore a valid measurement.

With this prescription,

$$\begin{aligned} M_x \left|\psi_x\right\rangle\!\left\langle\psi_x\right| &= \frac{1}{d_A}\left(\sum_{m,n=0}^{d_A-1} W_x^{d_A} \left|m\right\rangle\!\left\langle n\right| (W_x^{d_A})^\dagger \otimes \left|m\right\rangle\!\left\langle n\right|\right)\left(\sum_{i,j=0}^{s-1} \eta_i\eta_j W_x^{d_A} \left|i\right\rangle\!\left\langle j\right| (W_x^{d_A})^\dagger \otimes \left|i\right\rangle\!\left\langle j\right|\right) \\ &= \frac{1}{d_A}\sum_{m=0}^{d_A-1}\sum_{i,j=0}^{s-1} \eta_i\eta_j W_x^{d_A} \left|m\right\rangle\!\left\langle j\right| (W_x^{d_A})^\dagger \otimes \left|m\right\rangle\!\left\langle j\right|, \end{aligned}$$

Whereby

$$\begin{aligned} p_{\text{suc}} &= \frac{1}{d_A^2}\sum_{x=0}^{d_A^2-1} \operatorname{tr}\left(M_x \left|\psi_x\right\rangle\!\left\langle\psi_x\right|\right) \\ &= \frac{1}{d_A^3}\sum_{x=0}^{d_A^2-1}\sum_{i,j=0}^{s-1} \eta_i\eta_j \left\langle j\middle| W_x^\dagger \left(\sum_{l=0}^{d_A-1} \left|l\right\rangle\!\left\langle l\right|\right) W_x \middle| j\right\rangle \\ &= \frac{1}{d_A}\sum_{i,j=0}^{s-1} \eta_i\eta_j = \frac{1}{d_A}\left(\sum_{i=0}^{s-1}\eta_i^2 + \sum_{i\neq j}\eta_i\eta_j\right) = \frac{1+\Gamma}{d_A} \end{aligned}$$

□

Result 5.4 (Singlet fraction bound)**.** *The best probability of success achievable with a resource ρ is lower bounded as*

$$p_{suc} \geq \zeta(\rho). \tag{5.3}$$

Proof. Lemma B.3 lets us restrict the discussion to isotropic states $\chi(\alpha)$ having a singlet fraction $\zeta(\rho)$. Define the preparations

$$\rho_x = W_x\chi(\alpha)W_x^\dagger = (1-\alpha)\frac{1}{d^2} + \alpha W_x \left|\Phi^+\right\rangle\!\left\langle\Phi^+\right| W_x^\dagger,$$

where $\left|\Phi^+\right\rangle$ is a maximally entangled state. Applying the W_x will take it to another maximally entangled state, and the singlet fraction remains unchanged. Probing them with the measurement effects M_x^W defined in (B.3), we get that each

$$\begin{aligned} \operatorname{tr}\left(\rho_x M_x^W\right) &= \frac{1-\alpha}{d^2} + \alpha = \zeta(\rho_x) = \zeta(\rho) \\ \implies \quad p_{\text{suc}} &= \frac{1}{N}\sum_{x=0}^{N-1} \zeta(\rho) = \zeta(\rho) \end{aligned}$$

□

Result 5.5 (Multiple measurements witness)**.** *Given a shared bipartite resource with Schmidt number s, a prepare-and-measure scenario with N d_A-dimensional preparations labeled by $x \in \{0,\ldots,N-1\}$, and $N(N-1)/2$ dichotomic measurements labeled as (x,x') for $x > x'$ is such that*

$$\begin{aligned} V_N \equiv \sum_{x>x'} &\left|p\left(b=1 \mid x,(x,x')\right) - p\left(b=1 \mid x',(x,x')\right)\right|^2 \\ &\leq \frac{N^2}{2}\left(1 - \frac{1}{\min(d_A s, N)}\right). \end{aligned}$$

If $s = d_A$, and $N < d_A^2$ or N is an integer multiple of d_A^2, the inequality is tight.

