# On the optimization of nonlocality proofs in quantum statistics

**Stefan Zohren**

*Master Thesis in Mathematical Sciences*
*Utrecht University*
*Supervised by Prof. Dr. Richard D. Gill*
*August 23, 2007*

**Universiteit Utrecht**

# Preface

This master thesis is centered around the the mathematical framework of quantum probability and quantum statistics. Quantum statistics is the axiomatization of the physical theory of quantum mechanics pioneered by von Neumann in the 1930s. In contrast to many other physical theories which are described by deterministic models quantum mechanics is a stochastic theory. In fact, quantum probability can be viewed as an extension to Kolomogorov's "classical" probability theory.

Especially in the last decade, due to first physical realizations of quantum computational systems and the rise of quantum information theory, the subject of quantum statistics became more and more important and many applications for mathematical statisticians and probabilists opened up.

In this thesis we will be concerned with the optimization of so-called nonlocality proofs which are methods to show the "non-classicality" of certain probability distributions within the framework of quantum statistics. In particular, one is interested in measuring the statistical strength of such nonlocality proofs, sometimes called Bell tests. One of the main results of this thesis is the analysis of certain Bell tests where the corresponding measure space is described by an infinite dimensional separable Hilbert space, corresponding to infinitely many possible outcomes. In particular, this gives numerical evidence for a new quantum Bell inequality describing the boundary of the space of quantum probability distributions for the considered setting. These results have been published in the following letter, S. Zohren and R. D. Gill, *"Maximal violation of the Collins-Gisin-Linden-Massar-Popescu inequality for infinite dimensional states"* submitted to *Phys. Rev. Lett.*.

# Acknowledgments

The author would like to thank Prof. Dr. Richard D. Gill for supervising this research and for the enjoyable discussions related to it. Further, he is thankful for the invitation to the Thiele Center For Applied Mathematics in Natural Science at Århus University and the opportunity to give a talk in the "Research Kitchen on Quantum Statistics Meeting", 1-3 November 2006. The author also wants to thank the participants of the Imperial College Relativity Lunch Meetings for their comments on this work.

# Contents

# List of Figures

# Chapter 1

# Introduction

This thesis deals with the mathematical framework of quantum statistics. Before going into further details let us first clarify what we exactly mean by *quantum statistics*, namely, statistical interference for stochastic models for observational data obtained from quantum mechanical systems. Quantum mechanics is one of the best tested scientific theories describing a wide range of physical phenomena mostly at very short scales, such as sub-atomic scales. The mathematical framework underlying this theory has been developed by von Neumann [1] in the 1930s and is sometimes referred to as *quantum probability theory*, in contrast to the axiomatization of modern probability theory developed by Kolmogorov [2] shortly after to which we simply refer as *classical probability theory*. Whereas the framework of quantum statistics is well understood and its predictions match the experimental data perfectly, there are still many open questions regarding the interpretation and meaning of quantum mechanics. Especially the "nonclassical" aspects of quantum statistics are very counterintuitive and are still under debate. A famous example is Einstein, Podolsky and Rosen's (EPR) [3] notion of "elements of reality" questioning the completeness of quantum mechanics, as it could not be accounted for by any theory which defines its physical objects using *local* states. It was first shown by Bell [4] that statistical results predicted by quantum mechanics for measurements of spacelike separated perties cannot be reproduced by *local realistic theories*. The intuitive idea of local realistic theories is the condition that spacelike separated regions cannot influence each other. To make the above a bit clearer let us discuss it in the context considered by EPR and Bell. The scenario is shown in Figure 1.1. A source produces two particles which are send to two observers which in the literature are usually called Alice and Bob, or simply $\mathscr{A}$ and $\mathscr{B}$. Now Alice can perform a measurement $X$ on its particle obtaining

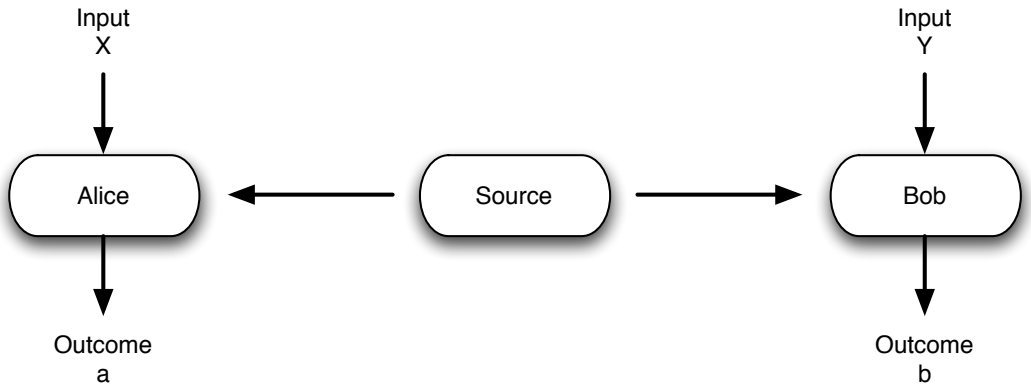


Figure 1.1: Typical Bell setting for two parties, called Alice and Bob: A source sends two particles to Alice and Bob. Alice can perform a measurement $X$ on its particle obtaining an outcome $a$ and Bob can independently perform a measurement $Y$ on its particle obtaining an outcome $b$.

an outcome $a$ and Bob can independently perform a measurement $Y$ on its particle obtaining an outcome $b$. The result of this experiment is characterized by the joint probability distribution $\mathrm{p}_{ab|XY}$, i.e. Alice measuring $a$ and Bob $b$ conditioned on Alice has chosen measurement $X$ and Bob $Y$. In general both measurement outcomes are not independent. i.e.

$$\mathrm{p}_{ab|XY} \neq \mathrm{p}_{a|X}\mathrm{p}_{b|Y} \tag{1.1}$$

This is due to the common past both particles have. If we would describe the common past by a variable $\lambda$ then however our notion of local realism requires that both measurements are independent conditioned on there common past described by $\lambda$, i.e.

$$P(a,b|X,Y,\lambda) = P(a|X,\lambda)P(b|Y,\lambda). \tag{1.2}$$

The parameter $\lambda$ is called a hidden variable, since it describes some hidden properties of the particle. Therefore, if we introduce a probability distribution $q(\lambda)$ for $\lambda$ we get the average

$$\mathrm{p}_{ab|XY} = \int d\lambda q(\lambda)P(a|X,\lambda)P(b|Y,\lambda), \tag{1.3}$$

which holds for any local realistic model, i.e. a model in which spacelike separated regions cannot influence each other.

In what sense quantum mechanics conflicts with the notion of local realism we want to discuss in the following. It was Bell who showed first that joint probability distributions arise from quantum mechanics that violate inequalities, now called *Bell inequalities*, which must hold for any local realistic model. In the following we want to demonstrate this using a certain Bell inequality introduced in [5], the so-called CHSH inequality. Therefore we consider the most simple case where Alice and Bob can choose between two measurements, $X_0, X_1$ and $Y_0, Y_1$, all having two possible outcomes $a, b = 0, 1$. As will be proven in Section 3.3.2 any local realistic joint probability distribution for this setting obeys the following inequality,

$$\begin{aligned} &\Pr(X_0 = Y_0) + \Pr(X_0 = Y_1) + \Pr(X_1 = Y_0) - \Pr(X_1 = Y_1) - \\ &- \Pr(X_0 \neq Y_0) - \Pr(X_0 \neq Y_1) - \Pr(X_1 \neq Y_0) + \Pr(X_1 \neq Y_1) \quad \leq \quad 2, \end{aligned} \tag{1.4}$$

where $\Pr\left(X_i = Y_j\right) = \mathrm{p}_{00|X_iY_j} + \mathrm{p}_{11|X_iY_j}$ and $\Pr\left(X_i \neq Y_j\right) = \mathrm{p}_{01|X_iY_j} + \mathrm{p}_{10|X_iY_j}$.

On the other hand we can consider the two particles to be spin-half particles send to Alice and Bob which are prepared in a so-called singlet state with spin alignment axis perpendicular to the direction of sending. Let Alice's measurements being spin measurements at angles 0 and $\pi/2$ form the spin alignment axis and Bob's measurements being spin measurements at $\pi/4$ and $-\pi/4$, then by using the machinery of quantum probability the joint probabilities can be derived to be

$$\mathrm{p}_{ab|X_0Y_0} = \mathrm{p}_{ab|X_0Y_1} = \mathrm{p}_{ab|X_1Y_0} = \begin{cases} \frac{1}{2}\cos^2\left(\frac{\pi}{8}\right) & \text{if } a = b, \\ \frac{1}{2}\sin^2\left(\frac{\pi}{8}\right) & \text{if } a \neq b, \end{cases} \tag{1.5}$$

and

$$\mathrm{p}_{ab|X_1Y_1} = \begin{cases} \frac{1}{2}\sin^2\left(\frac{\pi}{8}\right) & \text{if } a = b, \\ \frac{1}{2}\cos^2\left(\frac{\pi}{8}\right) & \text{if } a \neq b. \end{cases} \tag{1.6}$$

Hence, if we insert this probabilities into (1.4) we see that the left-hand-side equates to $2\sqrt{2}$ which violates the CHSH inequality. This shows that quantum mechanics violates local realism! We call this a *nonlocality proof*. The violation of Bell inequalities by certain quantum correlations can be seen as a nonclassical property of those correlations. This "quantum nonclassicality" has its roots in so-called quantum entanglement which is an important concept of quantum information theory [6].

The fact that quantum mechanics as a theory of nature does not respect the very fundamental notion of local realism is very controversial and might suggest

that quantum mechanics is not correct. However, inequalities like the CHSH inequality allows one in principle to experimentally test quantum mechanics versus local realism. And indeed, experiments [7, 8, 9, 10] have shown the violation of such Bell inequalities, ruling out (modulo certain loopholes) the possibility of finding a local realistic model alternative to quantum mechanics.

The remainder of this thesis is structured as follows: In Chapter 2 we introduce the reader to the mathematical framework of quantum statistics. All essential concepts will be defined and at the end of the chapter clarified by applying them to a simple example.

The following chapter (Chapter 3) discusses in more detail the aspect of quantum nonlocality in quantum statistics. In the first part of this chapter the basic properties of local realistic joint probability distributions will be given. Later certain Bell inequalities will be derived from this, thereunder the already discussed CHSH inequality and the CGLMP inequality, a generalization of the CHSH inequality for arbitrary many outcomes. In the second part of this chapter properties of the joint probability distributions arising from quantum probability theory will be given and a nonlocality proof will be defined. Further, an inequality for the quantum joint probability distributions for the case of the CHSH setting, the so-called Tsirelson inequality, will be derived.

Chapter 4 contains the main new results of this thesis. It discusses the optimization of nonlocality proofs in particular of the CGLMP setting; first with respect to statistical strength defined in terms of Kullback-Leibler divergence and second in terms of total variation distance, i.e. maximal violation of Bell inequalities. It is shown that whereas the optimal measurements are the same in both cases, the optimal states are not. In particular, numerical analysis suggests that the optimal states in both cases are not the maximally entangled state. On evidence of numerical data, an inequality for the quantum joint probability distributions for the case of the CGLMP setting with infinitely many outcomes is conjectured. At the end of this chapter a comparison of Kullback-Leibler divergence and total variation distance is given.

Finally, Chapter 5 gives a conclusion and summary of the results. Additional information on entropy measures in classical and in quantum information theory is provided in the Appendix.

# Chapter 2

# An invitation to quantum statistics

In this chapter we give a short introduction into quantum statistics. We will introduce the reader to the preliminaries and basic concepts which will be needed in the upcoming sections. The outline of this chapter is as follows: In Section 2.1 we will briefly introduce the basic notion of classical probability theory. After a short motivation in Section 2.2 we will define the concepts of quantum probability theory in terms of Hilbert spaces (Section 2.3) and measurements on them (Section 2.4). In Section 2.5 we will generalize the setup for tensor products of several Hilbert spaces. In Section 2.6 we will then show how this mathematical framework of quantum probability theory is used to model the physical theory of quantum mechanics. Finally, in the last section we will clarify those concepts by applying them to a simple example.

## 2.1 Classical probability theory

With classical probability theory we refer to the modern axiomatization of probability theory developed by Kolmogorov in the 1930s [2]. The central object in his formulation is the notion of a probability space. Before defining what we mean by a probability space we first have to introduce the elementary concepts of a $\sigma$-algebra and a measure space.

Let $\Omega$ be a set. Then we define

**Definition 2.1.1** (Algebra on $\Omega$)**.** *A collection $\mathscr{F}_0$ of subsets of $\Omega$ is called an algebra on $\Omega$ if*

*(i)* $\Omega \in \mathscr{F}_0$.
*(ii)* $A \in \mathscr{F}_0 \implies A^c := \Omega \backslash A \in \mathscr{F}_0$.
*(iii)* $A, B \in \mathscr{F}_0 \implies A \cup B \in \mathscr{F}_0$.

In particular, it is easy to see that this definition also implies that $\emptyset = \Omega^c \in \mathscr{F}_0$ and $A, B \in \mathscr{F}_0 \implies A \cap B \in \mathscr{F}_0$.

**Definition 2.1.2** ($\sigma$-algebra on $\Omega$)**.** *A collection $\mathscr{F}$ of subsets of $\Omega$ is called a $\sigma$-algebra on $\Omega$ if $\mathscr{F}$ is an algebra on $\Omega$ stable under countable set operations,*

$$\bigcup_n \mathscr{F}_n \in \mathscr{F}, \quad \forall \mathscr{F}_n \in \mathscr{F}, \quad n \in \mathbb{N}.$$

**Definition 2.1.3** (Measure space)**.** *Let $\Omega$ be a set equipped with a $\sigma$-algebra $\mathscr{F} = \sigma(\Omega)$, then we call the tuple $(\Omega, \mathscr{F})$ a measure space.*

Further, we define a $\sigma$-finite measure on $(\Omega, \mathscr{F})$ as follows:

**Definition 2.1.4** (($\sigma$-finite) measure)**.** *Let $(\Omega, \mathscr{F})$ be a measure space. A function $\mu : \mathscr{F} \to \bar{\mathbb{R}}$ is called measure on $\mathscr{F}$, if it has the following properties:*

*(i)* $\mu(\emptyset) = 0$.
*(ii)* $\mu(A) \geq 0$ *for all* $A \in \mathscr{F}$.
*(iii)* *For disjoint* $A_i \in \mathscr{F}$, $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$. *($\sigma$-additivity)*
*If there exist* $A_n \in \mathscr{F}$ *with* $\bigcup_{n=1}^{\infty} A_n = \Omega$ *and* $\mu(A_n) < \infty$ *for all n, then $\mu$ is called $\sigma$-finite.*

Now we can define a probability space as follows:

**Definition 2.1.5** (Probability space)**.** *Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a measure space, i.e. a basis set $\Omega$, a $\sigma$-algebra $\mathscr{F} = \sigma(\Omega)$ and a ($\sigma$-finite) measure $\mathbb{P}$, and let $\mathbb{P}(\Omega) = 1$, then $\mathbb{P}$ is called a probability measure and $(\Omega, \mathscr{F}, \mathbb{P})$ is called probability space.*

Since $\mathbb{P}(\Omega) = 1$, we have $0 \leq \mathbb{P}(A) \leq 1$ for all $A \in \mathscr{F}$ which agrees with the intuitive notion of a probability measure.

The space $(\Omega, \mathscr{F}, \mathbb{P})$ can be seen as a model for an experiment which one is able to repeat many times. The basis set $\Omega$ is in this context the space of all possible outcomes $\omega$ of the experiment. An outcome $\omega \in A \subseteq \Omega$ is measured by the experimentalist with probability $\mathbb{P}(A)$ according to the probability measure $\mathbb{P} : \mathscr{F} \to [0,1]$, where the collections of outcomes $A \in \mathscr{F}$ refers to a particular event.

In the following lemma we summarize some basic calculation rules for probability spaces. This rules will be needed to prove various Bell type inequalities.

**Lemma 2.1.6.** *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space and $A,B,A_n \in \mathscr{F}$, $n \in \mathbb{N}$. Then the following is true:*

*(i)* $\mathbb{P}(A^c) = 1-\mathbb{P}(A)$.
*(ii)* *For $A \subset B$, we have* $\mathbb{P}(A) \leq \mathbb{P}(B)$.
*(iii)* $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

**Proof.**

For $A \subset B$, $A \cup (B\backslash A) = B$ is a disjoint union and from the $\sigma$-additivity and non-negativity of $\mathbb{P}$ (Definition 2.1.4) it follows that $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B\backslash A) \geq \mathbb{P}(A)$ which proves (ii) and for $B = \Omega$ and using that $\mathbb{P}(\Omega) = 1$ it proves (i). Further we have $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \left(A_n \backslash \bigcup_{k=1}^{n-1} A_k\right)$, where on the right hand side the union is disjoint. It then follows using (ii) that $\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}\left(A_n \backslash \bigcup_{k=1}^{n-1} A_k\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ which proves (iii).

■

Let us clarify some of the terminology used in the above lemma: Here the complementary $A^c$ of the event $A$ means the non-realization of the event $A$. If $A \subset B$ we also say that the realization of the event $A$ implies the realization of the event $B$. Disjoint events $A$ and $B$ are also called non-compatible, i.e. $AB = 0$.

Important quantities in the theory of probability are random variables. Let us give the following definition:

**Definition 2.1.7** (Random variable, probability distribution)**.** *Let $(\Omega,\mathscr{F},\mathbb{P})$ be a probability space, then we call the $\mathscr{F}$-$\mathbb{R}$-measurable map*

$$X : \Omega \to \mathbb{R}$$

*a real random variable. Further, we call the measure image*

$$\mathbb{P}_X : \sigma(\mathbb{R}) \to [0,1] : x \mapsto \mathbb{P}_X(x) := \mathbb{P}(X^{-1}(x))$$

*a probability distribution of $X$.*

Using this definition we can also define the expectation value of a real random variable:

**Definition 2.1.8** (Expectation value)**.** *Let $X : \Omega \to \mathbb{R}$ be a quasi-integrable random variable, then*

$$\mathbb{E}(X) := \int X d\mathbb{P} = \int x d\mathbb{P}_X(x) \tag{2.1}$$

*is called the expectation value of $X$.*

Note that for the image of $X$ being a finite subset $T$ of $\mathbb{R}$ this definition agrees with the intuitive definition of the expectation value,

$$\mathbb{E}(X) = \sum_{x \in T} x\mathbb{P}(X = x), \tag{2.2}$$

where by $\mathbb{P}(X = x)$ we mean $\mathbb{P}(\{\omega \in \Omega | X(\omega) = x\})$. Random variables play an important role in the theory of probability and are essential for later developments.

## 2.2 From classical probability to quantum probability

In the previous section we gave a brief overview of the concepts of Kolmogorov's probability theory which we referred to as classical probability. However, one year before the publication of Kolmogrov's work, von Neumann formulated an axiomatization of quantum mechanics which is sometimes referred to as quantum probability [1]. In quantum probability the scheme of classical probability is loosen somewhat in the sense that one misses the probability space $\Omega$ with its elements, the outcomes $\omega$. However, as we will see in next sections, one can define a analogous quantity to the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ by a measurable space $(\mathscr{X}, \mathscr{A})$. In quantum probability $(\mathscr{X}, \mathscr{A})$ is organized around the principle mathematical object, a separable Hilbert space. Events for example are represented by projections onto closed subspaces of the Hilbert space. The details of von Neumann's axiomatization are described in the following sections for the special case of finite dimensional separable Hilbert spaces.

For further references on the topic of quantum probability the reader is referred to [6, 11, 12]. A nice exposition of how quantum probability forms an extension to classical probability theory and in what sense this relation is sometimes misunderstood see [13].

## 2.3 Hilbert spaces, subspaces, states and operators

*Gentlemen: There's lots of room left in Hilbert space.*

- S. MacLane

As already mentioned on the previous section, the basic object in von Neumann's formulation of quantum probability is a separable Hilbert space. Let us give the following definition:

**Definition 2.3.1** (Hilbert space)**.** *A Hilbert space is a complex linear space with a function*

$$\mathscr{H} \times \mathscr{H} \to \mathbb{C}: \quad (\psi, \chi) \mapsto \langle \psi, \chi \rangle$$

*called the inner product, with the following properties for all* $\psi, \chi, \phi \in \mathscr{H}$ *and* $\lambda \in \mathbb{C}$

*(i)* $\langle \psi, \psi \rangle \geq 0$ *and* $\langle \psi, \psi \rangle = 0$ *if and only if* $\psi = 0$
*(ii)* $\langle \psi, \chi + \phi \rangle = \langle \psi, \chi \rangle + \langle \psi, \phi \rangle$
*(iii)* $\langle \psi, \lambda \chi \rangle = \lambda \langle \psi, \chi \rangle$
*(iv)* $\langle \psi, \chi \rangle = \langle \chi, \psi \rangle^*$
*(v)* $\mathscr{H}$ *is complete under the norm* $\psi \mapsto \|\psi\| := \sqrt{\langle \psi, \psi \rangle}$

There are certain equivalence classes between Hilbert spaces which are isomorphic to each other:

**Definition 2.3.2** (Isomorphism)**.** *Two Hilbert spaces* $\mathscr{H}_1$ *and* $\mathscr{H}_2$ *are said to be isomorphic, if there exists a linear operator* $\mathscr{U} : \mathscr{H}_1 \to \mathscr{H}_2$ *such that* $\langle \mathscr{U}\psi, \mathscr{U}\chi \rangle_{\mathscr{H}_2} = \langle \psi, \chi \rangle_{\mathscr{H}_1}$*. Such an operator is called unitary.*

Let us now give some examples of Hilbert spaces:

**Example 2.3.1.** (i) Let $\mathbb{C}^d$ denote the set of $d$-dimensional complex vectors. The inner product of two vectors $\psi = (\psi_1, .., \psi_d)$ and $\chi = (\chi_1, ..., \chi_d)$ is defined as $\langle \psi, \chi \rangle = \sum_{i=1}^{d} \psi_1^* \chi_i$. (ii) Let $l_2$ denote the set of sequences of complex numbers $\psi = \{\psi_i\}_{i=1}^{\infty}$ which satisfy $\sum_{i=1}^{\infty} |\psi_i|^2 < \infty$ with the inner product $\langle \psi, \chi \rangle = \sum_{i=1}^{\infty} \psi_1^* \chi_i$.

In particular the usual definitions for vector spaces directly translate to the case of Hilbert spaces. For example we define:

**Definition 2.3.3.** *Two elements of* $\psi, \chi \in \mathscr{H}$ *are called orthogonal if* $\langle \psi, \chi \rangle = 0$*. Further, an element* $\psi \in \mathscr{H}$ *is called normalized if* $\|\psi\| = 1$*.*

Let us state two useful theorems without proof; the proofs can be found in every standard text book on functional analysis (see for instance [14]).

**Theorem 2.3.4.** *Every Hilbert space* $\mathscr{H}$ *has an orthonormal basis, a set of orthogonal and normalized elements of* $\mathscr{H}$ *that span* $\mathscr{H}$*.*

**Theorem 2.3.5.** *Let $S = \{\phi_i\}_{i\in A}$ be a orthonormal basis of a Hilbert space $\mathscr{H}$, then for each $\psi \in \mathscr{H}$,*

$$\begin{aligned} \psi &= \sum_{i\in A} \langle \phi_i, \psi\rangle \phi_i, \\ \|\psi\|^2 &= \sum_{i\in A} |\langle \phi_i, \psi\rangle|^2 . \end{aligned}$$

Having introduced the notion of Hilbert spaces we further have to specify what we mean by a separable Hilbert space. Using the common notion for metric spaces we define:

**Definition 2.3.6.** *A Hilbert space is called separable if it contains a dense countable subset.*

From this definition one can make the following useful statement:

**Theorem 2.3.7.** *A Hilbert space $\mathscr{H}$ is separable if and only if it has a countable orthonormal basis S. If the number of elements in S is $d < \infty$, then $\mathscr{H}$ is isomorphic to $\mathbb{C}^d$. If there are countable many elements in S, then $\mathscr{H}$ is isomorphic to $l_2$.*

**Proof.**

Consider first the case $d < \infty$. Using the Gram-Schmidt procedure we can find a complete orthonormal basis $\{\phi_i\}_{i=1}^d$ of the separable Hilbert space $\mathscr{H}$. We define the map $\mathscr{U} : \mathscr{H} \to \mathbb{C}^d : \psi \mapsto \{\langle \phi_i, \psi\rangle\}_{i=1}^d$. According to Theorem 2.3.5 this map is onto, further, it is unitary which proves that $\mathscr{H}$ is isomorphic to $\mathbb{C}^d$. The proof for $\mathscr{H}$ being isomorphic to $l_2$ is similar.

■

In this introduction we will only consider the case where the separable Hilbert space is finite dimensional. In this case, it follows from Theorem 2.3.7 that $\mathscr{H}$ is simply $\mathbb{C}^d$, where $d = \dim\mathscr{H}$. The elements $\psi \in \mathscr{H}$ are then $d$-dimensional column vectors of complex numbers and linear operators - linear maps from $\mathscr{H}$ to itself - are complex $d \times d$ matrices. Historically, the use of finite dimensional vector spaces in the formulation of quantum mechanics was first developed in the Heisenberg-Born-Jordan "Matrizenmechanik" (matrix mechanics) which was thought to be distinct to Schrödinger's "Wellenmechanik" (wavefunction mechanics). Later both formulations were unified in the framework of Hilbert spaces (for an exposition of this relation the reader is referred to [1]).

In the following we give some basic definitions of elements of $\mathscr{H}$, subspaces of $\mathscr{H}$ and self-adjoint operators.

**Definition 2.3.8** (Pure states)**.** *Consider an element $\psi \in \mathscr{H} = \mathbb{C}^d$, i.e. a d-dimensional complex valued column vector. The inner product is defined as $\langle \psi, \chi \rangle = \sum_{i=1}^{d} \psi_i^* \chi_i$, where $\psi_i$ and $\chi_i$ are the vector components of $\psi$ and $\chi$. Two elements are orthogonal, if their inner product is zero. The norm of a element reads $\|\psi\|^2 = \sum_{i=1}^{d} |\psi_i|^2$. If $\|\psi\| = 1$ it is said to be normalized. Normalized elements of a Hilbert space are called pure states. The span of a element $\psi \in \mathscr{H}$ is denoted by $[\psi] = \{\lambda\psi | \lambda \in \mathbb{C}\}$.*

As already mentioned earlier events will be described by closed subsystems $A \subseteq \mathscr{H}$, where $0 \leq \dim A \leq d$. Since $\mathscr{H}$ is finite dimensional, all subsystems are closed. Further, we say that two subsystems $A_1, A_2$ are orthogonal if and only if $\langle \psi, \chi \rangle = 0$ for all $\psi \in A_1$ and $\chi \in A_2$.

**Definition 2.3.9** (Decomposition)**.** *If subspaces $A_1, ..., A_r \subseteq \mathscr{H}$ are orthogonal and span $\mathscr{H}$ we say they form a decomposition of $\mathscr{H}$, written as*

$$\mathscr{H} = A_1 \oplus ... \oplus A_r.$$

*The decomposition is called proper if none of the subspaces is the zero subspace. Two subspaces $A, B$ are called compatible if there exist orthogonal subspaces $A_1, A_2, A_3$ such that*

$$\begin{aligned} A &= A_1 \oplus A_2, \\ B &= A_1 \oplus A_3. \end{aligned}$$

It is clear from Theorem 2.3.4 and 2.3.5 that for every Hilbert space $\mathscr{H}$ one can generate a proper decomposition by a partition of its orthonormal basis $S = \{\phi_i\}_{i \in A}$.

Essential quantities in von Neumann's formulation of quantum mechanics are self-adjoint operators. Let us in the following given some basic definitions and properties:

**Definition 2.3.10** (Operators)**.** *By an operator on $\mathscr{H}$ we mean a linear map $X : \mathscr{H} \to \mathscr{H} : \psi \mapsto X\psi$, represented by an $d \times d$ complex matrix. Let us define the adjoint of an operator, denoted by $X^\dagger$, by the unique solution of $\langle X^\dagger \psi, \chi \rangle = \langle \psi, X\chi \rangle$ for all $\psi, \chi \in \mathscr{H}$, that is the complex conjugate and transpose of the matrix $X$. An operator which satisfies $X = X^\dagger$ is called self-adjoint. Two operators $X, Y$ are said to be commuting if $XY = YX$. Further, we define the norm of an operator $X$ as*

$$\|X\| := \sup \{ \|X\psi\| \, | \, \psi \in \mathscr{H}, \|\psi\| = 1 \}.$$

As an important example of a self-adjoint operator we define the orthogonal projector:

**Definition 2.3.11** (Projectors)**.** *A projector is a operator* $\Pi$ *on* $\mathscr{H}$ *with the following property,* $\Pi\Pi = \Pi$*. The projector is called orthogonal if* $\Pi = \Pi^\dagger$*.*

Suppose we have a subspace $A$ spanned by $(\phi_i)_{i=1}^r$, $r \leq d$ then the orthogonal projector on this subspace is $\Pi_A = \sum_{i=1}^r \phi_i \phi_i^\dagger$ (here again, $\phi_i^\dagger$ means the complex conjugate and transpose vector of $\phi_i$). Hence, it acts on states within $A$ as the identity and on states of the orthogonal complement $A^\perp$ as the zero operator. In other words this means that $\Pi_A$ has two eigenvalues 1 and 0, where the corresponding eigenstates are $A$ and $A^\perp$. One sees that the orthogonal projector $\Pi_A$ defines a proper decomposition $\mathscr{H} = A \oplus A^\perp$. In the following we show that there is a general one-to-one correspondence between self-adjoint operators and proper decompositions of $\mathscr{H}$.

Let us first state some properties of the spectrum of self-adjoint operators. Let $x$ denote an eigenvalue of an operator $X$ and $[X = x]$ the corresponding eigenspace defined by $[X = x] = \{\phi \,|\, X\phi = x\phi\}$. If $x$ is not eigenvalue of $X$ then $[X = x]$ denotes the zero subspace. Further we define the spectrum of $X$, denoted by $\mathrm{spec}(X)$, as the set of all eigenvalues of $X$.

**Theorem 2.3.12.** *Let $X$ be a self-adjoint operator on $\mathscr{H}$. Then the following is true:*

*(i) All eigenvalues $x$ are real.*
*(ii) The corresponding eigenspaces $[X = x]$ are necessarily orthogonal.*

**Proof.**

From the eigenvalue equation we get $\langle \phi, X\phi \rangle = x\langle \phi, \phi \rangle$. Subtracting from this equation its adjoint and using that $X = X^\dagger$ yields $0 = (x - x^*)\langle \phi, \phi \rangle$ and hence $x = x^*$ which proves (i). (ii) is a consequence of the following powerful theorem.

■

**Theorem 2.3.13** (Spectral decomposition)**.** *Any self-adjoint operator $X$ on $\mathscr{H}$ is diagonal with respect to some orthonormal basis of $\mathscr{H}$.*

**Proof.**

Let us prove the statement by induction. For $d = 1$ it is trivially proven. Let $x$ denote an eigenvalue of $X$, $\Pi \equiv \Pi_{[X=x]}$ the projector onto the eigenspace $[X = x]$ and $\Pi^\perp = \mathbf{1} - \Pi$ the projector onto its orthogonal complement. We write $X = (\Pi + \Pi^\perp)X(\Pi + \Pi^\perp) = \Pi X \Pi + \Pi^\perp X \Pi^\perp + \Pi^\perp X \Pi + \Pi X \Pi^\perp$. Now $\Pi^\perp X \Pi = 0$, since $X$ takes $[X = x]$ into itself. Further, since $X$ is self-adjoint, also $\Pi X \Pi^\perp = 0$. Now we have $X = \Pi X \Pi + \Pi^\perp X \Pi^\perp$. By induction $\Pi^\perp X \Pi^\perp$ is already diagonal with respect to some orthonormal basis of the subspace $[X = x]^\perp$ and since $\Pi X \Pi = x\Pi$, the operator $\Pi X \Pi$ is also diagonal with respect to some orthonormal basis of $[X = x]$ which proves that $X$ is diagonal with respect to the union of those bases.

■

**Corollary 2.3.14.** *From the spectral decomposition it follows that we can write a self-adjoint operator $X$ on $\mathscr{H}$ as*

$$X = \sum_x x \Pi_{[X=x]}, \quad \mathbf{1} = \sum_x \Pi_{[X=x]},$$

*where the sum is taken over all $x \in \mathrm{spec}(X)$. Hence, there is an one-to-one correspondence between a self-adjoint operator $X$ and a proper decomposition $\mathscr{H} = \bigoplus_x [X = x]$.*

As we will discuss in more detail in the following sections what is actually measured in a quantum experiment are eigenvalues of self-adjoint operators. Because of this a self-adjoint operator is also called an observable. In that sense an self-adjoint operator is somewhat similar to a random variable in classical probability theory. The eigenvalues of a self-adjoint operator $X$ are the outcomes $x$ which define the sample space $\mathscr{X}$. In the next section we will show that $(\mathscr{X}, \mathscr{A})$ indeed defines a measurable space and we will see how to assign probabilities to those events. Before we can proceed with the definition of measurements and the assignment of probabilities, we first have to define some further quantities.

In analogy to Corollary 2.3.14 we can also define functions of observables. Suppose $f$ is a real function and $X$ a self-adjoint operator, then we define a new self-adjoint operator

$$f(X) = \sum_x f(x) \Pi_{[X=x]}. \tag{2.3}$$

Note that $X$ and $f(X)$ have the same eigenspaces, only the eigenvalues $x$ are replaced by $f(x)$.

In the formulation of product systems we often have to deal with several commuting observables. The following theorem is of essential importance:

**Theorem 2.3.15** (Simultaneous diagonalization)**.** *Let $X,Y$ denote two self-adjoint operators on $\mathscr{H}$. Then both operator commute, $XY = YX$, if and only if there exists an orthonormal basis of $\mathscr{H}$ such that both $X$ and $Y$ are diagonal in this basis. We say that commuting self-adjoint operators are simultaneously diagonalizable.*

**Proof.**

If $X$ and $Y$ are both diagonal in the same orthonormal basis it follows trivially that they both commute. Let us now show the converse: For all $\phi_x \in [X = x]$ we have

$$XY\phi_x = YX\phi_x = xY\phi_x$$

and hence $Y\phi_x \in [X = x]$. Let us now define the operator $Y_x := \Pi_{[X=x]}Y\Pi_{[X=x]}$. Clearly, $Y_x$ is self-adjoint on $[X = x]$, and therefore there exist, according to Theorem 2.3.13, an orthonormal basis of $[X = x]$ in which $Y_x$ is diagonal. Let us denote the subspace of $[X = x]$ for which $Y_x$ has eigenvalue $y$ by $[X = x, Y = y]$. Since for all $\phi_{x,y} \in [X = x, Y = y]$ we have $Y\phi_{x,y} \in [X = x]$, it follows that

$$Y\phi_{x,y} = \Pi_{[X=x]}Y\phi_{x,y} = \Pi_{[X=x]}Y\Pi_{[X=x]}\phi_{x,y} = y\phi_{x,y},$$

and hence $[X = x, Y = y]$ is also eigenspace of $Y$ with eigenvalue $y$ which completes the proof.

■

Using Theorem 2.3.15 we can also define functions of two commuting self-adjoint operators by

$$f(X,Y) = \sum_{x,y} f(x,y)\Pi_{[X=x,Y=y]}, \tag{2.4}$$

where again we sum over $x \in \operatorname{spec}(X)$ and $y \in \operatorname{spec}(Y)$.

**Theorem 2.3.16** (Trace)**.** *Let $S = \{\phi_i\}_{i=1}^d$ be an orthonormal basis of $\mathscr{H}$ and $X,Y$ positive operators on $\mathscr{H}$, i.e. operators which satisfy $\langle\psi, X\psi\rangle \geq 0$ for all $\psi \in \mathscr{H}$. We define the trace, $\operatorname{tr}X = \sum_{i=1}^d \langle\phi_i, X\phi_i\rangle$. The trace is independent of the choice of orthonormal basis of $\mathscr{H}$. Further it has the following properties:*

*(i)* $\operatorname{tr}(X+Y) = \operatorname{tr}X + \operatorname{tr}Y$.
*(ii)* $\operatorname{tr}(\lambda X) = \lambda \operatorname{tr}X$, *for* $\lambda \in \mathbb{C}$ *with* $\lambda \geq 0$.
*(iii)* $\operatorname{tr}(XY) = \operatorname{tr}(YX)$.

**Proof.**

Let us first show that the trace is independent of the choice of orthonormal basis. Let $S = \{\phi_i\}_{i=1}^d$ be an orthonormal basis of $\mathscr{H}$ and $S' = \{\psi_i\}_{i=1}^d$ be another orthonormal basis. Then

$$\begin{aligned} \mathrm{tr}X &= \sum_{i=1}^d \langle \phi_i, X\phi_i\rangle = \sum_{i=1}^d \left\| X^{1/2}\phi_i \right\|^2 \\ &= \sum_{i=1}^d \left( \sum_{j=1}^d \left| \langle \psi_j, X^{1/2}\phi_i\rangle \right| \right) = \sum_{j=1}^d \left( \sum_{i=1}^d \left| \langle X^{1/2}\psi_j, \phi_i\rangle \right| \right) \\ &= \sum_{j=1}^d \left\| X^{1/2}\psi_j \right\|^2 = \sum_{j=1}^d \langle \psi_j, X\psi_j\rangle, \end{aligned}$$

which proofs the independence of the choice of basis. Properties (i) and (ii) follow immediately from Definition 2.3.1 and (iii) is proven easily by a change of basis from $S = \{\phi_i\}_{i=1}^d$ to $S' = \{Y\phi_i\}_{i=1}^d$.

■

**Definition 2.3.17** (Density matrices, mixed states)**.** *A non-negative self-adjoint operator, i.e. a self-adjoint operator $\rho$ which satisfies $\langle \psi, \rho\psi\rangle \geq 0$, with $\mathrm{tr}\rho = 1$ is called a density matrix or mixed state.*

According to Theorem 2.3.13 we can chose an orthonormal basis $\{\psi_i\}_{i=1}^d$ of $\mathscr{H}$ in which

$$\rho = \sum_{i=1}^d p_i \Pi_{[\psi_i]}, \tag{2.5}$$

where all $p_i \geq 0$ and $\sum_{i=1}^d p_i = 1$. If $p_i = \delta_{i,j}$ the density matrix, $\rho = \Pi_{[\psi_j]} = \psi\psi^\dagger$, is called pure, which agrees with the above definition. Clearly, a mixture of density matrices according to some probability distribution $q_i$ is again a density matrix. Specifically, any density matrix can be written as a mixture of pure states, hence the name mixed state.

As we will see in the upcoming section, the last two concepts, the trace operation and density matrices, are essential to define a probability measure on $(\mathscr{X}, \mathscr{A})$. In fact, as we will see, there exits a theorem stating that any probability measure on $(\mathscr{X}, \mathscr{A})$ is defined via those quantities.

## 2.4 Measurements

After this short introduction into Hilbert spaces we want to introduce the concepts of measurements or in statistical language we want to define the measure space $(\mathscr{X}, \mathscr{A})$ and its probability measure Pr.

As mentioned earlier, measurements are mathematically described by a measurable space which in the context of quantum probability we denoted by $(\mathscr{X}, \mathscr{A})$. Here $\mathscr{X}$ is the set of possible outcomes $x$, equipped with a $\sigma$-algebra $\mathscr{A} = \sigma(\mathscr{X})$ of subsets $A \subseteq \mathscr{X}$.

We now define a measurement in the following way:

**Definition 2.4.1** (Measurement)**.** *A measurement $M$ on $\mathscr{H}$ is specified by the following operator-valued probability measure (oprm), that is a map $M : \mathscr{A} \to \mathscr{L}(\mathscr{H}) : A \mapsto M(A)$, where $\mathscr{L}(\mathscr{H})$ is the set of self-adjoint operators on $\mathscr{H}$ and $M$ has the following properties:*

*(i)* $M(\mathscr{X}) = \mathbf{1}$.
*(ii)* $\mathbf{0} \leq M(A) \leq \mathbf{1}$, *for all* $A \in \mathscr{A}$.
*(iii)* *For disjoint* $A_n \in \mathscr{A}$, $M\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} M(A_n)$.

Such measurements are also sometimes called generalized measurements, in contrast to a special case, called simple measurements, which we will define later. Another common name for generalized measurements are POVMs (positive operator valued measures). It is interesting to see that properties (i)-(iii) are basically equivalent to the properties of classical probability measures, as stated in Definition 2.1.4-2.1.5, apart from the fact that $M$ takes values in $\{M(A) \in \mathscr{L}(\mathscr{H}) | \mathbf{0} \leq M(A) \leq \mathbf{1}\}$ instead of in $[0,1]$. Another big difference is that $\mathscr{X}$ does not have to be the real numbers or a subset of it, with the Borel $\sigma$-algebra, but can be anything else. Basically, the events $A \in \mathscr{A}$ are subspaces in the Hilbert space $\mathscr{H}$.

Suppose we measure a system characterized by a density matrix $\rho$ using some measurement $M$. In terms of our statistical framework we would like to know what the probability distribution of a random outcome $x \in \mathscr{X}$ is. This is done by the following:

**Definition 2.4.2** (Born rule, trace rule)**.** *Let $\rho$ be a mixed state measured by $M$, then the probability distribution of an outcome $x \in A \subseteq \mathscr{X}$ is given by*

$$\Pr : \mathscr{A} \to [0,1] : A \mapsto \Pr(A) \equiv \Pr(x \in A) = \operatorname{tr}(\rho M(A)).$$

It is easy to see that Pr defines a probability measure on $(\mathscr{X},\mathscr{A})$. First, since $M(\mathscr{X})=\mathbf{1}$ and $\mathrm{tr}\rho=1$ it follows that $\Pr(\mathscr{X})=1$. Further, since $\mathrm{tr}\rho=1$ and $\mathbf{0}\leq M(A)\leq\mathbf{1}$, it is easy to see that $0\leq\Pr(A)\leq 1$ for all $A\in\mathscr{A}$. From Definition 2.4.1 (iii) and the linearity of the trace (Theorem 2.3.16) it follows that for disjoint $A_n\in\mathscr{A}$, $\Pr\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}\Pr(A_n)$.

Let us now define a special kind of measurement which in most application is sufficient:

**Definition 2.4.3** (Simple measurement)**.** *A simple measurement is a map* $\Pi:\mathscr{A}\to\mathscr{L}(\mathscr{H}):A\mapsto\Pi_A$*, where* $\Pi_A$ *is the orthogonal projector onto the subspace* $A\subseteq\mathscr{H}$*.*

It is easy to see that $\Pi$ indeed fulfills the above requirements of a measurement. Since $\dim\mathscr{H}=d$, the simple measurement can take at most $d$ different outcomes, each associated to an element $A_i$, $i\leq d$, of a proper decomposition of $\mathscr{H}$. According to Corollary 2.3.14 we can find a self-adjoint operator $X$ which has each subspace $A_i\in\mathscr{X}$ as an eigenspace and hence we can label the outcomes by the eigenvalues $x_i$, $i\leq d$. In this sense we can call $X=\sum_i x_i\Pi_{[X=x_i]}$ an observable, since as in the case of a random variable in classical probability theory $X$ defines a map $X:\mathscr{X}\to\mathbb{R}$ from its eigenspaces to the real eigenvalues.

One sees that the Born rule (Definition 2.4.2) for the case of a simple measurement defines a probability measure on the lattice of subspaces of a separable Hilbert space. Furthermore, it has been shown that Pr is defined in a unique way in the following sense:

**Theorem 2.4.4** (Gleason's theorem)**.** *Any probability measure* Pr *on the lattice of subspaces of a separable Hilbert space* $\mathscr{H}=\bigoplus_i A_i$ *has to be of the form* $\Pr(A_i)=\mathrm{tr}(\rho\Pi_{A_i})$ *for some density matrix* $\rho$ *on* $\mathscr{H}$*.*

Gleason's theorem was already conjectured by von Neumann. The first proofs (see [15]) of this theorem were not entirely rigorous and the original proof due to Gleason [16] was very long. For an elementary proof the reader is referred to [17].

In the previous sections we gave an introduction to quantum probability. We defined the measurable space $(\mathscr{X},\mathscr{A})$ and its measure Pr which were centered around the principle object of a separable Hilbert space. In finite dimensions this Hilbert space was simply $\mathscr{H}=\mathbb{C}^d$. In some cases it is however useful to to view the Hilbert space as a composition of several smaller dimensional Hilbert spaces. In the following section we want to make the latter more precise.

## 2.5 Tensor products and entanglement

An important aspect in the theory of quantum probability is the case where the Hilbert space is a tensor product of several smaller Hilbert spaces. Let us define the tensor product as follows:

**Definition 2.5.1** (Tensor product)**.** *Let $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ denote Hilbert spaces. For $\psi_{\mathscr{A}}, \chi_{\mathscr{A}} \in \mathscr{H}_{\mathscr{A}}$, $\psi_{\mathscr{B}}, \chi_{\mathscr{B}} \in \mathscr{H}_{\mathscr{B}}$ and $\lambda \in \mathbb{C}$, we define the conjugate bilinear form $\psi_{\mathscr{A}} \otimes \psi_{\mathscr{B}}$ with the following properties:*

*(i)* $\lambda(\psi_{\mathscr{A}} \otimes \psi_{\mathscr{B}}) = (\lambda \psi_{\mathscr{A}}) \otimes \psi_{\mathscr{B}} = \psi_{\mathscr{A}} \otimes (\lambda \psi_{\mathscr{B}})$.
*(ii)* $(\psi_{\mathscr{A}} + \chi_{\mathscr{A}}) \otimes \psi_{\mathscr{B}} = \psi_{\mathscr{A}} \otimes \psi_{\mathscr{B}} + \chi_{\mathscr{A}} \otimes \psi_{\mathscr{B}}$.
*(iii)* $\psi_{\mathscr{A}} \otimes (\psi_{\mathscr{B}} + \chi_{\mathscr{B}}) = \psi_{\mathscr{A}} \otimes \psi_{\mathscr{B}} + \psi_{\mathscr{A}} \otimes \chi_{\mathscr{B}}$.

*On the set of finite linear combinations of such conjugate bilinear forms we define the following inner product*

$$\langle \psi_{\mathscr{A}} \otimes \psi_{\mathscr{B}}, \chi_{\mathscr{A}} \otimes \chi_{\mathscr{B}} \rangle = \langle \psi_{\mathscr{A}}, \chi_{\mathscr{A}} \rangle_{\mathscr{A}} \langle \psi_{\mathscr{B}}, \chi_{\mathscr{B}} \rangle_{\mathscr{B}},$$

*where $\langle .,. \rangle_{\mathscr{A}}$ and $\langle .,. \rangle_{\mathscr{B}}$ are the inner products on $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively. The completion of the set of finite linear combinations under this inner product is called the tensor product $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$.*

There is an important theorem telling us how to construct a basis for $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$:

**Theorem 2.5.2.** *Let $\{\phi_i\}$ and $\{\psi_j\}$ be orthonormal basis for $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively, then $\{\phi_i \otimes \psi_j\}$ is a orthonormal basis for $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$.*

The proof is very easy, but not very instructive and is therefore omitted. In particular this theorem tells us that if $\{\phi_i\}$ and $\{\psi_j\}$ are orthonormal basis for $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively, then every element $\psi \in \mathscr{H} = \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ can be written in the following form

$$\psi = \sum_{i,j} c_{ij} \phi_i \otimes \psi_j. \tag{2.6}$$

This is the most general form for a element in $\mathscr{H}$. As a special case we define:

**Definition 2.5.3** (Entanglement, pure states)**.** *A element $\psi \in \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ is called separable if it can be written as a product*

$$\psi = \psi_{\mathscr{A}} \otimes \psi_{\mathscr{B}}, \tag{2.7}$$

*where $\psi_{\mathscr{A}} \in \mathscr{H}_{\mathscr{A}}$ and $\psi_{\mathscr{B}} \in \mathscr{H}_{\mathscr{B}}$. A pure state which is not separable, i.e. which is of the form of Equation (2.6) but cannot be written as Equation (2.7), is called entangled.*

The name entanglement goes back to the German word "Verschränkung" which was introduces by Schrödinger [18]. Entangled states play an important role in the theory of quantum statistics and they appear again at various places within this thesis. A reasonable question one might ask at this point is: Is there an easy way to see whether a pure state is entangled or not or is there a measure of entanglement? We will come back to this question towards the end of this section.

We have seen that an arbitrary element $\psi \in \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ is of the form of Equation (2.6). However there is a powerful theorem telling us that there is a particular basis in which this form simplifies significantly:

**Theorem 2.5.4** (Schmidt decomposition)**.** *Let $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ be Hilbert spaces of dimension $d$ and $d'$ respectively, with $d \leq d'$. For any pure state $\psi \in \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$, there exits a basis $\{e_i\}_{i=1}^{d}$ for $\mathscr{H}_{\mathscr{A}}$ and $\{f_j\}_{j=1}^{d'}$ for $\mathscr{H}_{\mathscr{B}}$ such that*

$$\psi = \sum_{i=0}^{d} \lambda_i \, e_i \otimes f_i, \tag{2.8}$$

*where the $\lambda_i$ are non-negative and are called the Schmidt coefficients. Since $\psi$ is normalized we also have $\sum_i \lambda_i^2 = 1$.*

**Proof.**

According to Equation (2.6) the general form of $\psi$ is

$$\psi = \sum_{i=0}^{d} \sum_{j=0}^{d'} c_{ij} \, \phi_i \otimes \psi_j,$$

where $\{\phi_i\}$ and $\{\psi_j\}$ are orthonormal basis for $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively and $C = \{c_{ij}\}$ is a real $d \times d'$ matrix. There is a theorem in linear algebra, called the singular value decomposition, stating that any $d \times d'$ matrix $C$ can be written as

$$C = U^T (D, \mathbf{0}) V,$$

where $U$ and $V$ are a unitary $d \times d$ and $d' \times d'$ matrix and $D$ is a positive semi-definite diagonal matrix. Hence we can write

$$\psi = \sum_{i,j,k=0}^{d} u_{ik} d_{kk} v_{kj} \, \phi_i \otimes \psi_j,$$

Due to the unitarity of $U$ and $V$, the defined $e_k = \sum_i u_{ik}\phi_i$ and $f_k = \sum_j v_{kj}\psi_j$ are again orthonormal sets.