Proof. Define the *trace distance* as $D(\rho_x, \rho_{x'}) = \frac{1}{2}\|\rho_x - \rho_{x'}\|_1$, where $\|\cdot\|_1$ is the trace norm. A known result says that ([22], sec. 9.2.1)

$$D(\rho_x, \rho_{x'}) = \max_{0 \leq P \leq \mathbf{1}} \operatorname{tr}\left[(\rho_x - \rho'_x)P\right].$$

P can thus be seen as a measurement effect. Intuitively, the trace distance measures the difference between the probabilities of a given result occurring for ρ_x and $\rho_{x'}$. That is way it is interpreted as a measure of distinguishability. It is also convex in $\rho_x - \rho_{x'}$.

So rewrite $V_N = \sum_{x>x'} \left|\operatorname{tr}\left[(\rho_x - \rho_{x'})M_{b=1|x,x'}\right]\right|^2$ and substitute the trace distance, obtaining

$$V_N \leq \sum_{x<x'} |D(\rho_x, \rho_{x'})|^2.$$

The inequality comes from the fact that we are not imposing completeness for each (x, x') measurement, but rather only maximizing over independent effects. As its r.h.s. is convex, the maximal values are at extremal points of its domain. Thus

$$V_N \leq \sum_{x<x'} |D\left(|\psi_x\rangle, |\psi_{x'}\rangle\right)|^2. \tag{B.6}$$

Another notion of distinguishability comes from the *fidelity*

$$F(\rho_x, \rho_{x'}) \equiv \operatorname{tr}\left(\sqrt{\rho_x^{1/2}\rho_{x'}\rho_x^{1/2}}\right).$$

Importantly, the fidelity is related to the trace distance ([22], sec. 9.2.3), and for pure states

$$D\left(|\psi_x\rangle, |\psi_{x'}\rangle\right) = \sqrt{1 - F^2(|\psi_x\rangle, |\psi_{x'}\rangle)} = \sqrt{1 - |\langle\psi_x|\psi_{x'}\rangle|^2}$$

Substituting back into eq. (B.6),

$$\begin{aligned} V_N &\leq \sum_{x<x'} |D\left(|\psi_x\rangle, |\psi_{x'}\rangle\right)|^2 = \sum_{x,x'} 1 - |\langle\psi_x|\psi_{x'}\rangle|^2 \\ &= \frac{N(N-1)}{2} - \frac{1}{2}\left(\sum_{x,x'} |\langle\psi_x|\psi_{x'}\rangle|^2 - N\right) \\ &= \frac{N^2}{2}\left[1 - \operatorname{tr}\left(\Omega^2\right)\right], \end{aligned}$$

with $\Omega = \frac{1}{N}\sum_{x=0}^{N-1} |\psi_x\rangle\langle\psi_x|$. Considering a resource with Schmidt rank s, all preparations can be written as in

$$|\psi_x\rangle = \sum_{i=0}^{s-1} \eta_i (U_x \otimes \mathbf{1}) |i_A\rangle \otimes |i_B\rangle,$$

showing Ω acts on an effective Hilbert space of dimension $d_A s$. Therefore, the purity $\operatorname{tr}\left(\Omega^2\right)$ is lower bounded by $\frac{1}{d_A s}$, and

$$V_N \leq \frac{N^2}{2}\left(1 - \frac{1}{d_A s}\right). \text{[To-Do: and the minimum of } d_A s \text{ and N?]}$$

The last part of the result claims that, for $s = d_A$, this bound can be saturated whenever

$N < d_A^2$ or $N = cd_A^2$ for $c \in \mathbb{Z}$. To see this is true, consider preparations

$$|\psi_x\rangle = \frac{1}{\sqrt{d_A}} \sum_{i=0}^{d_A-1} (U_x \otimes \mathbf{1}) |i\rangle \otimes |i\rangle ,$$

U_x being unitaries. Substituting back into $\Omega = \frac{1}{N} \sum_{x=0}^{N-1} |\psi_x\rangle\langle\psi_x|$,

$$\Omega^2 = \frac{1}{N^2 d_A^2} \sum_{x,x'=0}^{N-1} \sum_{i,j,k=0}^{d_A-1} U_x |i\rangle \langle j|U_x^\dagger U_{x'}|j\rangle |k\rangle U_{x'}^\dagger \otimes |i\rangle\langle k| .$$

From that,

$$\mathrm{tr}\left(\Omega^2\right) = \frac{1}{N^2 d_A^2} \sum_{x,x'=0}^{N-1} \mathrm{tr}\left(U_{x'}^\dagger U_x\right) \mathrm{tr}\left(U_x^\dagger U_{x'}\right) .$$