■

Let us now consider the case of measurements and operators on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$. For any measurement $M_{\mathscr{A}}$ and $M_{\mathscr{B}}$ defined on $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively there is a nature extension to $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ via $M_{\mathscr{A}} \otimes \mathbf{1}$ and $\mathbf{1} \otimes M_{\mathscr{B}}$. In particular the identity operator on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ is $\mathbf{1} \otimes \mathbf{1}$. Similar extensions can also be made for simple measurements or observables defined on the single Hilbert space. These kind of product measurements and observables will be the only ones we are interested in even though there are not the most general ones, since we can also have non-separable operators. One example for which we will often encounter such non-seperability will be in the case of density operators. Clearly, since density operators or mixed states are mixtures of pure states the notion of entanglement as defined in Definition 2.5.3 for pure states directly translates to mixed states:

**Definition 2.5.5** (Entanglement, mixed states)**.** *A mixed state or density operator $\rho$ on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ is called separable if it can be written as*

$$\rho = \sum_i p_i \rho_{\mathscr{A},i} \otimes \rho_{\mathscr{B},i}, \tag{2.9}$$

*where $\rho_{\mathscr{A},i}$ and $\rho_{\mathscr{B},i}$ are density operators on $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively. A non-seperable mixed state is called entangled.*

Let us consider for a moment a separable state $\rho_{\mathscr{A}} \otimes \rho_{\mathscr{B}}$ on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$. If we apply a measurement $M_{\mathscr{A}} \otimes \mathbf{1}$ on this product state, the probability of an event $A$ can be completely determined using only the state $\rho_{\mathscr{A}}$, i.e. $\Pr(A) = \mathrm{tr}\,(\rho_{\mathscr{A}} M_{\mathscr{A}}(A))$. What happens if we consider an entangled state $\rho$ on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$? One would expect that some part of $\rho$ is not effected by the measurement $M_{\mathscr{A}} \otimes \mathbf{1}$, exactly as $\rho_{\mathscr{B}}$ does in the previous example. Hence, we expect that there is some kind of marginal state on which the measurement $M_{\mathscr{A}}$ produces the same result as $M_{\mathscr{A}} \otimes \mathbf{1}$ on the whole state $\rho$. Let us define this marginal state as follows:

**Definition 2.5.6** (Marginal state, partial trace)**.** *Let $\rho$ be mixed state on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$, where $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ are Hlbert spaces with orthonormal bases $\{\phi_i\}$ and $\{\psi_j\}$ respectively. Then we define the marginal state on $\mathscr{H}_{\mathscr{A}}$ as*

$$\rho_{\mathscr{A}} = \mathrm{tr}_{\mathscr{B}}(\rho),$$

*where*

$$\mathrm{tr}_{\mathscr{B}}(\rho) = \sum_j \langle \mathbf{1} \otimes \psi_j, \rho\, \mathbf{1} \otimes \psi_j \rangle$$

*is called the partial trace and respectively for $\mathscr{H}_{\mathscr{B}}$.*

One can easily check that for the marginal state $\rho_{\mathscr{A}}$ one indeed gets indeed

$$\Pr(A) = \mathrm{tr}\left(\rho_{\mathscr{A}} M_{\mathscr{A}}(A)\right) = \mathrm{tr}\left(\rho\, M_{\mathscr{A}}(A) \otimes \mathbf{1}\right). \tag{2.10}$$

Further, it can be shown that the marginal state as defined in Definition 2.5.6 is the unique function of $\rho$ which fulfills this property.

The marginal state has several applications, where one of the most useful is that they can be used to define a measure of entanglement [19]:

**Definition 2.5.7** (Entropy of entanglement)**.** *Let $\rho$ be a mixed state on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ and let $\rho_{\mathscr{A}}$ and $\rho_{\mathscr{B}}$ be the marginal stated on $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively, where $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ have the same dimension. Then we define the entropy of entanglement of $\rho$ as*

$$E(\rho) = S(\rho_{\mathscr{A}}) = S(\rho_{\mathscr{B}}),$$

*where S is the usual von Neumann entropy,*

$$S(\rho_{\mathscr{A}}) = -\mathrm{tr}\left(\rho_{\mathscr{A}} \log \rho_{\mathscr{A}}\right).$$

*Here,* $\log(\cdot)$ *refers to the logarithm to the basis 2.*

From the properties of the von Neumann entropy (see Appendix A) it follows that $0 \leq E(\rho)/\log(d) \leq 1$, where $d$ is the dimension of $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$. One can easily see that for a separable state $E(\rho) = 0$. From this we make the following plausible definition:

**Definition 2.5.8** (Maximally entangled state)**.** *A state $\rho$ for which $E(\rho)/\log(d) = 1$ is called maximally entangled.*

One can easily see that the maximally entangled state on $\mathscr{H} = \mathbb{C}^d \otimes \mathbb{C}^d$ can be written as $\rho = \mathbf{1}/d$ and in particular for a pure state $\rho = \psi\psi^\dagger$ one gets using the Schmidt decomposition (Theorem 2.5.4) the maximally entangled state

$$\psi = \frac{1}{\sqrt{d}} \sum_{i=0}^{d} e_i \otimes f_i. \tag{2.11}$$

For the case of a pure state $\rho = \psi\psi^\dagger$ also the entanglement entropy can be simplified. First, note that using the Schmidt decomposition one obtains

$$\rho_{\mathscr{A}} = \mathrm{tr}_{\mathscr{B}}(\psi\psi^\dagger) = \sum_i \lambda_i^2 e_i e_i^\dagger, \tag{2.12}$$

and similarly for $\rho_{\mathscr{B}}$. The entanglement entropy of a pure state reads

$$E(\psi) = -\sum_i \lambda_i^2 \log \lambda_i^2, \tag{2.13}$$

where the $\lambda_i$ are again the Schmidt coefficients.

The notion of entanglement entropy will be used at several points in the later chapters and is the only measure of entanglement which will be used within this thesis. For other sufficient and necessary conditions for entanglement the reader is referred to [20].

In this section we introduced the tensor product of Hilbert spaces and we discussed various concepts like states, operators and measurements on the tensor products of two Hilbert spaces. However, the above notions can easily be generalized to the case of tensor products of more than two Hilbert spaces.

## 2.6 The postulates of quantum mechanics

> *... quantum phenomena do not occur in a Hilbert space, they occur in a laboratory.*
>
> - A. Peres

In the previous sections we have formulated quantum probability theory as a purely mathematical framework. Clearly, when von Neumann developed his axiomatic formulation of quantum probability theory he wanted to formulate a theory of quantum mechanics which should model the results of actual physical experiments. In this section we will state the underlying postulates of quantum mechanics relating the mathematical concepts with physical objects in quantum experiments.

**Postulate 1.** Associated to any closed physical system is a separable Hilbert space $\mathscr{H}$, also known as the state space of the system. The physical properties of the system are completely described by a density matrix $\rho$ on $\mathscr{H}$. A composite system is modeled by a tensor product of Hilbert spaces associated to each subsystem with a joint state $\rho$ on the tensor product of Hilbert spaces describing the properties of the composite system.

Having given the kinematical setting of a quantum system in terms of its state, it is important to now how the system evolves in time, i.e. how is the dynamics of the system.

**Postulate 2.** The evolution of a closed quantum system is described by a unitary transformation. That is, that state $\rho(t_1)$ at time $t_1$ is related to the state $\rho(t_2)$ at time $t_2$ by

$$\rho(t_2) = U(t_1,t_2)\rho(t_1)U^\dagger(t_1,t_2),$$

where $U$ is a unitary operator on $\mathscr{H}$, i.e. an operator whose inverse is equal to its adjoint, $U^\dagger U = \mathbf{1}$.

In the case of a pure state $\rho(t) = \psi(t)\psi^\dagger(t)$ we can write

$$\psi(t_2) = U(t_1,t_2)\psi(t_1). \tag{2.14}$$

Note that any unitary operator can be written as $U = \exp(iA)$, where $A$ is a self-adjoint operator. The precise form is given by

$$U(t_1,t_2) = \exp\left(-\frac{\mathbf{i}}{\hbar}H(t_2-t_1)\right), \tag{2.15}$$

where $\hbar$ is a physical constant called the Planck constant and $\mathbf{i} = \sqrt{-1}$ is the imaginary number, the self-adjoint operator $H$ is called the Hamitonian of the closed system. In particular we can view Equation (2.14) as the solution of the differential equation

$$i\hbar\frac{d}{dt}\psi(t) = H\psi(t) \tag{2.16}$$

which is known as the Schrödinger equation. The Schrödinger equation tells us how a closed system evolves under time when isolated from any environment. The real nonnegative eigenvalues of the Hamiltonian describe the energy of the closed system.

**Postulate 3.** A measurement on a closed system is described by a generalized measurement $M(A)$, where the probability of obtaining an outcome $x \in A \subseteq \mathscr{X}$ is given by the Born rule

$$\Pr(x \in A) = \operatorname{tr}(\rho M(A)).$$

This completes the set of postulates of quantum mechanics. We see that we have developed a sophisticated framework of quantum statistics, which we can use to model quantum mechanical experiments. As we can use classical statistics

á la Kolmogorov to model outcomes of "classical experiments", say the tossing of coins, we can use the machinery of quantum statistics to model quantum mechanical experiments, like for example the measurement of the spin of an electron.

In the following we want to illustrate the application of quantum statistics in a simple example, so that the reader gets familiar with the concepts introduced in this chapter.

## 2.7 An example: The qubit

In this section we want to make the above rather abstract concepts more concrete in a specific simple example. Before going into the details of the example let us first introduce some useful notation:

**Dirac notation.** We indicate an element $\psi$ of a Hilbert space by $|\psi\rangle$, sometimes called a ket. The adjoint $\psi^\dagger$ is denoted by a so-called bra $\langle\psi|$. The inner product of two elements $\psi, \chi \in \mathscr{H}$ is then simply written $\langle\psi|\chi\rangle$. For $\langle\psi, X\chi\rangle = \langle\psi X^\dagger, \chi\rangle$, where $X$ is an operator, we use the notation $\langle\psi|X|\chi\rangle$.

This notation is sometimes also called the "braket" notation. It can be in particular very useful to abbreviate certain notations with several labels and indices. For example for a orthonormal basis we simple write $\{|i\rangle\}$. For a basis vector of $\mathscr{H}_\mathscr{A} \otimes \mathscr{H}_\mathscr{B}$ we then write $|i\rangle_\mathscr{A} \otimes |j\rangle_\mathscr{B}$ or if the context is clear even $|ij\rangle$.

The most simple case to look at is that of $\mathscr{H} = \mathbb{C}^2$, where the dimension of the Hilbert space is $d = 2$. This setting is used in physics to model a spin-half particle like an electron (see standard quantum mechanics textbooks, e.g. [21]). In particular the number of possible outcomes is (maximally) two, referring to measuring the particle being spin down or spin up. In quantum information this case is referred to as the *qubit* (e.g. see [6]). As a *classical bit* which can have values 0 and 1 a *qubit* can be in states "spin down" or "spin up", sometimes also labeled as $|0\rangle$ and $|1\rangle$.

In the following we will see that a state in $\mathscr{H} = \mathbb{C}^2$ can be represented by a vector $\vec{a} \in \mathbb{R}^3$ with $|\vec{a}| \leq 1$ and a measurement corresponds to a unit vector $\vec{u} \in \mathbb{R}^3$.

Let us first make the following definition:

**Definition 2.7.1** (Pauli spin matrices)**.** *The following matrices are called Pauli spin matrices*

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -\mathbf{i} \\ \mathbf{i} & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

It is easy to check the following properties of the Pauli spin matrices:

**Theorem 2.7.2** (Properties of the Pauli spin matrices)**.** *The Pauli spin matrices satisfy:*

$$\begin{aligned}\sigma_x\sigma_y = -\sigma_y\sigma_x &= \mathbf{i}\sigma_z,\\ \sigma_y\sigma_z = -\sigma_z\sigma_y &= \mathbf{i}\sigma_x,\\ \sigma_z\sigma_x = -\sigma_x\sigma_z &= \mathbf{i}\sigma_y,\\ \sigma_x^2 = \sigma_y^2 = \sigma_z^2 &= \mathbf{1}.\end{aligned}$$

*Further, we have* $\sigma_i\dagger = \sigma_i$*,* $\mathrm{tr}\sigma_i = 0$ *and* $\det\sigma_i = -1$*,* $\forall i \in \{x,y,z\}$*. In particular this means that all three Pauli spin matrices have eigenvalues* $\pm 1$*.*

Now it is easy to see that any self-adjoint operator $X$ in $\mathscr{H} = \mathbb{C}^2$ can be written as

$$X = w\mathbf{1} + \vec{x}\cdot\vec{\sigma}, \tag{2.17}$$

where we defined $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ and $\vec{x} \in \mathbb{R}^3$. In particular, any density matrix $\rho$ can be written in the form Equation (2.17), where the condition $\mathrm{tr}\rho = 1$ requires that $w = 1/2$ and the non-negativity requires that $|\vec{x}| \leq 1/2$. Hence we can write,

$$\rho(\vec{a}) = \frac{1}{2}\left(\mathbf{1} + \vec{a}\cdot\vec{\sigma}\right), \tag{2.18}$$

with $a \in \mathbb{R}^3$ and $|\vec{a}| \leq 1$. For the case that $|\vec{a}| = 1$, $\rho$ has eigenvalues 0 and 1 and is thus a projection operator which refers to a pure state, namely the eigenvector corresponding to the eigenvalue 1. A state with $|\vec{a}| \leq 1$ can be written as

$$\rho(\vec{a}) = \alpha\rho(\vec{u}) + \beta\rho(-\vec{u}), \tag{2.19}$$

where we defined $\alpha = |\vec{a}|$, $\beta = 1-\alpha$ and $\vec{u} = \vec{a}/|\vec{a}|$. Hence, a state with $|\vec{a}| < 1$ refers to a mixed state, namely a mixture of two pure states, $\rho(\vec{u})$ and $\rho(-\vec{u})$.

Now it is easy to see that a simple measurement $M$ is determined by two projectors, generally written as $\Pi_{\vec{v}} \equiv \rho(\vec{v})$ and $\Pi_{-\vec{v}} \equiv \rho(-\vec{v})$, where $\vec{v}$ is a unit vector in $\mathbb{R}^3$. If we name the two possible outcomes $-1$ and $+1$, then $M$ can take values in $\mathscr{X} = \{-1,+1\}$.

Thus for a simple measurement $M(\pm 1) = \Pi_{\pm\vec{v}}$ of a system being in a mixed state $\rho(\vec{a})$ the probability of a outcomes $x \in \mathscr{X}$ is given by

$$\Pr(x = \pm 1) = \mathrm{tr}\left(\rho(\vec{a})\Pi_{\pm\vec{v}}\right) = \frac{1}{2}\left(1 + \vec{a}\cdot\vec{v}\right). \tag{2.20}$$

In particular, if we choose a measurements which is orthogonal to the direction of the state, i.e. $\vec{a}\cdot\vec{v} = 0$, then the outcome is completely random, $\Pr(x = \pm 1) = 1/2$.

# Chapter 3

# Hidden variables and nonlocality proofs

*... what is proved by impossibility proofs is lack of imagination.*

- J. Bell

After having given a short introduction into the subject of quantum statistics in the previous chapter we now want to discuss the aspect of non-classicality in quantum statistics. In the introduction we have already seen one example of such non-local joint probabilities in the setting of a specific Bell inequality. In this chapter we want to examine this structure in more detail and under more general conditions. First, in Section 3.1 we introduce the general framework of Bell scenarios and how to describe joint probability distributions, so-called behaviors, on those. Further, we will derive several geometrical properties of the set of all such behaviors. In Section 3.2 we will characterize under which conditions such behaviors are said to be local or locally deterministic. We will see in Section 3.3 that there are Bell type inequalities which specify whether a behavior is local or not. The CHSH and CGLMP inequality are derived as explicit examples. In Section 3.4 we specify the set of joint probabilities which arises from the theory of quantum probability. We will see that in general such quantum behaviors are non-local and can be used to define a nonlocality proof. In Section 3.5 a inequality for quantum behaviors of the CHSH setting is given and interesting relations to Grothendieck's inequality and Grothendieck's constant are presented.

For further discussions and references on the topics of this chapter the reader is referred to [22, 23, 24].

## 3.1 Bell scenarios, behaviors and some discrete geometry

In the introduction we have already seen an explicit example of a Bell experiment. In the following we want to put this kind of experiments in a general framework which will be the starting point for our further investigations.

A general Bell scenario is first characterized by a number of observers involved in the experiment. For the purpose of this thesis we will restrict ourself to the case of two observers called Alice and Bob or $\mathscr{A}$ and $\mathscr{B}$ for short.[1] Secondly, each observer has the choice to perform a certain measurement. Let us label the set of Alice possible measurements by $X \in \{0,...,m_{\mathscr{A}}-1\}$ and respectively for Bob, $Y \in \{0,...,m_{\mathscr{B}}-1\}$. The possible outcomes of each experiment are labeled by $a \in A_X = \{0,...,n_X-1\}$ and $b \in B_Y = \{0,...,n_Y-1\}$, here the $a$ and $b$ refer to outcomes of Alice's and Bob's experiments respectively. Essential for the analysis of the experiment is the joint probability $\mathrm{p}_{ab|XY}$, that is the probability of $\mathscr{A}$ obtaining the outcome $a$ after measuring $X$ and at the same time $\mathscr{B}$ obtaining $b$ after measuring $Y$.

Let us now state some properties of joint probabilities associated to Bell scenarios. Clearly, since $\mathrm{p}_{ab|XY} \in \mathbb{R}$ is a probability it is positive and normalized:

$$\mathrm{p}_{ab|XY} \geq 0, \quad \forall\, a,b,X,Y \tag{3.1}$$

$$\sum_{a,b} \mathrm{p}_{ab|XY} = 1, \quad \forall\, X,Y. \tag{3.2}$$

Further, in a Bell experiment we require the measurements $X$ and $Y$ of $\mathscr{A}$ and $\mathscr{B}$ to be made instantaneous meaning that $\mathscr{A}$, when making her choice of measurement has no knowledge of which measurement $\mathscr{B}$ has chosen and vice versa. Mathematically, this means that the marginal probabilities $\mathrm{p}_{a|X}$ and $\mathrm{p}_{b|Y}$ are independent of $Y$ and $X$ respectively, that is

$$\mathrm{p}_{a|X} = \sum_b \mathrm{p}_{ab|XY} = \sum_b \mathrm{p}_{ab|XY'} \quad \forall a,X,Y,Y', \tag{3.3}$$

$$\mathrm{p}_{b|Y} = \sum_a \mathrm{p}_{ab|XY} = \sum_a \mathrm{p}_{ab|X'Y} \quad \forall a,X,X',Y. \tag{3.4}$$

This condition is sometimes also called no-signaling condition, due to the fact that $\mathscr{A}$ and $\mathscr{B}$ are not able to signal their choice of measurement to the other party before those do their measurement.

[1] The framework presented in the following can be easily extended to a general number of observers. For an explicit discussion see [24].

Using the previous requirements on the joint probability let us give the following definition:

**Definition 3.1.1** (Behavior)**.** *A joint probability distribution* $\mathrm{p}_{ab|XY}$ *satisfying the conditions (3.1)-(3.4) is called a behavior over the given behavior scheme* $(A_1,...,A_{m_{\mathscr{A}}};B_1,...,B_{m_{\mathscr{B}}})$*. A behavior is called deterministic if all* $\mathrm{p}_{ab|XY} \in [0,1]$.

We see that a Bell scenario defines a certain behavior scheme, i.e. the number of observers, their sets of measurements and the possible outcomes. Any resulting joint probability distribution is then given by a behavior on this behavior scheme. Let us in the following denote the set of all behaviors on a given scheme $(A_1,...,A_{m_{\mathscr{A}}};B_1,...,B_{m_{\mathscr{B}}})$ by $\mathscr{P}$.[2] Later in the text we will also define the set of all local behaviors $\mathscr{L}$ and the set of all quantum behaviors $\mathscr{Q}$. Studying certain relations of these sets is a major aspect of this thesis. In particular we will see that in general $\mathscr{L} \subset \mathscr{Q} \subset \mathscr{P}$.

For the further discussion it is useful to look at $\mathscr{P}$ from a geometrical point of view. Let us therefore identify a behavior $\mathrm{p}_{ab|XY}$ with a point $p = (...,\mathrm{p}_{ab|XY},...)$ in $\mathbb{R}^t$ with dimension $t = \sum_{X=0}^{m_{\mathscr{A}}-1}\sum_{Y=0}^{m_{\mathscr{B}}-1} n_X n_Y$.

Clearly the elements of $\mathscr{P}$ are subject to the constraints (3.1)-(3.4) and hence only define a subspace $\mathscr{P} \subset \mathbb{R}^t$. To classify the geometric properties of $\mathscr{P}$ we have to introduce some concepts of the theory of discrete geometry.

**Definition 3.1.2** (Linear combinations)**.** *A point* $x \in \mathbb{R}^n$ *is called a linear combination of* $x_1,...,x_k \in \mathbb{R}^n$ *if there exists a* $\lambda = (\lambda_1,...,\lambda_k)^T \in \mathbb{R}^k$ *such that* $x = \sum_{i=1}^k \lambda_i x_1$*. For* $\lambda \geq 0$ *or* $\sum_i \lambda_i = 1$ *or* $\lambda \geq 0$ *and* $\sum_i \lambda_i = 1$ *we call the linear combination conic or affine or convex.*

**Definition 3.1.3** (Hull)**.** *For a given subset* $S \subseteq \mathbb{R}^n$ *we call the set of all point which can be written as a conic, affine, convex combination of finitely many point of S the conic, affine, convex hull of S, denoted by* $\mathrm{cone}(S)$, $\mathrm{aff}(S)$, $\mathrm{conv}(S)$.

Using this definitions we can make a further definition of some intuitive geometrical objects:

**Definition 3.1.4** (Half space, polyhedron, polytope)**.** *Let* $b \in \mathbb{R}$, $a \in \mathbb{R}^n \setminus \{0\}$, $A \in \mathbb{R}^{n\times m}$ *and* $c \in \mathbb{R}^m$*: We call* $H = \{x \in \mathbb{R}^n | \sum_i a_i x_i \geq b\} \subseteq \mathbb{R}^n$ *a half space. An intersection* $P = \bigcap_{j=1}^m \left\{x \in \mathbb{R}^n | \sum_i A_{ij} x_i \geq c_j\right\} \subseteq \mathbb{R}^n$ *is called a polyhedron. A polyhedron P is called a polytope if there exist a* $B \in \mathbb{R}$, $B > 0$ *such that* $P \subseteq \{x \in \mathbb{R}^n | \, \|x\| < B\}$.

[2] For simplicity we will not write $\mathscr{P}\left(A_1,...,A_{m_{\mathscr{A}}};B_1,...,B_{m_{\mathscr{B}}}\right)$, since usually the behavior scheme will be clear from the context.

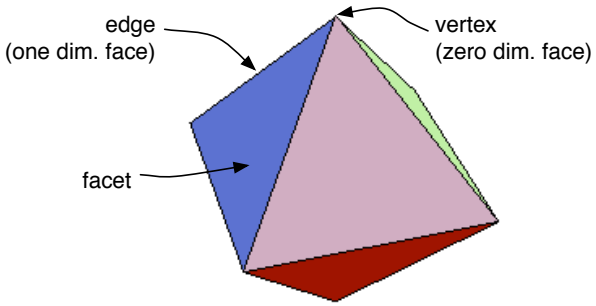


Figure 3.1: A three-dimensional illustration of a polytope, explaining also the concepts of vertices, edges and facets of the polytope.

Clearly, half spaces, the empty set and the whole space $\mathbb{R}^n$ are polyhedrons. From the Definition 3.1.4 it is also easy to see that every polyhedron is a finite intersection of half spaces. Polyhedrons are also of special interest in the theory of linear optimization (see for instance []). In particular we will be interested in bounded polyhedrons, i.e. polytopes. Figure 3.1 shows a typical polytope. From the figure one can already identify some intuitive subspaces, like edges, vertices etc. Let us strengthen this intuitive notion by some mathematical definition:

**Definition 3.1.5** (Faces, facets and vertices)**.** *Let $F \subseteq P$ be a subset of a polyhedron $P \subseteq \mathbb{R}^n$. We call $F$ a face of $P$ if there exists a inequality $\sum_i a_i x_i \geq b$ with $x, a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $F = P \cap \{x \in \mathbb{R}^n | \sum_i a_i x_i = b\}$. The face is called non-trivial if $\emptyset \neq F \neq P$. A non-trivial face of $P$ is called a facet if it is not included in one of the other faces of $P$. A zero dimensional face $F = \{x\}$ is called a vertex of $P$.*

One dimensional faces are then what is intuitively called a edge. Before coming back to our original motivation, a geometric understanding of the set $\mathscr{P}$, let us first state two important theorems which will be useful in the following sections.

**Theorem 3.1.6** (Minkowski)**.** *Every polytope can be written as a convex hull of a finite set.*

**Theorem 3.1.7** (Krein-Milman)**.** *Let $P \subseteq \mathbb{R}^n$ be a polyhedron and let $V \subseteq P$ denote the set of all vertices of $P$. Then the following is true, $P$ is a polytope if and only if it can be written as the convex hull of its vertices, i.e. $P = \mathrm{conv}(V)$.*

We will omit the proof of this theorems, since it is not instructive for our further proceeding. The proof can be found for instance in [25].

Let us now come back to our initial aim, finding a geometric understanding of the constraints defining the set $\mathscr{P}$ of all behaviors on a given behavior scheme.

**Theorem 3.1.8.** *The set $\mathscr{P}$ is a polytope of dimension* $\dim \mathscr{P} = t' \equiv \left[\sum_{X=0}^{m_{\mathscr{A}}-1}(n_X-1)+1\right]\left[\sum_{Y=0}^{m_{\mathscr{B}}-1}(n_Y-1)+1\right]-1$.

**Proof.**

The dimension of $\mathscr{P}$ is set by Equations (3.2)-(3.4). Let us first show that $\mathscr{P}$ is full-dimensional in $\mathbb{R}^{t'}$ with $t'$ defined above. Consider the vector $p' \in \mathbb{R}^{t'}$:

$$p' = \begin{pmatrix} \mathrm{p}_{a|X} \\ \mathrm{p}_{b|X} \\ \mathrm{p}_{ab|XY} \end{pmatrix} \quad \text{with} \quad a \neq n_X - 1, \quad b \neq n_Y - 1. \tag{3.5}$$

It is easy to see that by use of Equations (3.2)-(3.4) one can recover the full behavior $p$ from $p'$, therefore $\dim \mathscr{P} \leq t'$. Further, Equations (3.2)-(3.4) are trivially accounted in $p'$. Hence, $\dim \mathscr{P} \geq t'$ and therefore $\dim \mathscr{P} = t'$. Now, Equation (3.1) implies that $0 \leq p'_i \leq 1$ and therefore $\mathscr{P}$ is a intersection of half spaces which is clearly bounded. Hence, $\mathscr{P}$ is a polytope.

■

## 3.2 Local and local deterministic behaviors

After having represented the joint probabilities of a given Bell scenario as behaviors on a behavior scheme in the last section, we want to come back to the question motivated in the introduction, namely: Are those behaviors compatible with the description by a local model?

In the introduction we motivated that locality implies another constraint on the probability distributions:

$$\mathrm{p}_{ab|XY} = \int d\lambda q(\lambda) P(a|X,\lambda) P(b|Y,\lambda), \tag{3.6}$$

where $q(\lambda)$, $P(a|X,\lambda)$ and $P(b|Y,\lambda)$ were well-defined, i.e. positive and normalized, distributions. The parameter $\lambda$ was called the hidden variable. Using this locality constraint we define:

**Definition 3.2.1** (Local behavior)**.** *A behavior* $p_{ab|XY}$ *which in addition to conditions (3.1)-(3.4) also satisfies the locality constraint Equation (3.6) is called a local behavior. The set of local behaviors on a given scheme is denoted by* $\mathscr{L}$.

From this definition it is clear that in general $\mathscr{L} \subset \mathscr{P}$. Before going to study the geometrical properties of $\mathscr{L}$ let us first define another type of behavior which will be useful for the study of $\mathscr{L}$:

**Definition 3.2.2** (Local deterministic behavior)**.** *A local behavior for which the probability distributions* $P(a|X,\lambda)$ *and* $P(b|Y,\lambda)$ *in Equation (3.6) can only take values 0 and 1 is called a local deterministic behavior or hidden deterministic behavior.*

In the following we want to give a theorem, due to Fine [26], stating that the assumption of determinism in local behaviors is no more restrictive than that of arbitrary local behaviors.

**Theorem 3.2.3.** *The set of all local deterministic behaviors is equal to the set of all local behaviors* $\mathscr{L}$.

**Proof.**

Every local deterministic behavior is by definition also a local behavior. Let us now start off with an arbitrary local behavior. By introducing two new parameters $\mu, \nu \in [0,1]$ we can define a new hidden variable $\lambda' = (\lambda, \mu, \nu)$ with

$$P'(a|X,\lambda') = \begin{cases} 1 & \text{if } F(a-1|X,\lambda) \leq \mu \leq F(a|X,\lambda), \\ 0 & \text{otherwise,} \end{cases}$$

where we defined $F(a|X,\lambda) := \sum_{\alpha \leq a} P(\alpha|X,\lambda)$ is a new local probability for $\mathscr{A}$. We define a similar one for $\mathscr{B}$ using the parameter $\nu$. If we uniformly randomize over $\mu$ and $\nu$ and define the new hidden variable distributions as $q'(\lambda') = q'(\lambda,\mu,\nu) = q(\lambda)$ we recover the predictions of the original local behavior we started from. However, the new defined behavior is locally deterministic which shows the equivalence.

■

This is a big achievement, since the set of all local deterministic behaviors, where the $P(a|X,\lambda)$ and $P(b|Y,\lambda)$ only take values 0 and 1 is much easier to deal with than with the full expression without restrictions. In particular we can use it to prove the following theorem:

**Theorem 3.2.4.** *Let $\lambda_{\mathscr{A}}$ be an assignment of the outcome of $\mathscr{A}$ measurement to the choice of measurement, i.e. $\lambda_{\mathscr{A}}(X) = a_X$ and accordingly for $\mathscr{B}$, $\lambda_{\mathscr{B}}(Y) = b_Y$. Let $d^{\lambda}_{ab|XY} \in \mathscr{P}$ be the corresponding deterministic probability distributions*

$$d^{\lambda}_{ab|XY} = \begin{cases} 1 & \text{if } \lambda_{\mathscr{A}}(X) = a \text{ and } \lambda_{\mathscr{B}}(Y) = b, \\ 0 & \text{otherwise.} \end{cases} \tag{3.7}$$

*Then the behavior $\mathrm{p}_{ab|XY}$ is local if and only if it can be written as*

$$\mathrm{p}_{ab|XY} = \sum_{\lambda} q_\lambda d^{\lambda}_{ab|XY}, \quad q_\lambda \geq 0, \quad \sum_{\lambda} q_\lambda = 1. \tag{3.8}$$

**Proof.**

Let us set $d^{\lambda}_{ab|XY} = P(a|X,\lambda)P(b|Y,\lambda)$. Clearly there are only finitely many $d^{\lambda}_{ab|XY}$ and hence Equation (3.8) defines a local deterministic behavior. According to Theorem 3.2.3 every local deterministic behavior is equivalently represented by a local behavior. The converse step is similar.

■

From this we can make a statement about the geometrical properties of the set $\mathscr{L}$ of local behaviors.

**Corollary 3.2.5.** *The set $\mathscr{L}$ of local behaviors on a given behavior scheme defines a polytope*

$$\mathscr{L} = \mathrm{conv}\left( d^{\lambda}_{ab|XY} \,\middle|\, \lambda_{\mathscr{A}}(X) \in A_X, \lambda_{\mathscr{B}}(Y) \in B_Y \right), \tag{3.9}$$

*where $d^{\lambda}_{ab|XY}$ are the vertices of $\mathscr{L}$.*

**Proof.**

This is a direct consequence of Theorem 3.2.4, 3.1.6 and 3.1.7.

■

Let us now show that the polytope $\mathscr{L}$ is full-dimensional in $\mathbb{R}^{t'}$, where $t' = \dim \mathscr{P} = \left[\sum_{X=0}^{m_{\mathscr{A}}-1}(n_X - 1) + 1\right]\left[\sum_{Y=0}^{m_{\mathscr{B}}-1}(n_Y - 1) + 1\right] - 1$.

**Theorem 3.2.6.** $\dim \mathscr{L} = \dim \mathscr{P} \equiv t' = \left[\sum_{X=0}^{m_{\mathscr{A}}-1}(n_X - 1) + 1\right]\left[\sum_{Y=0}^{m_{\mathscr{B}}-1}(n_Y - 1) + 1\right] - 1$.

**Proof.**

To prove the claim we have to show that $\mathscr{L}$ has $t'+1$ affinely independent vertices. Let us write $d^{\lambda}_{ab|XY} = d^{\lambda_{\mathscr{A}}}_{a|X} \otimes d^{\lambda_{\mathscr{B}}}_{b|Y}$ with

$$d^{\lambda_{\mathscr{A}}}_{a|X} = \begin{cases} 1 & \text{if } \lambda_{\mathscr{A}}(X) = a, \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad d^{\lambda_{\mathscr{B}}}_{b|Y} = \begin{cases} 1 & \text{if } \lambda_{\mathscr{B}}(Y) = b, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, $d^{\lambda_{\mathscr{A}}}_{a|X}$ consists of $\sum_{X=0}^{m_{\mathscr{A}}-1}(n_X-1)+1$ linear independent vertices and $d^{\lambda_{\mathscr{B}}}_{b|Y}$ consists of $\sum_{Y=0}^{m_{\mathscr{B}}-1}(n_Y-1)+1$ linear independent vertices. Hence, $d^{\lambda}_{ab|XY}$ has $t'+1$ linear independent and thus also affinely independent vertices which completes the proof.

■

## 3.3 Bell inequalities

### 3.3.1 Generalities

In the last two sections we have derived several properties of the set $\mathscr{P}$ of all behaviors and the set $\mathscr{L}$ of all local behaviors. In the following we want to use this knowledge to determine a procedure which enables us to distinguish local from non local behaviors.

We have seen that both $\mathscr{P}$ and $\mathscr{L}$ are polytopes of the same dimension. Further, we know that in general $\mathscr{L} \subset \mathscr{P}$. Figure 3.2 shows a lower dimensional sketch of these quantities which gives us a good understanding of the relations between $\mathscr{P}$ and $\mathscr{L}$. The vertices of $\mathscr{L}$ are the deterministic behaviors $d^{\lambda}_{ab|XY}$ and

$$\mathscr{L} = \operatorname{conv}\left( d^{\lambda}_{ab|XY} \,\middle|\, \lambda_{\mathscr{A}}(X) \in A_X, \lambda_{\mathscr{B}}(Y) \in B_Y \right). \tag{3.10}$$

However, form Definition 3.1.4 we know that the polytope $\mathscr{L}$ can equivalently be written as an intersection of finitely many half spaces

$$\mathscr{L} = \bigcap_{j=1}^{m} \left\{ x \in \mathbb{R}^{t'} \,\middle|\, \sum_i A_{ij} x_i \geq c_j \right\} \tag{3.11}$$

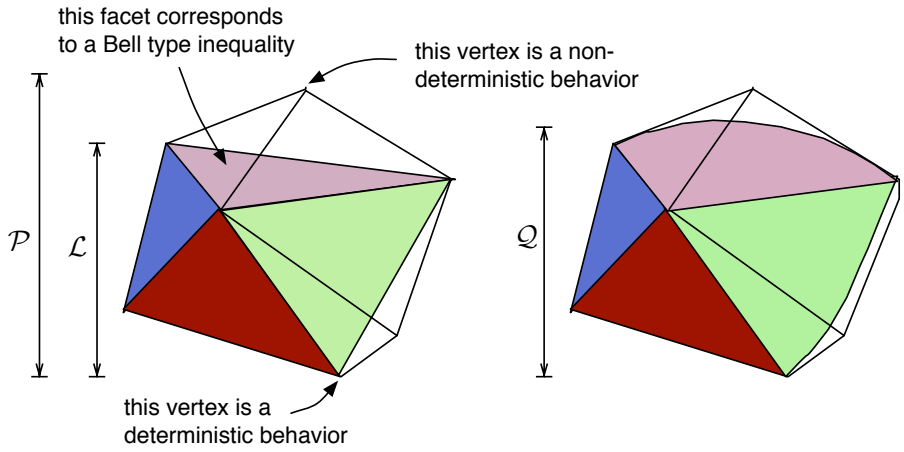


Figure 3.2: Left: A three-dimenional illustration of the space of local behaviors $\mathscr{L} \subset \mathscr{P}$ which is a polytope. Right: A illustration of the convex body of quantum behaviors $\mathscr{Q} \subset \mathscr{P}$.

If this description is minimal the inequalities $\sum_i A_{ij} x_i \geq c_j$ define the facets of $\mathscr{L}$, those inequalities are called extremal. Further, they provide us with a way of determine whether a given behavior is local or not: Every behavior $x \in \mathscr{P}$ that obeys the extremal inequalities $\sum_i A_{ij} x_i \geq c_j$ is local and every behavior that violates one of them is nonlocal. This inequalities are the so-called Bell inequalities which we already motivated in the introduction.

In the following we want to derive several type of Bell inequalities for several types of Bell scenarios.

For the derivation of the Bell inequalities it is useful to notice that according to Theorem 3.2.4 every element $p_{ab|XY} \in \mathscr{L}$ may be view as a (classical) probability distribution for the setting $\mathscr{S} = (X_0, ..., X_{m_{\mathscr{A}}-1}, Y_0, ..., Y_{m_{\mathscr{B}}-1})$ over the outcome space $\Omega = A_{X_0} \times ... \times A_{X_{m_{\mathscr{A}}-1}} \times A_{Y_0} \times ... \times A_{Y_{m_{\mathscr{B}}-1}}$. The different $X_i$ and $Y_j$ are then random variables on the different $A_i$ and $B_j$ respectively. We will see in the following how we can use this property to derive several Bell inequalities.

### 3.3.2 Explicit examples of Bell inequalities

In this section we want to derive several Bell type inequalities for different Bell scenarios. In all considered cases the number of possible measurements will be

equal for the two parties $\mathscr{A}$ and $\mathscr{B}$, hence we set $m = m_{\mathscr{A}} = m_{\mathscr{B}}$. Further, we simplify the scenario by requiring all measurements to have the same number of possible outcomes $n = n_{X_i} = n_{X_j}$. Hence we characterize a Bell scenario by the following tuple $(2,m,n)$, meaning that two parties, $\mathscr{A}$ and $\mathscr{B}$, each have a choice of $m$ measurements for which each measurement can have $n$ possible outcomes.

Let us first consider the scenario from the introduction.

### The CHSH inequality $(2,2,2)$

Let us consider the Bell scenario $(2,2,2)$, of two parties $\mathscr{A}$ and $\mathscr{B}$ which can both perform two different measurements, $X_0, X_1$ and $Y_0, Y_1$. Each of those four measurements can have two different outcomes which we label 0 and 1. As discussed earlier every local behavior can be viewed as a probability distribution for $(X_0, X_1, Y_0, Y_1)$ over the outcome space $\Omega = \{0,1\}^4$. We will now proof that this implies the following inequality of each probability distribution $\mathrm{p}_{ab|XY} \in \mathscr{L}(2,2,2)$:

**Theorem 3.3.1** (CHSH inequality)**.** *Every local behavior* $\mathrm{p}_{ab|XY} \in \mathscr{L}(2,2,2)$ *obeys the following inequality:*

$$\Pr(X_1 \neq Y_1) \leq \Pr(X_0 \neq Y_0) + \Pr(X_0 \neq Y_1) + \Pr(X_1 \neq Y_0), \tag{3.12}$$

*where* $\Pr(X_i \neq Y_j) = \mathrm{p}_{01|X_iY_j} + \mathrm{p}_{10|X_iY_j}$.

**Proof.**

As described above every local behavior $\mathrm{p}_{ab|XY} \in \mathscr{L}(2,2,2)$ is a probability distribution for $(X_0, X_1, Y_0, Y_1)$ over the outcome space $\Omega = \{0,1\}^4$. Let us consider events on this outcome space of the following type, $(X_i = Y_j)$. The complementary to this event would be $(X_i \neq Y_j)$. Let us start with is following implication

$$[(X_0 = Y_0) \cap (X_0 = Y_1) \cap (X_1 = Y_0)] \subset (X_1 = Y_1)$$

Taking the complementary of this equation gives $(X_1 \neq Y_1) \subset [(X_0 \neq Y_0) \cup (X_0 \neq Y_1) \cup (X_1 \neq Y_0)]$. According to our basic calculation rules (Lemma 2.1.6) this implies Equation (3.12) which completes the proof.

■

This inequality was first derived in [5] in a slightly modified version. Clearly, this inequality stands for a whole class of inequalities which can be generated form this one by permutation of the outcomes. Form our geometrical point of view it is interesting to note that $\mathscr{L}(2,2,2)$ defines the simplest non-trivial polytope. Further, it had been shown that this inequality and permutations of it uniquely define the facet inequalities of this polytope [26].

This inequality provides us with a test of seeing whether an underlying probability distribution of an experiment (actual or only gedankenexperiment) arises from a local model or not. As we have already seen in the introduction certain probability distributions of quantum statistics actually violate this inequality.

Before continuing with the next example of a Bell type inequality let us first show that there is a different form of writing the CHSH inequality which will be useful for later considerations. In this version the outcomes of the measurements are considered to be $\{-1,1\}$. In the form of the inequality Equation (3.12) the actual value or label of the outcomes does not matter. However, we are now going to rewrite it in terms of expectation values of the random variables $(X_0,X_1,Y_0,Y_1)$. Recall that the expectation value of a random variable is given by

$$\mathbb{E}(X) = \sum_{x\in\Omega} x \Pr(X = x).$$

Let us now rewrite Equation (3.12) using that $\Pr(X_i \neq Y_j) = 1 - \Pr(X_i = Y_j)$, yielding

$$\Pr(X_0 = Y_0) + \Pr(X_0 = Y_1) + \Pr(X_1 = Y_1) - \Pr(X_1 = Y_1) \leq 2.$$

Subtracting from this equation Equation (3.12) gives

$$\begin{aligned} &\Pr(X_0 = Y_0) + \Pr(X_0 = Y_1) + \Pr(X_1 = Y_0) - \Pr(X_1 = Y_1) - \\ &- \Pr(X_0 \neq Y_0) - \Pr(X_0 \neq Y_1) - \Pr(X_1 \neq Y_0) + \Pr(X_1 \neq Y_1) \quad \leq \quad 2 \end{aligned} \tag{3.13}$$

Since the random variables $(X_0,X_1,Y_0,Y_1)$ only take values of $\Omega = \{-1,1\}^4$ we can write for the expectation value

$$\mathbb{E}(X_iY_j) = \Pr(X_i = Y_j) - \Pr(X_i \neq Y_j). \tag{3.14}$$

Hence, Equation (3.28) reads

$$\mathbb{E}(X_0Y_0) + \mathbb{E}(X_0Y_1) + \mathbb{E}(X_1Y_0) - \mathbb{E}(X_1Y_1) \leq 2. \tag{3.15}$$

### The CGLMP inequality $(2,2,d)$

Let us now consider the case where there are $d$ possible outcomes for each measurement. We will label these outcomes by $(0,...,d-1)$. Hence, $(X_0,X_1,Y_0,Y_1)$ can be seen as random variables over the outcome space $\Omega = \{0,...,d-1\}^4$. We will now show that in this Bell scenario the requirement of locality implies the following Bell type inequality:

**Theorem 3.3.2** (CGLMP inequality)**.** *Every local behavior* $\mathrm{p}_{ab|XY} \in \mathscr{L}(2,2,d)$ *obeys the following inequality:*

$$\mathbb{E}\left([X_1 - Y_1] + [Y_1 - X_0] + [X_0 - Y_0] + [Y_0 - X_1 - 1]\right) \geqslant d-1, \tag{3.16}$$

*where* $[X]$ *denotes the X modulo d.*

**Proof.**

Let us start with the following trivial statement

$$[X_1 - Y_1 + Y_1 - X_0 + X_0 - Y_0 + Y_0 - X_1 - 1] = [-1] = d-1.$$

Clearly we have $[X+Y] \leqslant [X] + [Y]$ which completes the proof.

■

This inequality was first derived in [27] in a more complicated form. The compact form presented here is due to [28]. This is the inequality we will be most interested in in the remainder of this thesis. As we will show in the following there exists a equivalent form of this inequality in terms of probabilities. Let us first derive this different formulation independently and then show the equality to Theorem 3.3.2.

**Theorem 3.3.3** (CGLMP inequality (equivalent formulation))**.** *Every local behavior* $\mathrm{p}_{ab|XY} \in \mathscr{L}(2,2,d)$ *obeys the following inequality:*

$$\Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(Y_0 \leq X_1) \geqslant 1, \tag{3.17}$$

*where* $\Pr\left(X_i < Y_j\right) = \sum_{a<b} \mathrm{p}_{ab|X_iY_j}$.

**Proof.**

Let us start with the following statement

$$(X_1 \geq Y_1) \cap (Y_1 \geq X_0) \cap (X_0 \geq Y_0) \subset (X_1 \geq Y_0)$$

Taking the complementary we get $(X_1 < Y_0) \subset (X_1 < Y_1) \cup (Y_1 < X_0) \cup (X_0 < Y_0)$. This implies for the probabilities that

$$\begin{aligned} \Pr(X_1 < Y_0) &= 1 - \Pr(X_1 \geq Y_0) \\ &\leq \Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) \end{aligned}$$

which completes the proof.

■

Let us now show the equality of this Equation (3.16) and Equation (3.17). Starting form Equation (3.16) we note that

$$[X] = X - d \left\lfloor \frac{X}{d} \right\rfloor,$$

where we defined the Gauss bracket $\lfloor X \rfloor = \max\{Y \in \mathbb{Z} | Y \leq X\}$. Then we can write

$$\begin{aligned} & [X_1 - Y_1] + [Y_1 - X_0] + [X_0 - Y_0] + [Y_0 - X_1 - 1] = \\ = \; & -1 - d \left( \left\lfloor \frac{X_1 - Y_1}{d} \right\rfloor + \left\lfloor \frac{Y_1 - X_0}{d} \right\rfloor + \left\lfloor \frac{X_0 - Y_0}{d} \right\rfloor + \left\lfloor \frac{Y_0 - X_1 - 1}{d} \right\rfloor \right) \end{aligned}$$

Since the $X_i$ and $B_j$ only take possible values in $\{0, 1, ..., d-1\}$, we get using Equation (3.16)

$$\Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(Y_0 < X_1 + 1) \geq 1,$$

which is equivalent to Equation (3.17) for outcomes in $\{0, 1, ..., d-1\}$.

One sees that for the outcomes being in $\{0, 1, ..., d-1\}$ Equation (3.16) and Equation (3.17) are equivalent. However, whereas Equation (3.16) is only true for those outcomes, Equation (3.17) hold for arbitrary values of the outcomes. Only the relative ordering of the outcomes on the real line matters.

In Theorem 3.3.2 we have seen an explicit Bell type inequality for the $\mathscr{L}(2,2,d)$ Bell scenario. Further, it has been shown in [29] that this inequality, in its original formulation of [27], is indeed facet defining for the polytope $\mathscr{L}(2,2,d)$.

**Other cases**

Several other Bell type inequality for different Bell scenarios have been discussed within the literature. However, in this thesis we are only considered with the CHSH inequality and especially its generalization for general outputs, the CGLMP inequality. Another well known example is the three party scenario $(3,2,2)$ from which arises the GHZ paradox [30, 31] and Mermin's corresponding Bell type inequality [32]. However, we will not go into the details of these different Bell scenarios. One big advantage of the CHSH and the CGLMP over the other setting is that they are much easier realizable in actual physical experiments.

## 3.4 Quantum behaviors and nonlocality proofs

In this section we want to introduce the definition of quantum behaviors and its basic properties. In the previous sections we have studied the set of all local behavior $\mathscr{L} \subset \mathscr{P}$. Besides the various geometric properties of the polytope $\mathscr{L}$ which gave a greater understanding of it we also noticed that elements of $\mathscr{L}$ can be viewed as classical probability distributions for certain random variables. We will see in the following that quantum behaviors play a similar role in the framework of quantum probability, where now the random variables are replaced by observables.

As in the previous sections we only consider Bell scenarios of two parties $\mathscr{A}$ and $\mathscr{B}$. Let us now give a minimalists, meaning most general, definition of a quantum behavior:

**Definition 3.4.1** (Quantum behavior I)**.** *A quantum behavior is a joint probability distribution* $\mathrm{p}_{ab|XY}$ *with the following properties:*

$$\mathrm{p}_{ab|XY} = \mathrm{tr}\left(\rho E_{X,a} F_{Y,b}\right) \tag{3.18}$$

*where* $\rho$ *is a density matrix and* $E_{X,a}, F_{Y,b}$ *are operators on a Hilbert space* $\mathscr{H}$*, satisfying*

$$E_{X,a} \geq 0, \quad F_{Y,b} \geq 0, \quad \forall a,b,X,Y, \tag{3.19}$$

$$\sum_{a \in A_X} E_{X,a} = \mathbf{1}, \quad \sum_{b \in B_Y} F_{Y,b} = \mathbf{1}, \quad \forall X,Y, \tag{3.20}$$

$$E_{X,a} F_{Y,b} = F_{Y,b} E_{X,a}, \quad \forall a,b,X,Y. \tag{3.21}$$

*The set of all quantum behaviors is called* $\mathscr{Q}$*.*

Let us comment on this definition: As discussed earlier in Section 2.4 Gleason's theorem shows that an equation of type Equation (3.18) is the only possibility of defining a measure on $(\mathscr{X},\mathscr{A})$ and hence defining a "quantum probability" $\mathrm{p}_{ab|XY}$. According to Definition 2.4.1 Equations 3.19 and 3.20 define operator valued probability measures, namely the measurements $E_X = \{E_{X,a}\}_{a\in A_X}$, $X \in 0,...,m_{\mathscr{A}}-1$ for $\mathscr{A}$ and the measurements $F_Y = \{F_{Y,b}\}_{b\in B_Y}$, $Y \in 0,...,m_{\mathscr{B}}-1$ for $\mathscr{B}$. Equation (3.21) stays for the fact that the measurements on $\mathscr{A}$'s side do not interact with those on $\mathscr{B}$'s side and vice versa. Under this considerations Definition 3.4.1 is a natural definition for a joint probability arising from quantum statistics. Also note that in general different $E_{X_i}, E_{X_j}$ do not commute and similar for measurements on $\mathscr{B}$'s side.