By hypothesis, $s = d_A$, so if N were d_A^2, we could build d_A^2 orthogonal maximally entangled preparations, for instance by substituting discrete Weyl operators W_x for the U_x, where $x \in \{0, \ldots, d_A^2 - 1\}$. To do that while accounting for N possible preparations, the trick is to distribute the set of N preparations into d_A^2 *collections*, each associated to a Weyl operator. In that case, if $|\psi_x\rangle$ are $|\psi_{x'}\rangle$ are in the same collection, then $|\psi_x\rangle = |\psi_{x'}\rangle$. We label the collections with C, where $C \in \{0, \ldots, d_A^2 - 1\}$, and each C may be associated with several x's.

[**Review!** I made two minor corrections w.r.t. the paper. Is everything right?] To deal with all cases at once, define $c = \lfloor \frac{N}{d_A^2} \rfloor$, so making

$$N = cd_A^2 + N \bmod d_A^2.$$

Then put $c+1$ states into each of $N \bmod d_A^2$ collections, and c states into each of the remaining $d_A^2 - N \bmod d_A^2$. We then have

$$\begin{aligned}\mathrm{tr}\left(\Omega^2\right) &= \frac{1}{N^2 d_A^2} \sum_{x,x\prime=0}^{N-1} d_A^2 \frac{\delta_{C(x),C(x')}}{2} = \frac{1}{N^2} \left[\left(\sum_{x=0}^{N \bmod d_A^2 - 1} c+1 \right) + \sum_{x=N \bmod d_A^2}^{N-1} c \right] \\ &= \frac{1}{N^2} \left[(c+1)N \bmod d_A^2 + c(N - N \bmod d_A^2) \right] = \frac{1}{N^2} (cN + N \bmod d_A^2).\end{aligned}$$

If $N < d_A^2$, then $c = 0$ and we get $\mathrm{tr}\left(\Omega^2\right) = \frac{1}{N}$. Whereas if $N = cd_A^2$, for integer c, then $N \bmod d_A^2 = 0$ and we get $\frac{1}{d_A^2}$. Both situations saturate the bound on V_N. □

Bibliography

[1] Carlos de Gois, George Moreno, Ranieri Nery, Samuraí Brito, Rafael Chaves, and Rafael Rabelo. General method for classicality certification in the prepare and measure scenario. *arXiv preprint arXiv:2101.10459*, 2021.

[2] George Moreno, Ranieri Nery, Carlos de Gois, Rafael Rabelo, and Rafael Chaves. Semi-device-independent certification of entanglement in superdense coding. *Phys. Rev. A*, 103:022426, Feb 2021.

[3] C. de Gois. Code for "General method for classicality certification in the prepare and measure scenario" ([1]). `https://github.com/cgois/pam_classicality`, 2021.

[4] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47:777–780, May 1935.

[5] Erwin Schrödinger. Discussion of probability relations between separated systems. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 31–4, pages 555–563. Cambridge University Press, 1935.

[6] David Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. i. *Phys. Rev.*, 85:166–179, Jan 1952.

[7] J. S. Bell. On the einstein podolsky rosen paradox. *Physics Physique Fizika*, 1:195–200, Nov 1964.

[8] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science*, 560:7–11, 2014.

[9] Artur K. Ekert. Quantum cryptography based on bell's theorem. *Phys. Rev. Lett.*, 67:661–663, Aug 1991.

[10] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, March 1994.

[11] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86:419–478, Apr 2014.

[12] Jonathan Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, Mar 2007.

[13] Martin Plávala. General probabilistic theories: An introduction, 2021.

[14] R. W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A*, 71:052108, May 2005.

[15] H. M. Wiseman, S. J. Jones, and A. C. Doherty. Steering, entanglement, nonlocality, and the einstein-podolsky-rosen paradox. *Phys. Rev. Lett.*, 98:140402, Apr 2007.

[16] Roope Uola, Ana C. S. Costa, H. Chau Nguyen, and Otfried Gühne. Quantum steering. *Rev. Mod. Phys.*, 92:015001, Mar 2020.

[17] D Cavalcanti and P Skrzypczyk. Quantum steering: a review with focus on semidefinite programming. *Reports on Progress in Physics*, 80(2):024001, dec 2016.