Having given the minimal definition of a quantum behavior let us now give another, maximal definition of a quantum behavior:

**Definition 3.4.2** (Quantum behavior II)**.** *A quantum behavior is a joint probability distribution* $\mathrm{p}_{ab|XY}$ *with the following properties:*

$$\begin{aligned} \mathrm{p}_{ab|XY} &= \langle\psi|\Pi_{X,a}\otimes\Pi_{Y,b}|\psi\rangle \\ &= \mathrm{tr}(\rho\Pi_{X,a}\otimes\Pi_{Y,b}), \quad \textit{with} \quad \rho = |\psi\rangle\langle\psi|, \end{aligned} \tag{3.22}$$

*where* $\psi$ *is an entangled pure state on the product system,*

$$\psi \in \mathscr{H}_{\mathscr{A}}\otimes\mathscr{H}_{\mathscr{B}}, \quad \langle\psi|\psi\rangle = 1, \tag{3.23}$$

*and the* $\Pi_{X,a}$ *and* $\Pi_{Y,b}$ *are orthogonal projectors, i.e. simple measurements, on* $\mathscr{H}_{\mathscr{A}}$ *and* $\mathscr{H}_{\mathscr{B}}$ *respectively,*

$$\Pi_{X,a} : \mathscr{H}_{\mathscr{A}} \to \mathscr{H}_{\mathscr{A}}, \quad \Pi_{X,a}^{\dagger} = \Pi_{X,a}, \quad \Pi_{X,a}^2 = \Pi_{X,a}, \quad \forall a, X, \tag{3.24}$$

$$\Pi_{Y,b} : \mathscr{H}_{\mathscr{B}} \to \mathscr{H}_{\mathscr{B}}, \quad \Pi_{Y,b}^{\dagger} = \Pi_{Y,b}, \quad \Pi_{Y,b}^2 = \Pi_{Y,b}, \quad \forall b, Y, \tag{3.25}$$

$$\sum_{a\in A_X}\Pi_{X,a} = \mathbf{1}, \quad \sum_{b\in B_Y}\Pi_{Y,b} = \mathbf{1}, \quad \forall X, Y. \tag{3.26}$$

This definition is much more restrictive then Definition 3.4.1 in several aspects. First, in Definition 3.4.1 the state could have been mixed, whereas in the later definition it is pure. Secondly, the measurements in Definition 3.4.1 were generalized measurements while now they are simple. Further, the algebra in Definition 3.4.1 was not necessarily a tensor product while in the later definition it is. Although, Definition 3.4.2 is much more restrictive we still have the following theorem (cf. [23]):

**Theorem 3.4.3.** *The set of all quantum behaviors generated form Definition 3.4.1 is equal to that of Definition 3.4.2.*

For the proof of this theorem the reader is referred to [23] and references therein. The proof involves the theory of $C^*$-algebras. In the following we will use Definition 3.4.2 as a definition for a quantum behavior, since it is much simpler. It also gives us a much more natural interpretation of quantum behaviors. According to the considerations in Section 2.3 we can associate to every projector $\Pi_{X,a}$ an observable $X$ which has $\Pi_{X,a} \equiv \Pi_{[X=a]}$ as an eigenspace and respectively for $\Pi_{Y,b}$. Hence, we can write

$$X = \sum_{a \in A_X} a \Pi_{[X=a]}, \quad Y = \sum_{b \in B_Y} b \Pi_{[Y=b]}. \tag{3.27}$$

Thus we see that elements $\mathrm{p}_{ab|XY} \in \mathscr{Q}$ form a (quantum) probability distribution for the observables $(X_0, ..., X_{m_{\mathscr{A}}-1}, Y_0, ..., Y_{m_{\mathscr{B}}-1})$ over the outcome space $\Omega = A_{X_0} \times ... \times A_{X_{m_{\mathscr{A}}-1}} \times A_{Y_0} \times ... \times A_{Y_{m_{\mathscr{B}}-1}}$ exactly as elements $\mathrm{p}_{ab|XY} \in \mathscr{L}$ formed (classical) probability distribution for the random variables $(X_0, ..., X_{m_{\mathscr{A}}-1}, Y_0, ..., Y_{m_{\mathscr{B}}-1})$. We see that investigating nonlocality proofs also brings us to the heart of analyzing the differences between classical and quantum probability theory. In this sense, the fact that $\mathscr{L} \subseteq \mathscr{Q}$, as will be proven in the following, also emphasizes that quantum probability theory is a (non-commutative) extension to classical probability theory which includes the latter as a special case.

Let us now deduce some properties of the set of all quantum behaviors $\mathscr{Q}$.

**Theorem 3.4.4.** *The set of all quantum behaviors is a subset of or equal to the set of all behaviors, $\mathscr{Q} \subseteq \mathscr{P}$.*

**Proof.**

Clearly, due to the properties of the trace rule (Definition 2.4.2) and the properties of the measurements $\Pi_X = \{\Pi_{X,a}\}_{a \in A_X}$ and $\Pi_Y = \{\Pi_{Y,b}\}_{b \in B_Y}$ every element of $\mathscr{Q}$ is positive and normalized. Further, due to the tensor product structure and the properties of the trace (Theorem 2.3.16) it follows easily that elements of $\mathscr{Q}$ also fulfill the non-signaling condition. Hence, by Definition 3.1.1 we have $\mathscr{Q} \subseteq \mathscr{P}$.

■

Note that in general we have $\mathscr{Q} \subset \mathscr{P}$, as will also be clear from the next section. Further, we have the following relation with the set of local behaviors:

**Theorem 3.4.5.** *The set of all local behaviors is a subset of or equal to the set of all quantum behaviors,* $\mathscr{L} \subseteq \mathscr{Q}$.

**Proof.**

This is most easily seen using Definition 3.4.2. Let us only consider the case of finitely many hidden variables. The infinite dimensional case is similar. Let $d$ denote the number of hidden variables, then we choose $\mathscr{H} = \mathbb{C}^d$. Let the index be denoted by $\lambda = 0, .., d-1$. If we choose all measurements as diagonal operators with $E_{X,a} = \text{diag}(P(a|X,0), ..., P(a|X,\lambda), ...)$ and similarly for $F_{Y,b}$ and further choose $\rho = \text{diag}(q(0), ..., q(\lambda), ...)$, we recover, according to Definition 3.2.1, all elements of $\mathscr{L}$ from $\mathscr{Q}$ for this special choice. Hence, we have $\mathscr{L} \subseteq \mathscr{Q}$.

■

In general we also have $\mathscr{L} \subset \mathscr{Q}$. Let us now investigate the geometry of the set $\mathscr{Q}$. Using the above two theorems we can already make a statement about the dimension of $\mathscr{Q}$.

**Corollary 3.4.6.** $\dim \mathscr{Q} = t' = \left[\sum_{X=0}^{m_{\mathscr{A}}-1}(n_X - 1) + 1\right]\left[\sum_{Y=0}^{m_{\mathscr{B}}-1}(n_Y - 1) + 1\right] - 1.$

**Proof.**

We have already shown in Theorem 3.2.6 that $\dim \mathscr{P} = \dim \mathscr{L} = t'$ with $t'$ given above. From the last two theorems it follows that $\mathscr{L} \subseteq \mathscr{Q} \subseteq \mathscr{P}$ and hence $\dim \mathscr{L} \leq \dim \mathscr{Q} \leq \dim \mathscr{P}$ from which follows $\dim \mathscr{L} = \dim \mathscr{Q} = \dim \mathscr{P} = t'$.

■

What else can we learn about the geometrical properties of the set $\mathscr{Q}$? Clearly, the geometry of $\mathscr{Q}$ depends crucially on the properties of measurements, i.e. oprms on $(\mathscr{X}, \mathscr{A})$. Let us give the following theorem which we have omitted in Section 2.4:

**Theorem 3.4.7.** *The set of all oprms on* $(\mathscr{X}, \mathscr{A})$ *is convex, i.e. if* $M_1, M_2$ *are oprms on* $(\mathscr{X}, \mathscr{A})$ *so is* $M_3 = \lambda M_1 + (1-\lambda)M_2$ *for* $0 \leq \lambda \leq 1$.

**Proof.**

Recall from Definition 2.4.1 that condition (iii), i.e. for disjoint $A_n \in \mathscr{A}$, $M_3\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} M_3(A_n)$, hold if $M_3$ is any linear combination of $M_1$ and $M_2$. However, conditions (i) and (ii), i.e. $M_3(\mathscr{X}) = 1$ and $\mathbf{0} \leq M_3 \leq \mathbf{1}$, only hold if $M_3$ is a convex combination of $M_1, M_2$.

■

From this theorem we can make the following conclusion for the set $\mathscr{Q}$:

**Corollary 3.4.8.** *The set $\mathscr{Q}$ of quantum behaviors is convex.*

**Proof.**

We want to show that if $p^1, p^2 \in \mathscr{Q}$, i.e. $\mathrm{p}^1_{ab|XY} = \mathrm{tr}(\rho^1 \Pi^1_{X,a} \otimes \Pi^1_{Y,b})$ and $\mathrm{p}^2_{ab|XY} = \mathrm{tr}(\rho^2 \Pi^2_{X,a} \otimes \Pi^2_{Y,b})$, then also $p^3 = \lambda p^1 + (1-\lambda) p^2 \in \mathscr{Q}$ for $0 \leq \lambda \leq 1$, where the $\Pi_X \otimes \mathbf{1}$, $\mathbf{1} \otimes \Pi_Y$ and $\rho$'s are all operators on the same Hilbert space $\mathscr{H} = \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$. Since $\rho^1$ and $\rho^2$ are both density matrices we can write $\rho^1 = U^\dagger \rho^2 U$, where $U$ is a unitary operator, i.e. $U^\dagger U = \mathbf{1}$. Using the properties of the trace we can then write

$$\begin{aligned}
\mathrm{p}^3_{ab|XY} &= \lambda \mathrm{p}^1_{ab|XY} + (1-\lambda)\mathrm{p}^2_{ab|XY} \\
&= \mathrm{tr}\left(\lambda \rho^1 \Pi^1_{X,a} \otimes \Pi^1_{Y,b} + (1-\lambda)\rho^2 \Pi^2_{X,a} \otimes \Pi^2_{Y,b}\right) \\
&= \mathrm{tr}\left(\rho^1 \left[\lambda \Pi^1_{X,a} \otimes \Pi^1_{Y,b} + (1-\lambda) U \Pi^2_{X,a} \otimes \Pi^2_{Y,b} U^\dagger\right]\right). \\
&= \mathrm{tr}\Big(\rho^1 \Big[\Big(\lambda \Pi^1_{X,a} + (1-\lambda) U_{\mathscr{A}} \Pi^2_{X,a} U^\dagger_{\mathscr{A}}\Big) \otimes \\
&\qquad \otimes \Big(\lambda \Pi^1_{Y,b} + (1-\lambda) U_{\mathscr{B}} \Pi^2_{Y,b} U^\dagger_{\mathscr{B}}\Big)\Big]\Big),
\end{aligned}$$

where $U = U_{\mathscr{A}} \otimes U_{\mathscr{B}}$ is the decomposition of $U$ on $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$. $U_{\mathscr{A}}$ and $U_{\mathscr{B}}$ are again unitary operators on $\mathscr{H}_{\mathscr{A}}$ and $\mathscr{H}_{\mathscr{B}}$ respectively. It is readily checked that if $M$ is an oprm so is $UMU^\dagger$, where $U$ is unitary. Hence, the quantity in the bracket is a convex combinations of oprms which according to Theorem 3.4.7 again is a oprm. This completes the proof.

■

Let us summarize what we have gained so far. We have seen that in general $\mathscr{L} \subset \mathscr{Q} \subset \mathscr{P}$, where all sets have the same dimension (cf. Figure 3.2). Further, we showed that $\mathscr{Q}$ is a convex set. However, in contrary to $\mathscr{L}$ and $\mathscr{P}$, the set $\mathscr{Q}$ is not a ploytope. Hence, $\mathscr{Q}$ cannot be written as a convex hull of finitely many vertices or as an intersection of finitely many half spaces, specified by linear inequalities. However, since $\mathscr{Q}$ is convex it can still be fully characterized in terms of its boundary, this might be through infinitely extreme points or through a collection of non-linear inequality specifying the boundary. For the specific

case of the (2,2,2) Bell scenario a complete description of the boundary of $\mathscr{Q}$ has been given in terms of non-linear inequalities [33, 34, 29] which will be partly discussed in the following section. However, for other Bell scenarios like the (2,2,d) scenario very little is known.

Let us now come back to the importance of Bell inequalities. We have seen that in general $\mathscr{L} \subset \mathscr{P}$. This means that there are certain elements of $\mathscr{Q}$ which do not arise from a local model. Form a physical point of view this is a very interesting observation as already discussed in the introduction. Since quantum mechanics as a theory of nature like presented in Section 2.6 can produce such joint probabilities for certain settings, this means that quantum mechanics is a non-local theory. Bell inequalities are in this sense important, since they can be used to prove that quantum mechanics is non-local. As being facet defining for $\mathscr{L}$ the Bell inequalities define the boundary between the set $\mathscr{L}$ and $\mathscr{Q}$. Hence, an element $q \in \mathscr{Q}$ which is not element of $\mathscr{L}$ would violate a given Bell inequality, whereas an element of $\mathscr{L}$ would not. Let us make the above discussion precise:

**Definition 3.4.9** (Nonlocality proof)**.** *We call an element of $q \in \mathscr{Q}$ for a given Bell-scenario a nonlocality proof. Further, the nonlocality proof is called proper if $q$ is not element of $\mathscr{L}$, i.e. if it violates a given Bell type inequality.*

In the introduction we have already seen an explicit example of a nonlocality proof, namely a state $\rho$ together with a set of measurements of spin in certain directions producing a joint probability $q \in \mathscr{Q}$ which violated the CHSH inequality. The question one might ask at this point is: How big or how strong is this violation? Knowing the answer to this questions might have several implications. Firstly, from a physical point of view an experimentalist who would like to show in an experiment that quantum mechanics is nonlocal would prefer an nonlocality proof which produces the maximal violation such that he can clearly distinguish whether the measured probability distribution arose from a local theory or not. Secondly, since $\mathscr{Q}$ is convex we know that a linear function like the one defining a Bell inequality reaches its maximum on the boundary of $\mathscr{Q}$. Hence, knowing the maximum violation of a Bell inequality will give us further insides in the form of the boundary of $\mathscr{Q}$. Conversely, for the $(2,2,2)$ we stated earlier that a complete description of the boundary of $\mathscr{Q}$ is known and we will see in the following section how this enables us to give an optimal nonlocality proof for the CHSH inequality. Optimizing nonlocality proofs for the case of the $(2,2,d)$ Bell scenario of the CGLMP inequality is mayor aspect of this thesis in will be investigated in Chapter 4.

## 3.5 A quantum Bell inequality and Grothendieck's inequality

In this section we will further investigate the $(2,2,2)$ Bell scenario for the CHSH inequality. As already mentioned in the previous section the set $\mathscr{Q}(2,2,2)$ can be full described by a set of non-linear functions in this specific case, and due to the convexity one should be able to find a maximal violation of the CHSH inequality for elements of $\mathscr{Q}$.

Recall from Section 3.3.2 that the CHSH inequality can be written as follows:

$$\mathbb{E}(X_0Y_0)+\mathbb{E}(X_0Y_1)+\mathbb{E}(X_1Y_0)-\mathbb{E}(X_1Y_1)\leq 2, \tag{3.28}$$

where the random variables $X_0,X_1,Y_0,Y_1$ can take values in $\{-1,1\}$. Clearly, there exists an element of $\mathscr{P}$ for which the LHS of Equation (3.28) becomes 4 and which therefore violates this inequality. But what is the maximal violation for elements of $\mathscr{Q}\subseteq\mathscr{P}$? In the introduction we have already seen that there exists an element of $\mathscr{Q}$ for which the LHS of Equation (3.28) becomes $2\sqrt{2}$. In the following we want to show that this is actually the maximal violation.

According to Definition 3.4.2 every element of $\mathscr{Q}$ can be written as

$$\mathrm{p}_{ab|XY}=\langle\psi|\Pi_{X,a}\otimes\Pi_{Y,b}|\psi\rangle, \tag{3.29}$$

where $|\psi\rangle$ is a entangled state on $\mathscr{H}=\mathscr{H}_{\mathscr{A}}\otimes\mathscr{H}_{\mathscr{B}}$ and the $\Pi_{X_0},\Pi_{X_1}$ are orthogonal projectors onto the eigenspaces of the observables $X_0,X_1$ on $\mathscr{H}_{\mathscr{A}}$ and accordingly for $Y$. For the scenario $(2,2,2)$ we require in addition that the spectrum of the observables is $\mathrm{spec}(X_i),\mathrm{spec}(Y_j)\in[-1,1]$ for $i,j=0,1$. In other words this means that

$$X_i^2=\mathbf{1},\quad Y_j^2=\mathbf{1},\quad \forall i,j. \tag{3.30}$$

Using this simple property we can give the following theorem due to []:

**Theorem 3.5.1** (Tsirelson's inequality)**.** *All* $\mathrm{p}_{ab|XY}\in\mathscr{Q}(2,2,2)$ *satisfy the following inequality*

$$\mathbb{E}(X_0Y_0)+\mathbb{E}(X_0Y_1)+\mathbb{E}(X_1Y_0)-\mathbb{E}(X_1Y_1)\leq 2\sqrt{2}, \tag{3.31}$$

*where*

$$\mathbb{E}\left(X_iY_j\right)=\sum_{a,b\in\{-1,1\}}ab\,\mathrm{p}_{ab|X_iY_j}. \tag{3.32}$$

**Proof.**

Let us simplify the following term involving the operators operators $X_0, X_1, Y_0, Y_1$,

$$\begin{aligned} X_0Y_0 + X_0Y_1 + X_1Y_0 - X_1Y_1 &= \frac{1}{\sqrt{2}}\left(X_0^2 + X_1^2 + Y_0^2 + Y_1^2\right) - \\ - \quad & \frac{\sqrt{2}-1}{8}\left((\sqrt{2}-1)(X_0 - Y_0) + X_1 - Y_1\right)^2 - \\ - \quad & \frac{\sqrt{2}-1}{8}\left((\sqrt{2}-1)(X_0 - Y_1) - X_1 - Y_0\right)^2 - \\ - \quad & \frac{\sqrt{2}-1}{8}\left((\sqrt{2}-1)(X_1 - Y_0) + X_0 + Y_1\right)^2 - \\ - \quad & \frac{\sqrt{2}-1}{8}\left((\sqrt{2}-1)(X_1 - Y_1) - X_1 - Y_1\right)^2 - \\ \leq \quad & \frac{1}{\sqrt{2}}\left(X_0^2 + X_1^2 + Y_0^2 + Y_1^2\right) \leq 2\sqrt{2}\cdot\mathbf{1}, \end{aligned} \tag{3.33}$$

where in the last step we used that $X_i^2 = \mathbf{1}$ and $Y_j^2 = \mathbf{1}$. Taking the expectation value on both sides and using that $\mathbb{E}(X) + \mathbb{E}(Y) \leq \mathbb{E}(X+Y)$ gives what has to be proven.

■

From a mathematical point of view it is very interesting to see that this theorem also connects to the problem of determine Grothendieck's constant [35] which is known from Banach space theory. The connection was first seen in [33] and relies on the following definition and the two following theorems [36, 33]:

**Definition 3.5.2** (Quantum correlation matrix)**.** *For any quantum behavior* $\mathrm{p}_{ab|XY} \in \mathscr{Q}(2,m,2)$*, i.e. where we have* $X_i^2 = \mathbf{1}$ *and* $Y_j^2 = \mathbf{1}$*, we call* $c = \{c_{ij}\}$ *with*

$$c_{ij} = \mathbb{E}\left(X_i Y_j\right) \equiv \sum_{a,b\in\{-1,1\}} a\,b\,\mathrm{p}_{ab|X_iY_j}, \quad i,j = 0,\ldots,m-1,$$

*a quantum correlation matrix. The set of all quantum correlation matrices is called* $M_{\mathscr{Q}}$*.*

The following theorem is of special importance:

**Theorem 3.5.3.** *A matrix* $c = \{c_{ij}\}$ *is a quantum correlations matrix, i.e.* $c \in M_{\mathscr{Q}}$ *if and only if it admits a representation*

$$c_{ij} = \vec{x}_i \cdot \vec{y}_j, \quad i,j = 0,\ldots,m-1,$$

*where $\vec{x}_i$ and $\vec{y}_j$ are unit vectors in Euclidean space of dimension $2m$.*

For the proof the reader is referred to [33]. From the properties of $\mathscr{Q}$ one can see that $M_{\mathscr{Q}}$ is a bounded convex body in the $\mathbb{R}^{m\times m}$, exactly as $\mathscr{Q}$ was in $\mathbb{R}^{t'}$. Further, $M_{\mathscr{Q}}$ is not a polytope. Nevertheless, it admits a representation as a convex hull over a (infinite) set of extremal points, $M_{\mathscr{Q}} = \mathrm{conv}(\mathrm{ex}\, M_{\mathscr{Q}})$. These extremal points are of special interest, as we will see in the following. Let now state a theorem without proof [33]:

**Theorem 3.5.4.** *An extremal quantum correlations matrix $c \in \mathrm{ex}\, M_{\mathscr{Q}}$ admits a representation*

$$c_{ij} = \vec{x}_i \cdot \vec{y}_j, \quad i,j = 0,...,m-1,$$

*where $\vec{x}_i$ and $\vec{y}_j$ are unit vectors in Euclidean space of dimension $r$. Here $r$ is called the rank of $c$ an it satisfies the following inequalities,*

$$r \leq m \quad \textit{and} \quad r \leq -\frac{1}{2} + \sqrt{\frac{1}{4} + 4m}.$$

Having introduced the concept of quantum correlation matrices we can also define local determinist correlation matrices:

**Definition 3.5.5** (Local deterministic correlation matrix)**.** *A correlation matrix $c$ is called local deterministic of all components are $-1$ or $+1$.*

Let us now apply these concepts to the CHSH inequality. For the behavior scheme (2,2,2) we see that the rank of the correlation matrices is 2. Hence we can rewrite the CHSH inequality as

$$\sum_{i,j=0,1} M_{ij} c_{ij} \leq 1, \quad \text{for} \quad c_{ij} \in \{-1,1\} \tag{3.34}$$

where

$$M = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3.35}$$

Further we can rewrite Tsirelson's inequality

$$\sum_{i,j=0,1} M_{ij} c_{ij} \leq \sqrt{2}, \quad \text{for} \quad c_{ij} \in M_{\mathscr{Q}} \tag{3.36}$$

for the same matrix $M$. The quantum correlation matrices exceeding this bound are elements of the set of extremal quantum correlation matrices. Using Theorem 3.5.4 we can therefore further write

$$\sum_{i,j=0,1} M_{ij}\vec{x}_i \cdot \vec{y}_j \leq \sqrt{2}, \tag{3.37}$$

where the $\vec{x}_i$ and $\vec{y}_j$ are unit vectors in $\mathbb{R}^2$.

Let us now establish the connection to Grothendieck's constant which can be defined as follows using Grothendieck's inequality [37]:

**Definition 3.5.6** (Grothendieck's inequality, Grothendieck's constant of order n)**.** *Let n be an integer $n \geq 2$ and let M be any matrix $M \in \mathbb{R}^{m\times m}$ which satisfies*

$$\left|\sum_{i,j=0}^{m-1} M_{ij}x_i y_j\right| \leq 1 \tag{3.38}$$

*for $x_0,...,x_{m-1},y_0,...,y_{m-1} \in [-1,1]$. Then Grothendieck's constant of order n, denoted by $K_G(n)$, is defined as the smallest number satisfying the following inequality known as Grothendieck's inequality:*

$$\left|\sum_{i,j=0}^{m-1} M_{ij}\vec{x}_i \cdot \vec{y}_j\right| \leq K_G(n), \tag{3.39}$$

*for all unit vectors $\vec{x}_0,...,\vec{x}_{m-1},\vec{y}_0,...,\vec{y}_{m-1} \in \mathbb{R}^n$.*

**Definition 3.5.7** (Grothendieck's constant)**.** *Grothendieck's constant is defined as*

$$K_G = \lim_{n\to\infty} K_G(n). \tag{3.40}$$

To determine Grothendieck's constant is still a puzzling problem in mathematics. The only exact value known is $K_G(2) = \sqrt{2}$ [38]. Bounds are known for $K_G(3)$, $\sqrt{2} \leq K_G(3) \leq 1.5163$ and for Grothendieck's constant, $1.677 \leq K_G \leq \pi/(2\log(1+\sqrt{2})) = 1.7822$ [38, 39]. It is therefore very interesting to see that Grothendieck's inequality of second order,

$$\left|\sum_{i,j=0}^{m-1} M_{ij}\vec{x}_i \cdot \vec{y}_j\right| \leq K_G(2) = \sqrt{2}, \tag{3.41}$$

for

$$M = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{3.42}$$

and $\vec{x}_i$ and $\vec{y}_j$ being unit vectors in $\mathbb{R}^2$ is basically equivalent to Tsirelson's inequality as is clear from the considerations above. Investigations of different inequalities with correlation matrices of higher rank might therefore also lead to progress in the determination of Grothendieck's constant. For a recent contribution along this line the reader is referred to [40]. We might come back to this point in a future work.