[18] Marco Túlio Quintino, Tamás Vértesi, and Nicolas Brunner. Joint measurability, einstein-podolsky-rosen steering, and bell nonlocality. *Phys. Rev. Lett.*, 113:160402, Oct 2014.

[19] Roope Uola, Costantino Budroni, Otfried Gühne, and Juha-Pekka Pellonpää. One-to-one mapping between steering and joint measurability problems. *Phys. Rev. Lett.*, 115:230402, Dec 2015.

[20] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels. *Phys. Rev. Lett.*, 70:1895–1899, Mar 1993.

[21] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Phys. Rev. Lett.*, 69:2881–2884, Nov 1992.

[22] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[23] Stephen Barnett. *Quantum Information*. Oxford University Press, 2009.

[24] Bárbara Amaral, Alexandre Tavares Baraviera, and Marcelo O. Terra Cunha. *Mecânica Quântica para Matemáticos em Formação*. IMPA, 2011.

[25] Mark M Wilde. *Quantum information theory*. Cambridge University Press, 2013.

[26] Benjamin Schumacher and Michael Westmoreland. *Quantum processes systems, and information*. Cambridge University Press, 2010.

[27] E. Schrödinger. Discussion of probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4):555–563, 1935.

[28] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009.

[29] Lane P. Hughston, Richard Jozsa, and William K. Wootters. A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183(1):14–18, 1993.

[30] Marcelo O. Terra Cunha, Jacob A Dunningham, and Vlatko Vedral. Entanglement in single-particle systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2085):2277–2286, 2007.

[31] Reinhard F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phys. Rev. A*, 40:4277–4281, Oct 1989.

[32] Martin B. Plenio and Shashank Virmani. An introduction to entanglement measures. *Quantum Inf. Comput.*, 7(1):1–51, 2007.

[33] Marcelo O. Terra Cunha. Emaranhamento: caracterização, manipulação e consequências. *Universidade Federal de Minas Gerais*, 2005.

[34] Dagmar Bruß. Characterizing entanglement. *Journal of Mathematical Physics*, 43(9):4237–4251, 2002.

[35] M. A. Nielsen. Conditions for a class of entanglement transformations. *Phys. Rev. Lett.*, 83:436–439, Jul 1999.

[36] Barbara M. Terhal and Paweł Horodecki. Schmidt number for density matrices. *Phys. Rev. A*, 61:040301, Mar 2000.

[37] Asher Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, Aug 1996.

[38] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996.

[39] Rafael Luiz da Silva Rabelo. Não-localidade quântica: matemática e fundamentos. *Universidade Federal de Minas Gerais*, 2011.

[40] Antonio Acín, Nicolas Gisin, and Benjamin Toner. Grothendieck's constant and local models for noisy entangled quantum states. *Phys. Rev. A*, 73:062105, Jun 2006.

[41] T. Vértesi. More efficient bell inequalities for werner states. *Phys. Rev. A*, 78:032112, Sep 2008.

[42] Flavien Hirsch, Marco Túlio Quintino, Tamás Vértesi, Miguel Navascués, and Nicolas Brunner. Better local hidden variable models for two-qubit Werner states and an upper bound on the Grothendieck constant $K_G(3)$. *Quantum*, 1:3, April 2017.

[43] Jonathan Barrett. Nonsequential positive-operator-valued measurements on entangled mixed states do not always violate a bell inequality. *Phys. Rev. A*, 65:042302, Mar 2002.

[44] Michał Horodecki and Paweł Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Phys. Rev. A*, 59:4206–4216, Jun 1999.

[45] Marco Túlio Quintino, Tamás Vértesi, Daniel Cavalcanti, Remigiusz Augusiak, Maciej Demianowicz, Antonio Acín, and Nicolas Brunner. Inequivalence of entanglement, steering, and bell nonlocality for general measurements. *Phys. Rev. A*, 92:032107, Sep 2015.

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

[47] Christopher J. Wood, Jacob D. Biamonte, and David G. Cory. Tensor networks and graphical calculus for open quantum systems, 2015.

[48] A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.

[49] Min Jiang, Shunlong Luo, and Shuangshuang Fu. Channel-state duality. *Phys. Rev. A*, 87:022310, Feb 2013.