# Chapter 4

# Optimizing nonlocality proofs

As we have seen in the introduction, certain nonlocality proofs (cf. Section 3.4) can be implemented as an experimental proof showing that the physical reality is incompatible with local realism. In this sense it is particularly interesting to know the strength of such experimental nonlocality proofs, i.e. how sure one becomes that local realistic theories are false, after observing a certain disagreement for those theories in a certain number of experiments. An optimal nonlocality proof would then be the one which gives the most confidence that local realism is false with the least number of experiments.

In Section 4.1 we introduce a measure for the statistical strength of nonlocality proofs in terms of relative entropy. In particular, we use this measure to investigate the statistical strength of nonlocality proofs for the $(2,2,d)$ Bell scenario for different values of $d$. In Section 4.2 we analyze, in comparison to the previous section, the maximal violation of the Bell inequalities associated to the $(2,2,d)$ Bell scenario which gives us the total variation distance of the classical and quantum probability distributions. Afterwards, we give a comparison of the both measures and point out some disadvantages of the latter one with respect to measure statistical strength (Section 4.3).

## 4.1 From naive Bayesian analysis to statistical strength

In this section we analyze the statistical strength of nonlocality proofs in terms of a two-player game. Let QM be a pro-quantum-mechanics experimentalist and LR a pro-local-realistic theoretician. QM is equipped with a specific proper nonlocality

proof, meaning a Bell-type inequality together with a choice of state and measurements that determine a quantum behavior which violates this inequality. QM will perform various experiments and observe a certain amount of experimental data. The goal of QM is then to test the hypothesis $\mathscr{Q}$, "the observed data is governed by a quantum behavior (cf. Definition 3.4.2)", against the composite hypothesis $\mathscr{L}$, "the observed data can be described by a local behavior (cf. Definition 3.2.1)" [41].

### 4.1.1 Naive Bayesian analysis

In this subsection we do a Bayesian analysis of the statistics arising from the above described game between QM and LR for a *fixed* setting [42]. This can bee seen as a post-experimental analysis, meaning that given the results of a number of experiments QM has performed, we derive a value for the confidence of QM against LR. However, this statistical analysis cannot be used to give a pre-eperimental advise to QM about which experimental setup might be the best to arrive at a certain confidence with the least number of experiments. An analysis of this kind will be subject of the forthcoming subsection.

In the following we want to introduce the basic concepts of Bayesian inference which we need for our discussing. In Section 2.1 we gave a basic introduction to probability measures and probability spaces. Another important quantity which have not discussed in this section is the so-called conditional probability:

**Definition 4.1.1** (Conditional probability)**.** *Let* $(\Omega, \mathscr{F}, \mathbb{P})$ *be a probability space and* $\Pr(B) > 0$ *for a* $B \in \mathscr{F}$*, then*

$$\Pr(\cdot|B) : \mathscr{F} \to [0,1], \quad A \mapsto \Pr(A|B) := \frac{\Pr(A \cap B)}{\Pr(B)}$$

*is called the conditional probability under B.*

Let us now introduce the concept of Bayesian inference. Bayesian inference is a method to assign a numerical estimate of the degree of belief in a hypothesis after certain evidence has been provided, in our case in form of an experiment, based on a numerical estimate prior to the given evidence.

**Definition 4.1.2.** *Let a* $H_0$ *denote a hypothesis, often also called null hypothesis. We call* $\Pr(H_0)$ *the prior probability of the null hypothesis* $H_0$*. For a given evidence E we call* $\Pr(H_0|E)$ *the posterior probability.*

The prior and posterior probabilities are related by Bayes' theorem:

**Theorem 4.1.3** (Bayes' theorem)**.**

$$\Pr(H_0|E) = \frac{\Pr(E|H_0)\Pr(H_0)}{\Pr(E)}.$$

**Proof.**

Bayes' theorem can be easily proven by inserting Definition 4.1.1 for $\Pr(H_0|E)$ and $\Pr(E|H_0)$.

■

Here $\Pr(E|H_0)$ is the probability of $E$ given $H_0$ is true and $\Pr(H_0) = \sum_i \Pr(E|H_i)\Pr(H_i)$ is the probability of $E$ under all mutually exclusive hypotheses.

In the following we want to use Bayesian analysis to test the hypothesis $\mathscr{Q}$, the observed evidence is generated by quantum mechanics, against the composite hypothesis $\mathscr{L}$, the evidence can be modeled by a local realistic theory. Suppose QM sets up an experiment which gathers some evidence in form of a sample of events $\mathscr{Z} = \{z_1, ..., z_n\}$. Using Bayes' theorem we can now relate the prior and posterior probabilities for both hypotheses [42]

$$\frac{\Pr(\mathscr{Q}|\mathscr{Z})}{\Pr(\mathscr{L}|\mathscr{Z})} = \frac{\Pr(\mathscr{Z}|\mathscr{Q})}{\Pr(\mathscr{Z}|\mathscr{L})}\frac{\Pr(\mathscr{Q})}{\Pr(\mathscr{L})}. \tag{4.1}$$

The confidence in this decision is given by the posterior odds $\mathrm{Post}(\mathscr{Q},\mathscr{L})$ of $\mathscr{Q}$ against $\mathscr{L}$, defined, for a given evidence $\mathscr{Z}$, as

$$\mathrm{Post}(\mathscr{Q},\mathscr{L}) := \frac{\Pr(\mathscr{Z}|\mathscr{Q})}{\Pr(\mathscr{Z}|\mathscr{L})}, \tag{4.2}$$

where large $\mathrm{Post}(\mathscr{Q},\mathscr{L})$ means larger confidence.

Let us now apply this Bayesian analysis to the specific example. Imaging QM and LR are setting up an experiment to test their hypotheses $\mathscr{Q}$ and $\mathscr{L}$. Due to the simple setting they choose the experiment discussed in the introduction. LR was so convinced by Einstein, Podolsky and Rosen's [3] notion of "elements of reality" that prior to the experiment he bets 100 Euros against 1 Euro that $\mathscr{L}$ is correct, yielding the following ratio of prior probabilities $\Pr(\mathscr{Q})/\Pr(\mathscr{L}) = 0.01$. However, QM as a true believer in quantum mechanics

and is willing to run as many experiments as necessary to convince his friend by $\Pr(\mathscr{Q}|\mathscr{Z})/\Pr(\mathscr{L}|\mathscr{Z}) = 100$. Thus, due to relation (4.1), he needs to provide a confidence of $\mathrm{Post}(\mathscr{Q},\mathscr{L}) = 10000$. We can now use the Bayesian analysis described above to determine how many experimental runs QM has to perform until he arrives at the desired confidence.

Recall the details of the experiment, QM prepares a singlet state of two spin-half particles and sends on particle to Alice and one to Bob. Alice can perform two spin measurements at an angle of 0 and $\pi/2$ and Bob at an angle of $\pi/4$ and $-\pi/4$. The corresponding CHSH inequality was derived in Theorem 3.3.1,

$$\Pr(X_1 \neq Y_1) \leq \Pr(X_0 \neq Y_0) + \Pr(X_0 \neq Y_1) + \Pr(X_1 \neq Y_0). \tag{4.3}$$

In the introduction we had shown that $\mathscr{Q}$ predicts all probabilities of the right-hand-side Equation (4.3) to be equal to $q$ and the probability on the left hand side to be equal to $1-q$, where $q = \sin^2(\pi/8) \approx 0.146$. From the rotational symmetry of the measuring devices one can see that an alternative description of the probabilities in Equation (4.3) from a local realistic model also yields that the probabilities on the right-hand-side are equal, let us call this common value $p$, and that the left-hand side is then $1-p$. For a local realistic model which mimics quantum mechanics as close as possible we have equality in (4.3), and hence we find $p = 1/4$.

We now assume that the probabilities for the experimentally found results, in terms of $m$ correct predictions in $n$ experiments are given by the a binomial distribution ,

$$\begin{aligned} \Pr(\mathscr{Z}|\mathscr{Q}) &= \frac{n!}{m!(n-m)!} q^m (1-q)^{n-m} \\ \Pr(\mathscr{Z}|\mathscr{L}) &= \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}. \end{aligned}$$

If we assume that $\mathscr{Q}$ is experimentally correct, i.e. $m \approx qn$, we arrive at the following expression for that posterior odds

$$\mathrm{Post}(\mathscr{Q},\mathscr{L}) = \left( \left(\frac{q}{p}\right)^q \left(\frac{1-q}{1-p}\right)^{1-q} \right)^n$$

Thus, to achieve the desired confidence of $\mathrm{Post}(\mathscr{Q},\mathscr{L}) = 10000$ we would need $n = 288$ experimental runs.

In this subsection we investigated the statistical strength of a certain nonlocality proof for the CHSH Bell setting. As we have seen the above calculation

depended strongly on the symmetries of the setting (which we used for example to determine the value of $p$). More generally one would like to formulate a notion of statistical strength which is not constraint to a particular setting nor does require any special symmetries. In the following subsection we want to introduce the relative entropy as such a measure for statistical strength.

### 4.1.2 Statistical strength and the Kullback-Leibler divergence

After the short discussion of a Bayesian analysis of nonlocality proofs as done in the previous subsection we now want to give a more sophisticated notion of statistical strength of nonlocality proofs. This is done in terms of the Kullback-Leibler divergence or relative entropy.

**Definition 4.1.4** (Kullback-Leibler divergence)**.** *Let $\mathscr{Z}$ be a finite set and $\{z\}$ be an event in $\mathscr{Z}$. For two probability distributions, $Q(z)$ and $L(z)$, over $\mathscr{Z}$ we define the Kullback-Leibler divergence from $Q$ to $L$ as*

$$H(Q||L) := \sum_{z\in\mathscr{Z}} Q(z)\log\frac{Q(z)}{L(z)},$$

*where* $\log(\cdot)$ *refers to the logarithm to the basis 2, as through out this text, and we adopted the convention* $0\log 0 = 0$*.*

The Kullback-Leibler divergence is also known as relative entropy or cross-entropy and was first introduced in [43]. A detailed discussion of its properties is given in Appendix A.1.2. In particular, it is shown that $H(Q||L) \geq 0$ with equality at $Q = L$. This is typical for the notion of a distance, however, in general we have $H(Q||L) \neq H(L||Q)$, and hence $H(Q||L)$ is formally not a distance. A further discussion of this asymmetry is presented in Section 4.3.

The Kullback-Leibler divergence was first used in [41] to quantify the statistical strength of nonlocality proofs.

In terms of simple hypothesis testing one can get an intuition of the relation between Kullback-Leibler divergence and statistical strength. Let $Z_1, Z_2, \ldots$ be a sequence of random variables independently generated by either $Q$ or $L$, with $Q \neq L$. Given some observed data $z_1, z_2, \ldots$ we can find out whether this data is generated by $Q$ or $L$ by comparing the likelihood of the data according to the two distributions. This is done by looking at the ratio

$$\frac{Q(z_1, z_2, \ldots)}{L(z_1, z_2, \ldots)} = \prod_{i=1}^{n} \frac{Q(z_i)}{L(z_i)}. \tag{4.4}$$

We now want to look at the behavior of the logarithm of this ratio assuming that $Q$ is true. From the Law of Large Numbers we know that if the $Z_i$ only take values in the finite set $\mathscr{Z}$ and $Q(z), L(z) > 0$ for all $z \in \mathscr{Z}$ then

$$\frac{1}{n}\sum_{i=1}^{n} P_i \to \mathbb{E}_Q(P) \quad \text{as } n \to \infty. \tag{4.5}$$

Here $P_i := \log(Q(Z_i)/L(Z_i))$, and

$$\begin{aligned} \mathbb{E}_Q(P) &= \mathbb{E}_Q(P_1) = \ldots = \mathbb{E}_Q(P_n) \\ &= \mathbb{E}_Q\left(\log\frac{Q}{L}\right) = \sum_{z\in\mathscr{Z}} Q(z)\log\frac{Q(z)}{L(z)} = H(Q||L). \end{aligned} \tag{4.6}$$

Hence,

$$\frac{1}{n}\log\frac{Q(Z_1,\ldots,Z_n)}{L(Z_1,\ldots,Z_n)} \to H(Q||L). \tag{4.7}$$

This means that that the average log-likelihood ratio between $L$ and $Q$, which can be seen as the amount of evidence for $Q$ against $L$, is asymptotically given by the Kullback-Leibler divergence.

It is also interesting to see that the Kullback-Leibler divergence has also a natural interpretation in terms of confidence, measured by the posterior odds of $Q$ against $L$. In Ref. [41] it was shown that

$$\mathrm{Post}(Q,L) = 2^{nH(Q||L)+\mathscr{O}(\log n)}, \quad \text{as } n \to \infty. \tag{4.8}$$

Thus the Bayesian confidence is measured by the Kullback-Leibler divergence up to first order in the exponent.

In this sense, in the previous subsection, we indirectly calculated an approximation for the Kullback-Leibler divergence for two probability distributions $Q = (q,q,q,1-q)$ and $L = (p,p,p,1-p)$. But who tells us that the distribution $L$ was the probability distribution in $\mathscr{L}$ which, in terms of Kullback-Leibler divergence, was the "closest" to any given $Q \in \mathscr{Q}$ or that the distribution $Q$ was the most far from this distribution $L$? The answer is simple nobody, it was just a good guess. We assumed that the element $L \in \mathscr{L}$ which yielded equality in the considered Bell inequality was the one closes to $\mathscr{Q}$ and that the element $Q$ which lead to the maximal violation of this inequality was the one most distant to $L$. However, as we will see in the following sections, this is simply not true in general. The nonlocality proof with the highest statistical strength (yielding to a bigger confidence after a fixed number of experimental runs) is not the one which leads to a maximal violation of the corresponding Bell inequality.

Now we see that to calculate the statistical strength of a given nonlocality proof (a fixed experimental setting) against the composite hypothesis $\mathscr{L}$ we should determine

$$\min_{L\in\mathscr{L}} H(Q||L), \tag{4.9}$$

or, if we leave the setting arbitrary,

$$\max_{Q\in\mathscr{Q}} \min_{L\in\mathscr{L}} H(Q||L). \tag{4.10}$$

The solution $Q^s \in \mathscr{Q}$ which maximizes $\min_{L\in\mathscr{L}} H(Q||L)$ then determines the optimal experimental setting for the given Bell setting $\mathscr{Q}$.

In [41] such optimizations of the above type were performed to numerically determine the statistical strength of several nonlocality proofs. Like this totally different Bell settings like CHSH and GHZ for example could be compared in terms of statistical strength. In the following we do not want to discuss all this results, but rather concentrate on the specific example of the CGLMP inequality and discuss the statistical strength for the different numbers of outcomes.

### 4.1.3 An application to the CGLMP inequality

Recall, that in the case of the $(2,2,d)$ Bell setting [27] we consider two parties, $\mathscr{A}$ and $\mathscr{B}$, each having a choice of two measurements for which each measurement can have $d$ possible outcomes.

We now want to discuss the statistical strength of such nonlocality proofs for different values of the number of outcomes [28]. In general the problem of determine the optimal setting for a given number of measurement outcomes, i.e. the optimal measurements and the optimal state, is numerically very difficult. To make the numerics more feasible we further want to fix the measurements to the following projective measurements $\Pi_{X,a}$ and $\Pi_{Y,b}$ with eigenvectors [44, 27],

$$|i\rangle_{X,a} = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} \exp\left(\mathbf{i}\frac{2\pi}{d}k(i+\alpha_a)\right)|k\rangle_{\mathscr{A}}, \tag{4.11}$$

$$|j\rangle_{Y,b} = \frac{1}{\sqrt{d}} \sum_{l=0}^{d-1} \exp\left(\mathbf{i}\frac{2\pi}{d}l(-j+\beta_b)\right)|l\rangle_{\mathscr{B}}, \tag{4.12}$$

where the phases read $\alpha_0 = 0$, $\alpha_1 = 1/2$, $\beta_0 = 1/4$ and $\beta_1 = -1/4$. These are the conjectured measurements which maximally violate the CGLMP inequality,

$$\Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(X_1 \geq Y_0) \geq 1, \tag{4.13}$$

for the case of the maximally entangled state

$$|\Phi\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_{\mathscr{A}} \otimes |i\rangle_{\mathscr{B}}, \tag{4.14}$$

as we will discuss in more detail in the following section. In [28] it was indeed shown that at least in the (numerical feasible) case of $d = 2,3,4$ these measurements also corresponded to the optimal measurements maximizing the Kullback-Leibler divergence. Conjecturing that this also holds for higher values of the number of outcomes the problem of finding the solution

$$H^s = H\left(Q^s||L^s\right) = \max_{Q\in\mathscr{Q}} \min_{L\in\mathscr{L}} H\left(Q||L\right). \tag{4.15}$$

becomes very similar to a standard Bell type inequality. In particular, if we consider small variations from this solution we find

$$\begin{aligned} H\left(Q^s + \delta Q||L^s\right) &\leq H^s & (4.16)\\ H\left(Q^s||L^s + \delta L\right) &\geq H^s. & (4.17) \end{aligned}$$

This implies that values $(Q,L)$ close to $(Q^s,L^s)$ must obey

$$\begin{aligned} \sum_i \log\left(\frac{Q_i^s}{L_i^s}\right) Q_i &\leq H^s & (4.18)\\ \sum_i \log\frac{Q_i^s}{L_i^s} L_i &\leq 1, & (4.19) \end{aligned}$$

where the index $i$ runs over the $4d^2$ components of the corresponding probability vector. Further, by convexity arguments one can see that the above requirement must be fulfilled by all probability distributions $Q$ and $L$.

Note that the left-hand-side of inequality (4.18) can be seen as the mean of a certain Bell operator and the left-hand-side of inequality (4.19) as a Bell type inequality. In particular, in [28] is was numerically checked that at least in the case of $d = 2,3,4$ inequality (4.19) is equivalent to the corresponding CGLMP inequality (4.13). To simplify the numerics this is further assumed also to hold for $d > 4$.

Now the numerical problem of finding the optimal state which maximizes the Kullback-Leibler divergence has hugely simplified and can be carried out for very large vales of $d$. The results from [28] are summarized in Figure 4.1. Shown is

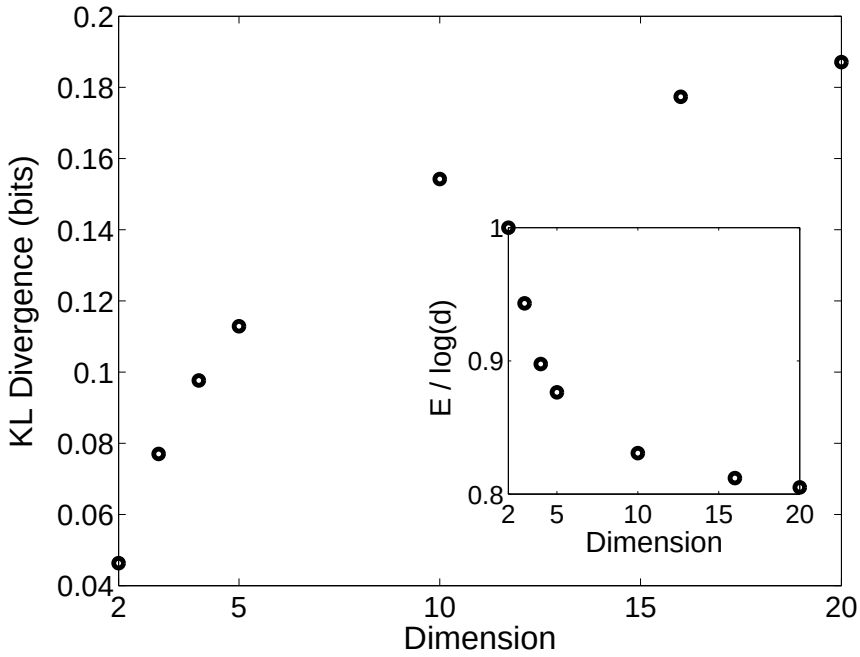


Figure 4.1: Plot of the maximized Kullback-Leibler divergence for Bell tests using the CGLMP setting with the described "best measurements" as a function of the dimension $d$. Inside: entanglement entropy $E/\log d$ of the corresponding optimal state as a function of $d$.

the numerically maximized Kullback-Leibler divergence $H^s(d)$ as a function of the number of measurement outcomes as well as the entanglement entropy of the corresponding optimal state.

It is interesting to see that only the optimal state for $d = 2$ is maximally entangled and all others for $d > 2$ are not. In particular, it was shown that for large values of $d$ the entanglement entropy behaves like $\lim_{d\to\infty} E(\psi) = \ln d \approx 0.69 \log d$. We will see in the forthcoming section that qualitatively this is also true for the optimal state maximally violating the CGLMP inequality. However, interestingly the optimal state in terms of the statistical strength of the corresponding Bell experiment is not the same as the one maximally violating the corresponding Bell type inequality, as was long believed.

The fact that the optimal state both in the sense of Kullback-Leibler divergence as in the sense of maximal violation of Bell type inequalities is not maximally

entangled for the CGLMP setting with $d > 2$ was already observed in [28] and [45]. Nevertheless, this fact is still puzzling, since violation of Bell inequalities and entanglement are both seen as measures of "quantum nonclassicality". In the forthcoming section we will further investigate this differences.

## 4.2 On the maximal violation of the CGLMP inequality

In the previous section we introduced the Kullback-Leibler divergence as a statistical strength of non-locality proofs. In this section we also want to concentrate on nonlocality proofs that use the CGLMP inequality. Whereas in the previous section we determined the optimal setting via maximizing the Kullback-Leibler divergence, in this section want to determine the optimal setting by maximizing the violation of the CGLMP inequality over the set of quantum behaviors $\mathscr{Q}(2,2,d)$. This is equivalent to maximize the so-called total variation distance [46]

**Definition 4.2.1** (Total variation distance)**.** *Let $\mathscr{Z}$ be a finite set and $\{z\}$ be an event in $\mathscr{Z}$. For two probability distributions, $Q(z)$ and $L(z)$, over $\mathscr{Z}$ we define the total variation distance between Q and L as*

$$D(Q||L) = \sum_{z\in\mathscr{Z}} |Q(z) - L(z)| .$$

It is easy to see that $D(P||Q)$ defines a proper measure of distance (in contrast to the Kullback-Leibler divergence for which in general $H(Q||P) \neq H(P||Q)$).

As said above in this section we want to determine the optimal setting which maximizes the violation of the CGLMP inequality, i.e. find the solutions

$$D^s = D(Q^s||L^s) = \max_{Q\in\mathscr{Q}(2,2,d)} \min_{L\in\mathscr{L}(2,2,d)} D(Q||L). \tag{4.20}$$

As already discussed in previous chapter, since $\mathscr{Q}(2,2,d)$ is a convex and since the CGLMP inequality is linear, it reaches its maximum on the boundary. Hence, knowing the maximal violation would give us also further inside in the geometrical properties of the set $\mathscr{Q}(2,2,d)$.

For the case of the CHSH inequality we have derived a bound over the set of quantum behaviors in terms of Tsirelson's inequality. Thereby we proved that its maximal violation corresponds to the value $2\sqrt{2}$. In the introduction we have

shown a specific choice of measurements that for the maximally entangled state exhibited this maximal value.

At this point one might ask several questions regarding the more general case of the $(2,2,d)$ Bell setting. First, what is the maximal violation of the CGLMP inequality and can one prove the existence of an exact bound? Second, what are the best measurements and the best state that produce the maximal violation? Is the best state maximally entangled?

In the following we want to investigate this problem and provide answers to the above questions [47].

### 4.2.1 The maximally entangled state

Before considering the case of arbitrary quantum behaviors let us first concentrate on the maximal violation of the CGLMP inequality,

$$\Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(X_1 \geq Y_0) \geq 1, \tag{4.21}$$

for the case of the maximally entangled state.

For the maximally entangled state,

$$|\Phi\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_{\mathscr{A}} \otimes |i\rangle_{\mathscr{B}} \tag{4.22}$$

it has long been conjectured that the measurements which maximally violate the CGLMP inequality are described by the projective measurements $\Pi_{X,a}$ and $\Pi_{Y,b}$ with the following eigenvectors [44, 27],

$$|i\rangle_{X,a} = \frac{1}{\sqrt{d}} \sum_{k=0}^{d-1} \exp\left(\mathbf{i}\frac{2\pi}{d} k(i+\alpha_a)\right) |k\rangle_{\mathscr{A}}, \tag{4.23}$$

$$|j\rangle_{Y,b} = \frac{1}{\sqrt{d}} \sum_{l=0}^{d-1} \exp\left(\mathbf{i}\frac{2\pi}{d} l(-j+\beta_b)\right) |l\rangle_{\mathscr{B}}, \tag{4.24}$$

where the phases read $\alpha_0 = 0$, $\alpha_1 = 1/2$, $\beta_0 = 1/4$ and $\beta_1 = -1/4$.