[50] Roope Uola, Tobias Moroder, and Otfried Gühne. Joint measurability of generalized measurements implies classicality. *Phys. Rev. Lett.*, 113:160403, Oct 2014.

[51] Marco Túlio Quintino, Joseph Bowles, Flavien Hirsch, and Nicolas Brunner. Incompatible quantum measurements admitting a local-hidden-variable model. *Phys. Rev. A*, 93:052115, May 2016.

[52] Flavien Hirsch, Marco Túlio Quintino, and Nicolas Brunner. Quantum measurement incompatibility does not imply bell nonlocality. *Phys. Rev. A*, 97:012129, Jan 2018.

[53] Erika Bene and Tamás Vértesi. Measurement incompatibility does not give rise to bell violation in general. *New Journal of Physics*, 20(1):013021, jan 2018.

[54] Teiko Heinosaari, Takayuki Miyadera, and Mário Ziman. An invitation to quantum incompatibility. *Journal of Physics A: Mathematical and Theoretical*, 49(12):123001, feb 2016.

[55] Leonardo Guerini. Simulating quantum measurements and quantum correlations. *Universidade Federal de Minas Gerais*, 2018.

[56] Leonardo Guerini, Jessica Bavaresco, Marcelo Terra Cunha, and Antonio Acín. Operational framework for quantum measurement simulability. *Journal of Mathematical Physics*, 58(9):092102, 2017.

[57] Erkka Haapasalo, Teiko Heinosaari, and Juha-Pekka Pellonpää. Quantum measurements on finite dimensional systems: relabeling and mixing. *Quantum Information Processing*, 11(6):1751–1763, 2012.

[58] Michał Oszmaniec, Leonardo Guerini, Peter Wittek, and Antonio Acín. Simulating positive-operator-valued measures with projective measurements. *Phys. Rev. Lett.*, 119:190501, Nov 2017.

[59] Reinhold A Bertlmann and Philipp Krammer. Bloch vectors for qudits. *Journal of Physics A: Mathematical and Theoretical*, 41(23):235303, may 2008.

[60] Gen Kimura. The bloch vector for n-level systems. *Physics Letters A*, 314(5):339–349, 2003.

[61] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, 1970.

[62] Branko Grünbaum. *Convex Polytopes*. Springer New York, 2003.

[63] Günter M. Ziegler. *Lectures on Polytopes*. Springer New York, 1995.

[64] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004.

[65] Stefan LÃ¶rwald and Gerhard Reinelt. Panda: a software for polyhedral transformations. *EURO Journal on Computational Optimization*, pages 1–12, 2015.

[66] lrs home page. =http://cgm.cs.mcgill.ca/ avis/C/lrs.html.

[67] cddlib repository. =https://github.com/cddlib/cddlib.

[68] Roope Uola, Tristan Kraft, Jiangwei Shang, Xiao-Dong Yu, and Otfried Gühne. Quantifying quantum resources with conic programming. *Phys. Rev. Lett.*, 122:130404, Apr 2019.

[69] Christos H Papadimitriou and Kenneth Steiglitz. *Combinatorial optimization: algorithms and complexity*. Courier Corporation, 1998.

[70] Jiri Matousek and Bernd Gärtner. *Understanding and using linear programming.* Springer Science & Business Media, 2007.

[71] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, March 1996.

[72] Bernd Gärtner and Jiri Matousek. *Approximation algorithms and semidefinite programming.* Springer Science & Business Media, 2012.

[73] Glpk (gnu linear programming kit) homepage. =https://www.gnu.org/software/glpk/.

[74] LLC Gurobi Optimization. Gurobi optimizer reference manual.

[75] MOSEK ApS. Mosek optimizer documentation.

[76] J. Lofberg. Yalmip : a toolbox for modeling and optimization in matlab. In *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, pages 284–289, 2004.

[77] Michael Grant and Stephen Boyd. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

[78] Guillaume Sagnol and Maximilian Stahlberg. Picos: A python interface to conic optimization solvers.

[79] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016.

[80] Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Phys. Rev. Lett.*, 98:010401, Jan 2007.

[81] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, jul 2008.

[82] Cynthia Vinzant. WHAT IS...a spectrahedron? *Notices of the American Mathematical Society*, 61(5):1, May 2014.

[83] SDPA Project. Sdpa (semidefinite programming algorithms).

[84] Hans D. Mittelmann. Several sdp-codes on sparse and other sdp problems.

[85] Teiko Heinosaari, Jukka Kiukas, and Daniel Reitzner. Noise robustness of the incompatibility of quantum measurements. *Phys. Rev. A*, 92:022115, Aug 2015.

[86] Sébastien Designolle, Máté Farkas, and Jędrzej Kaniewski. Incompatibility robustness of quantum measurements: a unified framework. *New Journal of Physics*, 21(11):113053, nov 2019.

[87] Claudio Carmeli, Teiko Heinosaari, and Alessandro Toigo. Quantum random access codes and incompatibility of measurements. *EPL (Europhysics Letters)*, 130(5), June 2020.

[88] Nicolas Brunner, Miguel Navascués, and Tamás Vértesi. Dimension Witnesses and Quantum State Discrimination. *Physical Review Letters*, 110(15), April 2013.

[89] Joseph Bowles, Nicolas Brunner, and Marcin Pawłowski. Testing dimension and nonclassicality in communication networks. *Physical Review A*, 92(2), August 2015.

[90] Davide Poderini, Samuraí Brito, Ranieri Nery, Fabio Sciarrino, and Rafael Chaves. Criteria for nonclassicality in the prepare-and-measure scenario. *Physical Review Research*, 2(4), October 2020.

[91] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, October 2009.

[92] Marcin Pawłowski and Valerio Scarani. Information Causality. *arXiv:1112.1142 [quant-ph]*, December 2011. arXiv: 1112.1142.

[93] Nicholas Harrigan, Terry Rudolph, and Scott Aaronson. Representing probabilistic data via ontological models, 2008.

[94] Ernesto F. Galvão. Economical ontological models for discrete quantum systems. *Phys. Rev. A*, 80:022106, Aug 2009.

[95] Julio I. de Vicente. A general bound for the dimension of quantum behaviours in the prepare-and-measure scenario. *Journal of Physics A: Mathematical and Theoretical*, 52(9):095304, February 2019. Publisher: IOP Publishing.

[96] Julio I. de Vicente. Shared randomness and device-independent dimension witnessing. *Physical Review A*, 95(1):012340, January 2017.

[97] Michele Dall'Arno, Elsa Passaro, Rodrigo Gallego, and Antonio Acin. Robustness of Device Independent Dimension Witnesses. *Physical Review A*, 86(4):042312, October 2012. arXiv: 1207.2574.

[98] Arthur Fine. Hidden variables, joint probability, and the bell inequalities. *Phys. Rev. Lett.*, 48:291–295, Feb 1982.

[99] Valerio Scarani. *Bell Nonlocality*. Oxford University Press, 1 edition, August 2019.

[100] Rodrigo Gallego, Nicolas Brunner, Christopher Hadley, and Antonio Acín. Device-Independent Tests of Classical and Quantum Dimensions. *Physical Review Letters*, 105(23):230501, November 2010.

[101] Armin Tavakoli, Jef Pauwels, Erik Woodhead, and Stefano Pironio. Correlations in entanglement-assisted prepare-and-measure scenarios. *arXiv:2103.10748 [quant-ph]*, March 2021. arXiv: 2103.10748.

[102] Marcin Pawłowski and Marek Żukowski. Entanglement-assisted random access codes. *Physical Review A*, 81(4):042326, April 2010.

[103] Leonardo Guerini, Marco Túlio Quintino, and Leandro Aolita. Distributed sampling, quantum communication witnesses, and measurement incompatibility. *Physical Review A*, 100(4):042308, 2019. arXiv: 1904.08435.

[104] Jamie Sikora, Antonios Varvitsiotis, and Zhaohui Wei. Device-independent dimension tests in the prepare-and-measure scenario. *Physical Review A*, 94(4), October 2016. Publisher: American Physical Society.

[105] Joseph Bowles, Marco Túlio Quintino, and Nicolas Brunner. Certifying the Dimension of Classical and Quantum Systems in a Prepare-and-Measure Scenario with Independent Devices. *Physical Review Letters*, 112(14), April 2014.

[106] Stephanie Wehner, Matthias Christandl, and Andrew C. Doherty. A lower bound on the dimension of a quantum system given measured data. *Physical Review A*, 78(6), December 2008. arXiv: 0808.3960.