We evaluate the left-hand-side of inequality (4.21) for the joint probabilities arising from a quantum behavior in the case of the maximally entangled state and the just described measurements. From Definition 3.4.2 we get,

$$\Pr(X_a < Y_b) = \sum_{i<j} \mathrm{Tr}\left(\Pi^i_{X,a} \otimes \Pi^j_{Y,b} |\Phi\rangle\langle\Phi|\right), \tag{4.25}$$

where the $\Pi^i_{X,a} = |i\rangle_{X,a}\langle i|_{X,a}$ are the projectors on the corresponding eigenspaces defined in (4.23)–(4.24) and similarly for $\Pi^j_{Y,b}$. For later purposes we will leave the Schmidt coefficients unspecified throughout this calculation and only equate them to $1/\sqrt{d}$ at the end. Hence considering the state

$$|\psi\rangle = \sum_{i=0}^{d-1} \lambda_i |i\rangle_{\mathscr{A}} \otimes |j\rangle_{\mathscr{B}}, \tag{4.26}$$

one obtains for the joint probability for the measurement outcome being $X_a = k$ and $Y_b = l$

$$\begin{aligned} P(X_a = k, Y_b = l) &= \left|\langle\psi|k\rangle_{X,a} \otimes |l\rangle_{Y,b}\right|^2 \\ &= \frac{1}{d^3}\left|\sum_{i=0}^{d-1} \exp\left(\mathbf{i}\frac{2\pi}{d} i(k-l+\alpha_a+\beta_b)\right)\right|^2, \end{aligned} \tag{4.27}$$

where we used that $\langle j|j'\rangle = \delta_{j,j'}$. Using the values for the phases we get

$$\begin{aligned} \mathscr{A}_d(\psi) &\equiv \Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(X_1 \geq Y_0) \\ &= \sum_{i=0}^{d-1}\sum_{j=0}^{d-1} M_{ij}\lambda_i\lambda_j, \end{aligned} \tag{4.28}$$

where the $d \times d$-matrix $M$ can be simplified to

$$M_{ij} = 2\,\delta_{ij} - \frac{1}{d}\cos^{-1}\left(\frac{(i-j)\pi}{2d}\right). \tag{4.29}$$

Putting $\lambda_i = 1/\sqrt{d}$, i.e., looking at the maximally entangled state, we obtain for $d = 2$,

$$\mathscr{A}_2(\Phi) = \frac{3-\sqrt{2}}{2} \approx 0.79289 \tag{4.30}$$

which corresponds to the maximal violation of the CHSH inequality know from Csirelson's inequality [36].

It is also interesting to look at the conjectured (it is not known that these are the best measurements) maximal violation of Equation (4.21) with the infinite dimensional maximally entangled state. We get

$$\begin{aligned} \lim_{d\to\infty} \mathscr{A}_d(\Phi) &= 2 - \frac{1}{L^2}\int_0^L\int_0^L \cos^{-1}\left(\frac{\pi}{2L}(x-y)\right) dxdy \\ &= 2 - \frac{16Cat^2}{\pi^2} \approx 0.515 \end{aligned} \tag{4.31}$$

where Cat is Catalan's constant, reproducing the result obtained in [27] for the original version of the CGLMP inequality.

In this section we described what are believed to be the best measurements for the CGLMP inequality with the maximally entangled state. Though it is often thought that the maximally entangled state $|\Phi\rangle$ represents the most nonlocal quantum state, evidence has been given in [45] and [28] that the states which maximally violate inequality (4.21) are not maximally entangled. In the following section we provide further evidence for this and investigate several properties of the optimal state especially in the case of large values of $d$.

### 4.2.2 Optimizing over states and measurements

In the previous section we described the measurements which in the case of the maximally entangled state appear to give the maximal violation of inequality (4.21). However, as mentioned above, it has already been seen that in the case of $d \geq 3$ the state which causes the maximum violation of the inequality is actually not the maximally entangled state [45, 28].

Natural questions which arise at this point are: How can the optimal state be described for larger $d$ and are the corresponding best measurements the same as in the case of the maximally entangled state? Further, what is the maximal violation of inequality (4.21) as $d$ tends to infinity?

To address the above questions we want to optimize the left-hand-side of inequality (4.21) over all possible measurements and states. For this purpose we assume that the state of Alice's and Bob's composite system is a pure state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ and that the measurements $\Pi_{X,a}$ and $\Pi_{Y,b}$ describing Alice's and Bob's measurement are projective and nondegenerate as also considered above. In this case the problem of finding the minimal value of the left-hand-side of inequality (4.21) becomes a nonlinear $(4d^2+d)$-dimensional optimization problem with respect to certain constraints.

For small values of $d$ we can numerically perform the optimization. The results for the first values are summarized in Table 4.1. Shown are the minimal values of the left-hand-side of inequality (4.21), denoted by $\min \mathscr{A}_d(\psi, \Pi_{X,a}, \Pi_{Y,b})$, and the Schmidt coefficients of the optimal state for which $\mathscr{A}_d(\psi, \Pi_{X,a}, \Pi_{Y,b})$ reaches its minimum.

One observes that for $d \geq 3$ the optimal state is not maximally entangled. More precisely, as we will see later the entanglement entropy decreases as $d$ becomes bigger. The optimal states arising from the numerical optimization over $\mathscr{A}_d(\psi, \Pi_{X,a}, \Pi_{Y,b})$ agree with results obtained in [45], but differ from the results

discussed in the in Section 4.1 (see also [28]). That is because in [28] the quantity to be optimized was not the CGLMP inequality, but the Kullback-Leibler divergence (relative entropy) which quantifies the average amount of support in favour of QM against LR per trial.

Closer analysis of the optimal measurements $\Pi_{X,a}$ and $\Pi_{Y,b}$ shows that even though the optimal state is not the maximally entangled state the best measurements seem to be the best measurements (4.23) and (4.24) of the previous case. Further numerical optimizations for higher values of $d$ give strong evidence that this true in general.

### 4.2.3 On a quantum Bell inequality for infinite dimensional states

If we assume that (4.23) and (4.24) are the best measurements for all values of $d$ we can further simplify the optimization. We have already derived in Equation (4.28) that in the case of the measurements (4.23) and (4.24) we can write

$$\begin{aligned} \mathscr{A}_d(\psi) &\equiv \Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(X_1 \geq Y_0) \\ &= \sum_{i=0}^{d-1}\sum_{j=0}^{d-1} M_{ij}\lambda_i\lambda_j, \end{aligned} \tag{4.32}$$

where $|\psi\rangle = \sum_{i=0}^{d-1} \lambda_i |i\rangle_{\mathscr{A}} \otimes |i\rangle_{\mathscr{B}}$ and the $d \times d$-matrix $M$ was given by

$$M_{j,j'} = 2\,\delta_{j,j'} - d^{-1}\cos^{-1}\left(\frac{(j-j')\pi}{2d}\right). \tag{4.33}$$

Hence under this assumption, finding the maximal violation of (4.21) reduces to finding the smallest eigenvalue of the matrix $M$. The corresponding eigenvector $\{\lambda_i\}_{i=0}^{d-1}$ gives us the optimal state.

Before going on with the search for the smallest eigenvalue of $M$ let us first proof some general properties of all eigenvalues.

Note that the matrix has the following form

$$M = \begin{pmatrix} m_0 & m_1 & m_2 & \cdots & m_{d-1} \\ m_1 & m_0 & m_1 & \cdots & m_{d-2} \\ m_2 & m_1 & m_0 & \cdots & m_{d-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{d-1} & m_{d-2} & m_{d-3} & \cdots & m_0 \end{pmatrix} \tag{4.34}$$

with $m_0 > 0 > m_1 > m_2 > \cdots > m_{d-1}$.

Table 4.1: Optimization over all observables and states. Only the minimal value of the left-hand-side of inequality (4.21) and the Schmidt coefficients are shown.

| $d$ | $\min A_d(\psi, \Pi_{X,a}, \Pi_{Y,b})$ | $\lambda_0$ | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ | $\lambda_6$ | $\lambda_7$ |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.792893 | 0.707107 | 0.707107 | - | - | - | - | - | - |
| 3 | 0.695048 | 0.616895 | 0.488753 | 0.616895 | - | - | - | - | - |
| 4 | 0.635238 | 0.568573 | 0.420394 | 0.420381 | 0.568572 | - | - | - | - |
| 5 | 0.593716 | 0.536835 | 0.385935 | 0.385908 | 0.385908 | 0.536842 | - | - | - |
| 6 | 0.562626 | 0.513699 | 0.364399 | 0.321401 | 0.321417 | 0.364453 | 0.513706 | - | - |
| 7 | 0.538159 | 0.495734 | 0.349317 | 0.301086 | 0.288313 | 0.301076 | 0.349346 | 0.495742 | - |
| 8 | 0.518208 | 0.481196 | 0.337879 | 0.286858 | 0.268132 | 0.267839 | 0.287501 | 0.337887 | 0.481191 |

**Theorem 4.2.2.** *The eigenvectors* $\vec{\lambda}^T = (\lambda_0, \lambda_1, ..., \lambda_{d-1})^T$ *of the matrix M with positive components* $\lambda_i > 0$, $\forall i \in \{0, ..., d-1\}$ *have the following symmetry for* $d \geq 3$*:* $\lambda_i = \lambda_{d-1-i}$.

**Proof.**

Let use only prove the Theorem for even $d$. The proof for odd values of $d$ is similar. For even $d$ we can write the matrix $M$ in terms of four $d/2 \times d/2$ matrices,

$$M = \begin{pmatrix} M_1 & M_2 \\ \tilde{I}^T M_2 \tilde{I} & M_1 \end{pmatrix}, \tag{4.35}$$

where we have $\tilde{I}_{i,j} = \delta_{i,d/2-1-j}$, $i, j = 0, ..., d/2-1$. $\tilde{I}$ and $M_1$ have the following poperties $\tilde{I}^T = \tilde{I}$ and $M_1^T = M_1$. If we write $\vec{\lambda}^T = \{\vec{\lambda}_1^T, \vec{\lambda}_2^T\}$, where both $\vec{\lambda}_1$ and $\vec{\lambda}_2$ have $d/2$ components, the eigenvalue problem reads

$$\begin{aligned} M_1\vec{\lambda}_1 + M_2\vec{\lambda}_2 &= \lambda\vec{\lambda}_1 \\ (\tilde{I}^T M_2 \tilde{I})\vec{\lambda}_1 + M_1\vec{\lambda}_2 &= \lambda\vec{\lambda}_2. \end{aligned} \tag{4.36}$$

Using the conditions stated above one gets

$$\vec{\lambda}_2^T M_2 \vec{\lambda}_2 = \vec{\lambda}_1^T (\tilde{I}^T M_2 \tilde{I})\vec{\lambda}_1 \tag{4.37}$$

and hence $\vec{\lambda}_2 = \pm\tilde{I}\vec{\lambda}_1$. But since all components of $\vec{\lambda}$ are positive we get $\vec{\lambda}_2 = \tilde{I}\vec{\lambda}_1$ which completes the proof.

■

For small values of $d$ we can determine the eigenvectors of $M$ analytically and hence also obtain $\min \mathscr{A}_d = \vec{\lambda}^T M \vec{\lambda}$, where $\vec{\lambda}$ is the eigenvector with the smallest eigenvalue. In particular, for $d = 2, 3$ we obtain $\min \mathscr{A}_2 = (3 - \sqrt{2})/2$, with $\vec{\lambda} = (1, 1)^T/\sqrt{2}$, and $\min \mathscr{A}_3 = (12 - \sqrt{33})/9$, with $\vec{\lambda} = (1, \gamma, 1)^T/(\sqrt{2 + \gamma^2})$, where $\gamma = (\sqrt{11} - \sqrt{3})/2$, agreeing with results presented in [45].

More interesting becomes the search for eigenvectors with minimal eigenvalue for a large number of possible measurement outcomes. Numerical search for those eigensystems is feasible for very large values of $d$ by use of Arnoldi iteration.

The results of the numerical optimizations are summarized in Figure 4.2. Shown is the minimal target value, $\mathscr{A}_d(\psi)$, as a function of the dimension $d$ for a range from 2 to $10^6$ both for the case of the maximally entangled state and the

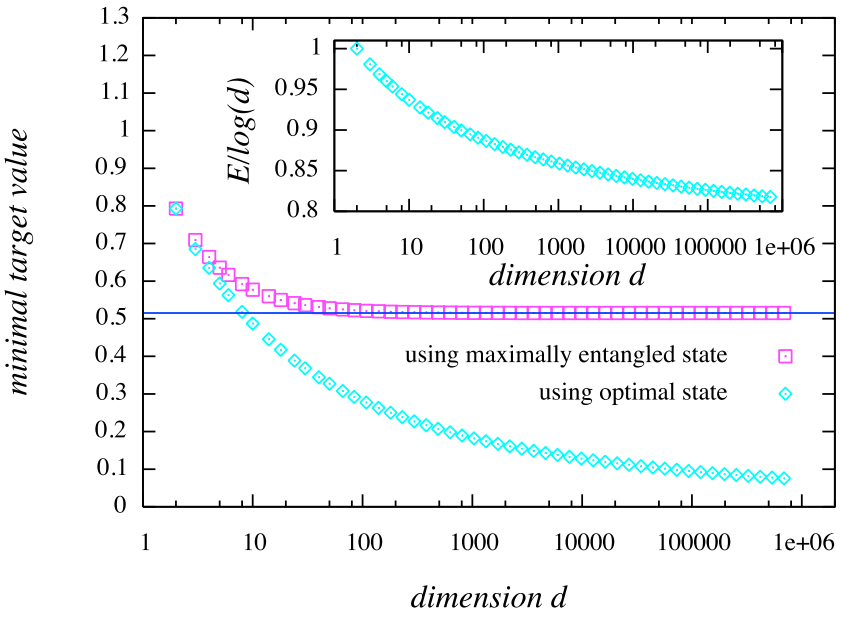


Figure 4.2: Minimal value of the left-hand-side of inequality (4.21) as a function of the dimension $d$: (i) for the maximally entangled state and (ii) for the optimal state. Inside: Entanglement entropy $E/\log d$ of the optimal state as a function of the dimension $d$.

optimal state. In the case of the maximally entangled state, $\mathscr{A}_d(\Phi)$ approaches very quickly the asymptotic value $\mathscr{A}_\infty(\Phi) \approx 0.515$ derived above.

In the case of the optimal state it is interesting that the maximal violation of (4.21) does not approach an asymptote very quickly. In fact, for very large $d$ it falls off slower than logarithmically with the dimension. The numerical data shown in Figure 4.2 do suggest that the minimal value of $\mathscr{A}_d(\psi)$ approaches zero as $d$ tends to infinity. This is very interesting since zero is the absolute minimum of $\mathscr{A}_d(\psi)$ on the boundary of the polytope of all probability vectors. If one could show analytically that there exists a optimal state which actually causes $\mathscr{A}_d(\psi)$ to approach zero as $d$ tends to infinity, one would have proven a new tight quantum Bell inequality for the $(2,2,\infty)$ scenario (see conjecture at the end of this section).

Let us now investigate further properties of the optimal states causing the maximal violation of inequality (4.21). Figure 4.3 shows the typical shape of a optimal state for $d \geq 3$, namely in the case of $d = 10000$. Plotted are the Schmidt coeffi-

cients $\lambda_i$ as a function of the index $i$. The reflection symmetry around $(d-1)/2$ is the one proven in Theorem 4.2.2 as a general property of all eigenvectors of $M$. As $d$ increases the Schmidt coefficient get more and more peaked at $i=0$ and $i=d-1$.

It is also interesting to look at the entanglement entropy of the optimal state. Whereas for the maximally entangled state $E(\Phi)/\log d = 1$ for all values of $d$, in the case of the optimal state the entanglement entropy decreases with the dimension. As in the case of the minimal value of $\mathscr{A}_d(\psi)$ the entanglement entropy decreases slower than logarithmically, but we are not able to give an asymptotic bound for it. This is contrary to work of [28] presented in the previous section, where the entanglement entropy seemed to approach the asymptotic value $\lim_{d\to\infty} E(\psi) = \ln d \approx 0.69 \log d$. Again, the disagreement is due to the fact that in the latter the quantity to be optimized was not the CGLMP inequality, but rather the Kullback-Leibler divergence.

From the insights gained in this section we state the following conjecture:

**Conjecture 1** (Quantum Bell inequality)**.** *For $d \to \infty$ the minimal value of* $\Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(X_1 \geq Y_0)$ *converges to zero, where the best measurements for each d are the ones presented above,* (4.23) *and* (4.24)*, and the optimal states are of the form shown in Fig. 4.3. Hence,*

$$\Pr(X_1 < Y_1) + \Pr(Y_1 < X_0) + \Pr(X_0 < Y_0) + \Pr(X_1 \geq Y_0) \geq 0, \tag{4.38}$$

*is a tight quantum Bell inequality for the* $(2,2,\infty)$ *Bell setting.*

A proof of this conjecture is work in progress and will be hopefully reported soon.

### 4.2.4 Approximating the optimal state

Under the assumption that the conjectured best measurements are optimal the problem of proving the above conjecture reduces to find a analytical expression for the optimal state for which $\mathscr{A}_d(\psi)$ should become zero as $d$ tends to infinity. In this section we try to model the optimal state by various approximations and looks at the value of $\mathscr{A}_d(\psi)$ for this approximate states.

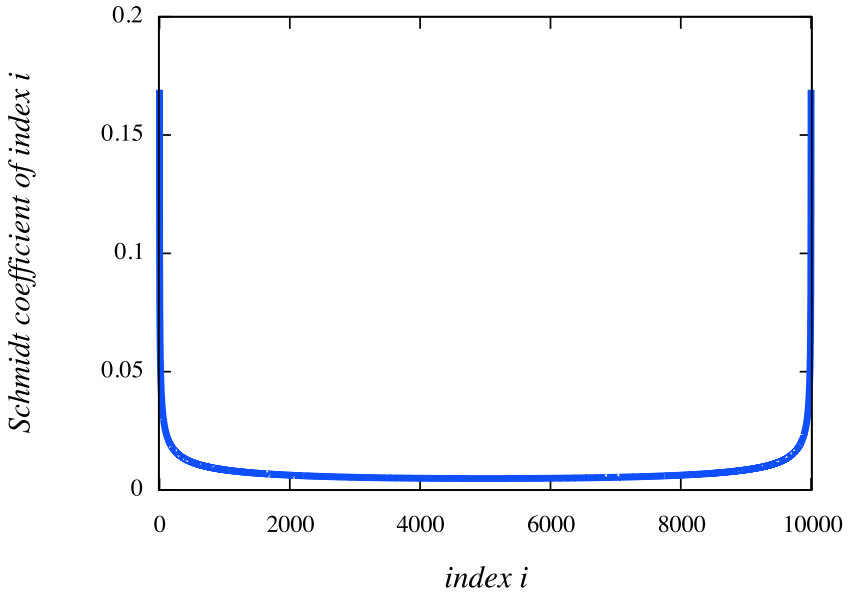


Figure 4.3: The typical shape of a optimal state for $d \geq 3$. Shown are the Schmidt coefficients $\lambda_i$ of the optimal state for $d = 10000$ as a function of the index $i$.

### The NOPA state

To model the infinite dimensional maximally entangled state the following, so-called NOPA state has been proposed in the literature [48]

$$|\psi\rangle_{NOPA} = \sqrt{1-\lambda^2}\sum_{i=0}^{\infty}\lambda^i|i\rangle_{\mathscr{A}}\otimes|i\rangle_{\mathscr{B}}, \tag{4.39}$$

for $\lambda \in (0,1)$. In particular, for $\lambda \to 1$ the NOPA state becomes maximally entangled. The NOPA state is also very interesting, since there are ways of experimentally realizing this state for certain values of $\lambda$ [49].

Since the optimal state we want to approximate has symmetrical Schmidt coefficients, we try to model it with a superposition of two NOPA states of which one has reflected Schmidt coefficients, i.e.

$$|\psi_1(\lambda)\rangle \sim \sum_{j=0}^{d-1}\left(\lambda^j + \lambda^{d-1-j}\right). \tag{4.40}$$

which for small values of $\lambda$ qualitatively looks like the conjectured optimal state.

In particular, it is possible to perform the limit $d \to \infty$ and turn the summations into integrals yielding

$$\begin{aligned} \mathscr{A}_\infty(\psi_1(\lambda)) &= 2 - \frac{\log\lambda}{2\lambda\log\lambda + \lambda^2 - 1}\int_0^1\int_0^1 \cos^{-1}\left(\frac{\pi}{2}(x-y)\right)\times \\ &, \quad \times\left(\lambda^{x+y} + \lambda^{x-y-1} + \lambda^{y-x-1} + \lambda^{2-x-y}\right)dxdy \end{aligned} \tag{4.41}$$

The result can be numerically integrated and further minimized over all vales of $\lambda$. One obtains that $\lambda^m = 3.6510^{-6}$ corresponds to the minimal value $\mathscr{A}_\infty(\psi_1(\lambda^m)) = 0.352$. It is also easy to check that for $\lambda = 1$ one recovers the result of the maximally entangled state $\mathscr{A}_\infty(\psi_1(1)) \approx 0.515$. Clearly, $\psi_1$ is not a very good approximation for the optimal state, however, it is still much better than the maximally entangled state and due to the possibility of modeling the NOPA state in the laboratory the above described approximation might still be a good candidate to perform actual Bell experiments.

A slight modification of $\psi_1(\lambda)$ is the state

$$|\psi_2(\lambda,a)\rangle \sim \sum_{j=0}^{d-1}\left(\lambda^j + \lambda^{d-1-j} + a\right), \tag{4.42}$$

i.e. the superposition of $\psi_1(\lambda)$ with a maximally entangled state. A similar analysis as in the above example shows that the $\lambda^m = 5.4810^{-17}$ and $a^m = 0.672$ minimize $\mathscr{A}_\infty(\psi_2(\lambda^m, a^m)) = 0.2306$.

In the following we want to discuss one further more accurate approximation to the optimal state.

### A better approximation

In this subsection we want to discuss a different state as an approximation for the optimal state maximally violating the CGLMP inequality (in the case of the conjectured best measurements). Consider the following state which was first given in [50] as a good approximation to the optimal state:

$$|\psi_3\rangle \sim \sum_{i=0}^{d-1}\frac{1}{\sqrt{(1+i)(d-i)}}|i\rangle_{\mathscr{A}}\otimes|i\rangle_{\mathscr{B}} \tag{4.43}$$

Qualitatively, this state has the shape as shown in Figure 4.3 for the optimal state and therefore serves as a good candidate.

Ideally, we would be interested in knowing the value of $\mathscr{A}_d(\psi_3)$ as $d$ tends to infinity. It is straightforward to turn the summations in this expression into integral in the limit as $d$ goes to infinity, however, due to the poles in the integrals the normalization factor becomes zero and the integral becomes infinite. To deal with the problem one has to introduce a regulator for the divergences which one wants to take to zero again after performing the integration. One possible choice of regulator would be

$$\mathscr{A}_\infty(\psi_3) = 2 - \lim_{\varepsilon\to 0} \frac{1}{\mathscr{N}_\infty(\varepsilon)} \int_0^1 \int_0^1 \frac{\cos^{-1}\left(\frac{\pi}{2}(x-y)\right) dx\, dy}{\sqrt{(x+\varepsilon)(1-x-\varepsilon)(y+\varepsilon)(1-y-\varepsilon)}} \quad (4.44)$$

with

$$\mathscr{N}_\infty(\varepsilon) = \log\left(\frac{\varepsilon^2-1}{\varepsilon^2}\right). \quad (4.45)$$

However, it turns out that in this case the integration becomes analytically and, due to the problems with the regulator, also numerically non-feasable.

The only thing left we can do is numerically evaluate the sums in $\mathscr{A}_d(\psi_3)$ for various values for $d$. The result is shown in Figure 4.4 in comparison to the numerical values for the optimal state. In addition, Figure 4.5 displays the entanglement entropy of both, the approximate and the optimal state.

One can see that whereas for small vales of $d$ the minimal value of the left-hand-side of inequality (4.21) for both the approximate state and the optimal state are very close, for bigger values of $d$ the deviation becomes much bigger. For approximately $d = 500000$ the minimal target value for the approximate state is already more than 25% above the minimal target value for the optimal state. This also suggest that even if we would be able to evaluate $\mathscr{A}_\infty(\psi_3)$ exactly it would probably not tell us too much about the corresponding value for the optimal state. In particular, where has the former might give a non-zero result the latter might still be zero as conjectured above.

The difference between the approximate state $\psi_3$ and the optimal state is nicely observed in Figure 4.5 in which the entanglement entropy $E/\log d$ as a function of $d$ is displayed for both the approximate as well as the optimal state. One sees that for large values of $d$ the approximate state is much more entangled than the optimal state.