[107] Martin Hendrych, Rodrigo Gallego, Michal Mičuda, Nicolas Brunner, Antonio Acín, and Juan P. Torres. Experimental estimation of the dimension of classical and quantum systems. *Nature Physics*, 8(8), August 2012.

[108] Johan Ahrens, Piotr Badziag, Adán Cabello, and Mohamed Bourennane. Experimental device-independent tests of classical and quantum dimensions. *Nature Physics*, 8(8):592–595, 2012.

[109] Vincenzo D'Ambrosio, Fabrizio Bisesto, Fabio Sciarrino, Johanna F. Barra, Gustavo Lima, and Adán Cabello. Device-Independent Certification of High-Dimensional Quantum Systems. *Physical Review Letters*, 112(14):140503, April 2014.

[110] Armin Tavakoli, Jędrzej Kaniewski, Tamás Vértesi, Denis Rosset, and Nicolas Brunner. Self-testing quantum states and measurements in the prepare-and-measure scenario. *Physical Review A*, 98(6), December 2018.

[111] Máté Farkas and Jędrzej Kaniewski. Self-testing mutually unbiased bases in the prepare-and-measure scenario. *Physical Review A*, 99(3), March 2019.

[112] Shi-Hui Wei, Fen-Zhuo Guo, Xin-Hui Li, and Qiao-Yan Wen. Robustness self-testing of states and measurements in the prepare-and-measure scenario with \textdollar3\textbackslashto 1\textdollar random access code. *Chinese Physics B*, 28(7), July 2019. Publisher: IOP Publishing.

[113] Piotr Mironowicz and Marcin Pawłowski. Experimentally feasible semi-device-independent certification of four-outcome positive-operator-valued measurements. *Phys. Rev. A*, 100:030301, Sep 2019.

[114] Armin Tavakoli, Massimiliano Smania, Tamás Vértesi, Nicolas Brunner, and Mohamed Bourennane. Self-testing nonprojective quantum measurements in prepare-and-measure experiments. *Science Advances*, 6(16), April 2020.

[115] Nikolai Miklin and Michał Oszmaniec. A universal scheme for robust self-testing in the prepare-and-measure scenario. *Quantum*, 5:424, Apr 2021.

[116] Miguel Navascués and Tamás Vértesi. Bounding the Set of Finite Dimensional Quantum Correlations. *Physical Review Letters*, 115(2), July 2015.

[117] Miguel Navascués, Adrien Feix, Mateus Araújo, and Tamás Vértesi. Characterizing finite-dimensional quantum behavior. *Physical Review A*, 92(4), October 2015.

[118] Karthik Mohan, Armin Tavakoli, and Nicolas Brunner. Sequential random access codes and self-testing of quantum measurement instruments. *New Journal of Physics*, 21(8):083034, August 2019.

[119] Nikolai Miklin, Jakub J. Borkała, and Marcin Pawłowski. Semi-device-independent self-testing of unsharp measurements. *Physical Review Research*, 2(3):033014, July 2020.

[120] Yukun Wang, Ignatius William Primaatmaja, Emilien Lavie, Antonios Varvitsiotis, and Charles Ci Wen Lim. Characterising the correlations of prepare-and-measure quantum networks. *npj Quantum Information*, 5(1), February 2019.

[121] Marcin Pawlowski and Nicolas Brunner. Semi-device-independent security of one-way quantum key distribution. *Physical Review A*, 84(1), July 2011. arXiv: 1103.4105.

[122] Christian Schmid, Pavel Trojek, Mohamed Bourennane, Christian Kurtsiefer, Marek Żukowski, and Harald Weinfurter. Experimental single qubit quantum secret sharing. *Phys. Rev. Lett.*, 95:230505, Dec 2005.

[123] Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Reviews of Modern Physics*, 82(1):665–698, March 2010.

[124] Andris Ambainis, Ashwin Nayak, Ammon Ta-Shma, and Umesh Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the Thirty-First Annual ACM Symposium on Theory of Computing*, STOC '99, pages 376–383, New York, NY, USA, 1999. Association for Computing Machinery.

[125] Andris Ambainis, Debbie Leung, Laura Mancinska, and Maris Ozols. Quantum Random Access Codes with Shared Randomness. *arXiv:0810.2937 [quant-ph]*, June 2009. arXiv: 0810.2937.

[126] Stephen Wiesner. Conjugate coding. *ACM Sigact News*, 15(1):78–88, 1983.

[127] Ashwin Nayak. Optimal lower bounds for quantum automata and random access codes. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 369–376. IEEE, 1999.

[128] Andris Ambainis, Ashwin Nayak, Amnon Ta-Shma, and Umesh Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM (JACM)*, 49(4):496–511, 2002.

[129] Masahito Hayashi, Kazuo Iwama, Harumichi Nishimura, Rudy Raymond, and Shigeru Yamashita. (4, 1)-quantum random access coding does not exist—one qubit is not enough to recover one of four bits. *New Journal of Physics*, 8(8):129, 2006.

[130] Robert W Spekkens, Daniel H Buzacott, Anthony J Keehn, Ben Toner, and Geoff J Pryde. Preparation contextuality powers parity-oblivious multiplexing. *Physical review letters*, 102(1):010401, 2009.

[131] Pierre-Emmanuel Emeriau, Mark Howard, and Shane Mansfield. Quantum advantage in information retrieval, 2020.

[132] Alexander Semenovich Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973.

[133] Adriano Barenco and Artur K. Ekert. Dense coding based on quantum entanglement. *Journal of Modern Optics*, 42(6):1253–1259, 1995.

[134] Ashwin Nayak and Henry Yuen. Rigidity of superdense coding, 2020.

[135] D. Cavalcanti, L. Guerini, R. Rabelo, and P. Skrzypczyk. General method for constructing local hidden variable models for entangled quantum states. *Phys. Rev. Lett.*, 117:190401, Nov 2016.

[136] Flavien Hirsch, Marco Túlio Quintino, Tamás Vértesi, Matthew F. Pusey, and Nicolas Brunner. Algorithmic construction of local hidden variable models for entangled quantum states. *Phys. Rev. Lett.*, 117:190402, Nov 2016.

[137] Sandu Popescu. Bell's inequalities and density matrices: Revealing "hidden" nonlocality. *Phys. Rev. Lett.*, 74:2619–2622, Apr 1995.

[138] Flavien Hirsch, Marco Túlio Quintino, Joseph Bowles, and Nicolas Brunner. Genuine hidden quantum nonlocality. *Phys. Rev. Lett.*, 111:160402, Oct 2013.

[139] Rodrigo Gallego, Lars Erik Würflinger, Rafael Chaves, Antonio Acín, and Miguel Navascués. Nonlocality in sequential correlation scenarios. *New Journal of Physics*, 16(3):033037, 2014.

[140] Miguel Navascués and Tamás Vértesi. Activation of nonlocal quantum resources. *Phys. Rev. Lett.*, 106:060403, Feb 2011.

[141] Carlos Palazuelos. Superactivation of quantum nonlocality. *Phys. Rev. Lett.*, 109:190401, Nov 2012.

[142] Daniel Cavalcanti, Mafalda L Almeida, Valerio Scarani, and Antonio Acin. Quantum networks reveal quantum nonlocality. *Nature communications*, 2(1):1–6, 2011.

[143] Joseph Bowles, Flavien Hirsch, and Daniel Cavalcanti. Single-copy activation of bell nonlocality via broadcasting of quantum states, 2020.

[144] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, Jan 2006.

[145] Mathieu Fillettaz, Flavien Hirsch, Sébastien Designolle, and Nicolas Brunner. Algorithmic construction of local models for entangled quantum states: Optimization for two-qubit states. *Phys. Rev. A*, 98:022115, Aug 2018.

[146] Otfried Gühne, Yuanyuan Mao, and Xiao-Dong Yu. Geometry of faithful entanglement. *Phys. Rev. Lett.*, 126:140503, Apr 2021.

[147] Daniel Collins, Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu. Bell inequalities for arbitrarily high-dimensional systems. *Phys. Rev. Lett.*, 88:040404, Jan 2002.

[148] Péter Diviánszky, Erika Bene, and Tamás Vértesi. Qutrit witness from the grothendieck constant of order four. *Phys. Rev. A*, 96:012113, Jul 2017.

[149] John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[150] K. G. H. Vollbrecht and R. F. Werner. Why two qubits are special. *Journal of Mathematical Physics*, 41(10):6772–6782, 2000.