In this subsection we analyst different approximate states of the optimal state. The NOPA state might be particularly interesting for actual experimental realizations of the CGLMP Bell experiment in the case of infinitely many dimensions. The other approximation was very accurate for small values of $d$. However, none

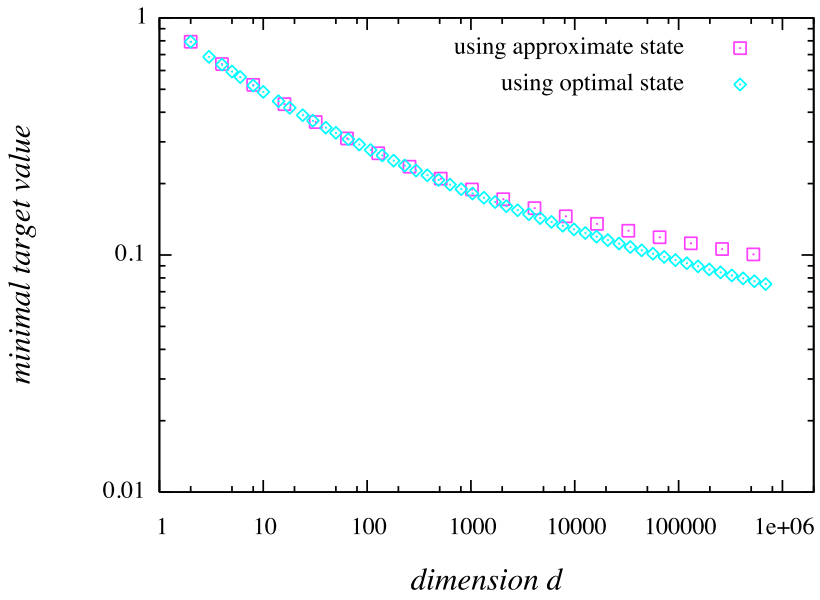


Figure 4.4: Minimal value of the left-hand-side of inequality (4.21) as a function of the dimension $d$: (i) for the approximate state Equation (4.43) and (ii) for the optimal state.

of the considered approximations was very accurate for the case of large values of $d$.

## 4.3 Kullback-Leibler divergence versus total variation distance

In this chapter we introduced two divergences for probability distributions, namely the Kullback-Leibler divergence and the total variational distance. Whereas the Kullback-Leibler divergence was shown to give a proper measure for statistical strength in the context of nonlocality proofs, the total variational distance was more useful to study certain geometrical properties of the convex body of quantum behaviors, e.g. finding a generalized Tsirelson inequality.

In particular, this means that if one is performing some actual Bell experiment and one is interested in showing most efficiently that the outcomes of the experiment cannot be obtained by local realistic models the settings one should

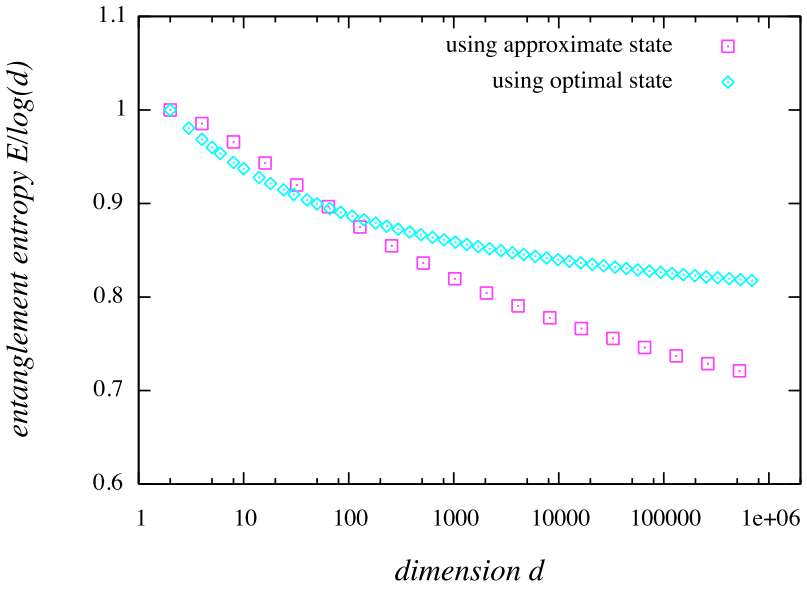


Figure 4.5: Entanglement entropy $E/\log d$ as a function of $d$: (i) for the approximate state Equation (4.43) and (ii) for the optimal state.

use are the optimal settings derived from maximizing the Kullback-Leibler divergence and *not* the optimal settings derived from maximizing the total variation distance, i.e. from finding maximal violations of Bell inequalities. This fact is sometimes stated wrong in the literature, where some authors equate statistical strength with total variation distance. We therefore want to point out explicitly that Kullback-Leibler divergence defines a proper measure of statistical strength, but not the total variation distance. The latter is mainly interesting to study the boundary of the convex body of quantum behaviors.

To make the misuse of total variation distance as statistical strength clearer let us consider the following simple example.

**Example 4.3.1.** Consider a set of two possible outcomes, say the events spin up and spin down, $\mathscr{X} = \{\uparrow, \downarrow\}$. Image now we would have the following probability distributions $L(z)$ and $Q(z)$:

$$L(\uparrow) = 0.99, \quad L(\downarrow) = 0.01 \quad \textit{and} \quad Q(\uparrow) = 1, \quad Q(\downarrow) = 0.$$

The total variation distance is just $D(L||Q) = 0.02$. Further it does not depend on whether we would test $L(z)$ against $Q(z)$ or the other way around. Consider

for example the case that the observed data would be generated from $L(z)$, then after a finite amount of experimental runs we would observe one event $\downarrow$ and we would be 100% sure that the data was sampled from $L(z)$ and $Q(z)$. This can be seen from the Kullback-Leibler divergence. If we take "the data is generated from $L(z)$" as our hypothesis we would naturally calculate $H(L||Q) = \infty$ which is infinite in our example, reflecting the fact that after a certain amount of runs we would have infinite confidence that the hypothesis was correct. If we one the other hand would assume that the data is generated from $Q(z)$ we would rather calculate $H(Q||L) = 0.015$, meaning that if the data was really sampled by $Q(z)$ we would need many experimental runs to gain a certain amount of confidence.

In this section we commented on the differences between Kullback-Leibler divergence and total variation distance. The main massage one should take from this discussion is that, against common belief, the optimal setting one should use to perform a Bell experiment most efficiently, i.e. gaining the highest confidence that the data was generated from quantum mechanics rather than from a local realistic model, is not the one which maximally violates the corresponding Bell inequality, but rather the one which maximizes the Kullback-Leibler divergence.

# Chapter 5

# Conclusion

Despite many efforts there are still many mysteries surrounding the nonclassical aspects of quantum probability distributions. The mayor task of this thesis was to introduce the reader to this mysteries and to further investigate some of those. This was done by analyzing the relations between several notions or measures of nonclassicallity often used in the literature, namely entanglement (entropy), maximal violation of Bell inequalities (total variance distance) and statistical strength of nonlocality proofs (Kullback-Leibler divergence).

In Chapter 2 we introduced the reader to the mathematical framework of quantum statistics. We will briefly introduced the basic notion of classical probability theory and defined the concepts of quantum probability theory in terms of Hilbert spaces and measurements on them. Further, we will generalized the setup for tensor products of several Hilbert spaces. Finally, we showed how this mathematical framework of quantum probability theory is used to model the physical theory of quantum mechanics and clarified the basic concepts by applying them to a simple example.

In Chapter 3 we discussed the aspect of non-locality in quantum statistics. In the first part of this chapter we introduced the general framework of Bell scenarios and how to describe joint probability distributions on those. Further, we will derived several geometrical properties of the set of all such joint probability distributions. We will characterized under which conditions such behaviors are said to be local or locally deterministic. The notion of locality was used to derive the CHSH and CGLMP inequality. In the second part we specified the set of joint probabilities which arises from the theory of quantum probability. We showed that in general such quantum behaviors are nonlocal and can be used to define a nonlocality proof. At the end of the chapter an inequality for quantum behav-

iors of the CHSH setting was given and interesting relations to Grothendieck's inequality and Grothendieck's constant were presented.

Chapter 4 was the main new part of this thesis. It discusses the optimization of nonlocality proofs in particular of the CGLMP setting; first with respect to statistical strength defined in terms of Kullback-Leibler divergence and second in terms of total variation distance, i.e. maximal violation of Bell inequalities. It is shown that whereas the optimal measurements are the same in both cases, namely the (conjectured) best measurements in the case of the maximally entangled state for violation of the CGLMP inequality, the optimal states are not. In particular, the optimal states in both cases are not the maximally entangled state.

We investigated the maximal violation of the CGLMP inequality for very large numbers of measurement outcomes. We analysed the specific form of the best states and their entanglement entropy. It turned out that for increasing dimension the entanglement entropy of the optimal state decreases, agreeing qualitatively with the observations made by optimizing the Kullback-Leibler divergence. Interestingly, the numerics indicated that the maximal violation of the inequality tends, as the number of measurement outcomes tends to infinity, to the absolute bound imposed by the polytope of probability vectors. We conjectured from this a tight quantum Bell inequality for the $(2, 2, \infty)$ Bell scenario. Analytical proof of the tightness of this inequality is work in progress which will hopefully appear soon. Several approximate state for the optimal state were analyst.

At the end of the chapter some major differences between Kullback-Leibler divergence and total variation distance were discussed coming to the conclusion that, against common belief, the optimal setting one should use to perform a Bell experiment most efficiently is not the one which maximally violates the corresponding Bell inequality, but rather the one which maximizes the Kullback-Leibler divergence.

# Appendix A

# Entropy and Information

Entropy is an important quantity in classical and quantum information theory (e.g. see [51] and [6]). In this appendix we want to summarize some basic definitions and properties of notions of entropy in classical and quantum probability theory.

## A.1 The notion of entropy for classical probability theory

### A.1.1 Shannon entropy

Shannon entropy is an important concept in classical information theory. Consider a finite set $\mathscr{Z}$ and a random variable $Z$ defined over this set. The Shannon entropy quantifies how much information on average we gain when we learn $Z$.

Mathematically, the Shannon entropy is defined as follows:

**Definition A.1.1** (Shannon entropy)**.** *Let $\mathscr{Z}$ be a finite set and $\{z\}$ be an event in $\mathscr{Z}$. For the probability distribution $p(z)$ we define the Shannon entropy as*

$$H(p(z)) := -\sum_{z\in\mathscr{Z}} p(z)\log p(z). \tag{A.1}$$

*Here, as throughout the text, $\log(\cdot)$ refers to the logarithm with basis 2 and we adopted the convention $0\log 0 = 0$.*

The Shannon entropy has many interesting properties. In this appendix we will only state two basic properties which will be of interest for us. For further discussion the reader is referred to the above literature.

**Theorem A.1.2** (Properties of Shannon entropy)**.** *Let $\mathscr{Z}$ be a finite set, $\{z\}$ be an event in $\mathscr{Z}$ and $p(z)$ a probability distribution defined over this set. The Shannon entropy $H(p(z))$ has the following properties:*

*(i)* $H(p(z)) \geq 0$*, with equality for* $p(z) = 0\ \forall z \in \mathscr{Z}$.
*(ii)* $H(p(z)) \leq \log n$*, where n is the number of outcomes.*

**Proof.**

(i) is clear since $0 \leq p(z) \leq 1$ and $0 \log 0 = 0$. (ii) In the next section we show that the so-called relative entropy $H(p(z)||q(z)) := -H(p(x)) - \sum_{z\in\mathscr{Z}} p(z) \log q(z)$ is always positive, $H(p(z)||q(z)) \geq 0$. Hence, if we set $q(z)$ to be the uniform distribution $q(z) = 1/n$, we get: $0 \leq H(p(z)||1/n) = -H(p(x)) - \sum_{z\in\mathscr{Z}} p(z) \log 1/n = \log n - H(p(x))$ which completes the proof.

■

## A.1.2 Relative entropy

Another important entropy measure is the so-called relative entropy or Kullback-Leibler divergence. The relative entropy can be seen as a measure for the closeness of two probability distributions, $p(z)$ and $q(z)$, over the same set $\mathscr{Z}$.

Formally we define:

**Definition A.1.3** (Relative entropy/Kullback-Leibler divergence)**.** *Let $\mathscr{Z}$ be a finite set and $\{z\}$ be an event in $\mathscr{Z}$. For two probability distributions, $p(z)$ and $q(z)$, over $\mathscr{Z}$ we define the relative entropy or Kullback-Leibler divergence from $p(z)$ to $q(z)$ as*

$$H(p(z)||q(z)) := \sum_{z\in\mathscr{Z}} p(z) \log \frac{p(z)}{q(z)} = -H(p(x)) - \sum_{z\in\mathscr{Z}} p(z) \log q(z).$$

In Subsection 4.1.2 we use the Kullback-Leibler divergence as a measure of statistical strength between two distributions $Q(z)$ and $L(z)$. An intuitive justification for this usage is given in this subsection, as well as the relation to confidence in terms of a Bayesian approach.

A first motivation for the use of the Kullback-Leibler divergence as a measure for the closeness of two probability distributions is given by the following theorem:

**Theorem A.1.4** (Klein's inequality for the Kullback-Leibler divergence)**.** *The Kullback-Leibler divergence is non-negative,* $H(p(z)||q(z)) \geq 0$*, with equality iff* $p(z) = q(z)\ \forall z$*.*

**Proof.**

The proof uses the following inequality from standard text book analysis, $\log x \ln 2 = \ln x \leq x - 1,\ \forall x \geq 0$ with equality iff $x = 1$. Using this inequality we can write

$$\begin{aligned} H(p(z)||q(z)) &= \sum_{z\in\mathscr{Z}} p(z) \log \frac{p(z)}{q(z)} \\ &\geq \frac{1}{\ln 2} \sum_{z\in\mathscr{Z}} p(z) \left(1 - \frac{q(z)}{p(z)}\right) \\ &= \frac{1}{\ln 2} \sum_{z\in\mathscr{Z}} (p(z) - q(x)) = 0, \end{aligned}$$

which proofs the inequality. Further, it is easily seen that equality occurs only when $q(z)/p(z) = 1\ \forall z$, which completes the proof.

■

In this sense the Kullback-Leibler divergence has similar properties as the notion of distance, however, in general we have $H(p(z)||q(z)) \neq H(q(z)||p(z))$, for which $H(p(z)||q(z))$ is formally not a distance. Some understanding of this asymmetry is provided in Section 4.3.

# A.2 The notion of entropy for quantum probability theory

## A.2.1 Von Neumann entropy

In the previous section we introduced the Shannon entropy as a measure for the information gained on average when learning a probability distribution within the framework of classical probability theory. In context of quantum probability theory we can define a analogues quantity in terms of the so-called von Neumann entropy which generalizes the concept of Shannon entropy fro quantum states $\rho$.

Let us formally define:

**Definition A.2.1** (von Neumann entropy)**.** *Let $\rho$ be a state on a Hilbert space $\mathscr{H}$. We define the von Neumann entropy as*

$$S(\rho) := -\mathrm{tr}\,(\rho \log \rho)\,. \tag{A.2}$$

*If $\rho$ has eigenvalues $\mu_i$, we can write*

$$S(\rho) := -\sum_i \mu_i \log \mu_i. \tag{A.3}$$

In analogy to Theorem A.1.2 we can prove the following properties of the von Neumann entropy:

**Theorem A.2.2** (Properties of von Neumann entropy)**.** *Let $\rho$ be a state on a Hilbert space $\mathscr{H}$. The von Neumann entropy $S(\rho)$ has the following properties:*

*(i)* $S(\rho) \geq 0$*, with equality for the pure state.*
*(ii)* *For $\mathscr{H} = \mathbb{C}^d$, $S(\rho) \leq \log d$, with equality iff $\rho = \mathbf{I}/d$.*

**Proof.**

(i) is clear, since for the pure state $\rho = \mathrm{diag}(0,...,0,1,0,...,0)$ and hence $S(\rho) = -1\log 1 = 0$. (ii) In the next section we show that the so-called quantum relative entropy $S(\rho||\sigma) := -S(\rho) - \mathrm{tr}\,(\rho \log \sigma)$ is always positive, $S(\rho||\sigma) \geq 0$. Hence, if we set $\sigma$ to be $\sigma = \mathbf{I}/d$, we get: $0 \leq S(\rho||\mathbf{I}/d) = -S(\rho) - \mathrm{tr}\,(\rho \log 1/d) = \log d - S(\rho)$ which completes the proof.

■

The following theorem will be interesting for the discussion of entanglement entropy:

**Theorem A.2.3** (Von Neumann entropy of a composite system)**.** *Let $\rho$ be a pure state on a composite Hilbert space $\mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ and let $\rho_{\mathscr{A}} = \mathrm{tr}_{\mathscr{B}}\rho$ and $\rho_{\mathscr{B}} = \mathrm{tr}_{\mathscr{A}}\rho$ denote the marginal states, then*

$$S(\rho_{\mathscr{A}}) = S(\rho_{\mathscr{B}}) = -\sum_i \lambda_i^2 \log \lambda_i^2,$$

*where the $\lambda_i$ are the Schmidt coefficients.*

**Proof.**

From the Schmidt decomposition it is clear that both $\rho_{\mathscr{A}}$ and $\rho_{\mathscr{B}}$ have the same eigenvalues, namely the Schmidt coefficients squared, i.e. $\mu_i = \lambda_i^2$. The results follows then from the definition of the von Neumann entropy.

■

The quantity $E(\rho) = S(\rho_{\mathscr{A}}) = S(\rho_{\mathscr{B}})$ is what we called the entanglement entropy of $\rho$. In particular we can see that the pure state with the maximal entanglement entropy is the the one with equal Schmidt coefficients, the so-called maximally entangled state.

### A.2.2 Quantum relative entropy

As we defined the von Neumann entropy as a quantum analog of the Shannon entropy we now also want to define the quantum relative entropy as a analog of the Kullback-Leibler divergence:

**Definition A.2.4** (Quantum relative entropy)**.** *Let $\rho$ and $\sigma$ be a states on a Hilbert space $\mathscr{H}$. We define the quantum relative entropy as*

$$S(\rho||\sigma) := \operatorname{tr}(\rho\log\rho) - \operatorname{tr}(\rho\log\sigma). \tag{A.4}$$

As for the Kullback-Leibler divergence one can show that the quantum relative entropy non-negative which is known as Klein's inequality:

**Theorem A.2.5** (Klein's inequality for the quantum relative entropy)**.** *The quantum relative entropy is non-negative, $S(\rho||\sigma) \geq 0$, with equality iff $\rho = \sigma$.*

**Proof.**

Let us first decompose $\rho$ and $\sigma$ into orthonormal decompositions $\rho = \sum_i p_i|i\rangle\langle i|$ and $\sigma = \sum_j q_j|j\rangle\langle j|$. Hence, from the definition of the trace we have

$$S(\rho||\sigma) = \sum_i p_i \log p_i - \sum_i \langle i|\rho\log\sigma|i\rangle.$$

The term under the second sum becomes

$$\langle i|\rho\log\sigma|i\rangle = p_i\langle i|\left(\sum_j \log q_j|j\rangle\langle j|\right)|i\rangle = p_i\sum_j \log q_j M_{ij},$$

where we defined $M_{ij} = \langle i|j\rangle\langle j|i\rangle$. Clearly, $M_{ij} \geq 0$ and $\sum_i M_{ij} = \sum_j M_{ij} = 1$. Noting that $\log(\cdot)$ is a concave function we have $\sum_j \log q_j M_{ij} \leq \log r_i$, where $r_i = \sum_j M_{ij} q_j$ with equality iff there exist a value for j for which $M_{ij} = 1$. Hence we got

$$S(\rho||\sigma) \geq \sum_i p_i \log p_i - \sum_i p_i \log r_i \geq 0,$$

where the second inequality sign come from the fact that the middle term is just the Kullback-Leibner divergence $H(p_i||q_i) \geq 0$. It is not difficult to see that equality for the both relations above occurs exactly when $p_i = r_i \; \forall i$, but this means nothing but that $\rho$ and $\sigma$ must have the same eigenvalues and hence are identical. Thus we have shown that $S(\rho||\sigma) \geq 0$ with equality iff $\rho = \sigma$

■

# Bibliography


[1] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin, 1932.

[2] A. Kolmogoroff, *Grunbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933.

[3] A. Einstein, B. Podolsky, and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?," *Phys. Rev.* **47** (1935) 777.

[4] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics* **1** (1964) 195–200.

[5] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.* **23** (1969), no. 15, 880–884.

[6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[7] A. Aspect, P. Grangier, and G. Roger, "Experimental tests of realistic local theories via Bell's theorem," *Phys. Rev. Lett.* **47** (1981) 460.

[8] W. Tittel, J. Brendel, H. Zbinden, and N. Gisin, "Violation of Bell inequalities by photons more than 10 km apart," *Phys. Rev. Lett.* **81** (1998) 3563.

[9] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, "Violation of Bell's inequality under strict Einstein locality conditions," *Phys. Rev. Lett.* **81** (1998) 5039.

[10] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, "Experimental violation of a Bell's inequality with efficient detection," *Nature* **409** (2001) 791.

[11] B. Kümmerer and H. Maassen, "Elements of quantum probability," in *Quantum Probability Communications X*, R. Hudson and J. Lindsay, eds. World Scientific, Singapore, 1996.

[12] R. Gill, "Lecture notes on discrete quantum systems." preprint, 1995.

[13] R. Gill, "Critique of 'Elements of quantum probability'," in *Quantum Probability Communications XI*, R. Hudson and J. Lindsay, eds. World Scientific, Singapore, 1996.

[14] M. Reed and B. Simon, *Methods Of Modern Mathematical Physics. I: Functional Analysis*. Academic Press, New York, 1972.

[15] C. Piron, *Foundations of Quantum Physics*. Benjamin, Reading, 1976.

[16] A. M. Gleason, "Measures on the closed subspaces of a Hilbert space," *J. Math. Mech.* **6** (1957) 885–893.

[17] R. Cooke, M. Keane, and W. Moran, "An elementary proof of Gleason's theorem," *Math. Proc. Cambr. Phil. Soc.* **98** (1985) 117–128.

[18] E. Schrödinger, "Die gegenwärtige Situation in der Quantenmechanik," *Naturwissenschaften* **48-50** (1935).

[19] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," *Phys. Rev. A* **53** (1996), no. 4, 2046–2052, `quant-ph/9511030`.

[20] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of mixed states: Necessary and sufficient conditions," *Phys. Lett. A* **223** (1996) 1–8, `quant-ph/9605038`.

[21] P. A. M. Dirac, *The Principles of Quantum Mechanics*. Oxford University Press, Oxford, fourth ed., 1958.

[22] R. Gill, "Lecture notes on hidden variables." preprint, 1995.

[23] B. S. Tsirelson, "Some results and problems on quantum Bell-type inequalities.," *Hadronic Journal Supplement* **8** (1993), no. 4, 329–345.

[24] S. Pironio, *Aspects of quantum non-locality*. PhD thesis, Université Libre de Bruxelles, Faculté des Sciences, Service de Physique Théorique, 2004.

[25] G. M. Ziegler, *Lectures on Polytopes*, vol. 152 of *Graduate texts in Mathematics*. Springer, Berlin, 1995.

[26] A. Fine, "Hidden variables, joint probability, and the Bell inequalities," *Phys. Rev. Lett.* **48** (1982), no. 5, 291–295.

[27] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, "Bell inequalities for arbitrarily high-dimensional systems," *Phys. Rev. Lett.* **88** (2002), no. 4, 040404, `quant-ph/0106024`.

[28] A. Acin, R. Gill, and N. Gisin, "Optimal bell tests do not require maximally entangled states," *Phys. Rev. Lett.* **95** (2005) 210402, `quant-ph/0506225`.

[29] L. Masanes, "Tight Bell inequality for d-outcome measurements correlations," *Quantum Information and Computation* **3** (2002) 345, `quant-ph/0210073`.

[30] D. Greenberger, M. Horne, and A. Zeilinger, "Going beyond Bell's theorem," in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, M. Kafatos, ed., pp. 73–76. Kluwer, Dordrecht, 1989.

[31] N. D. Mermin *Physics Today* **43** (1990), no. 6, 9.

[32] N. D. Mermin, "Extreme quantum entanglement in a superposition of macroscopically distinct states," *Phys. Rev. Lett.* **65** (1990), no. 15, 1838–1840.

[33] B. S. Tsirl'son, "Quantum analogues of the Bell inequalities. The case of two spatially separated domains," *J. of Soviet Math.* **36** (1987), no. 4, 557–570.

[34] L. J. Landau, "Empirical two-point correlation functions," *Foundations of Physics* **18** (1988), no. 4, 449–460.

[35] A. Grothendieck, "Resume de la thorie metrique des produits tensoriels topologiques," *Boletim da Sociedade de Matemaática de São Paulo* **8** (1956) 1–79.

[36] B. S. Cirel'son, "Quantum generalizations of Bell's inequality," *Lett. Math. Phys.* **4** (1980), no. 2, 93–100.

[37] S. R. Finch, "Grothendieck's constant," in *Mathematical Constants*, ch. 3.11, pp. 235–237. Cambridge University Press, Cambridge, 2003.

[38] J.-L. Krivine, "Constantes de Grothendieck et fonctions de type positif sur les spheres.," *Adv. Math.* **31** (1979) 16–30.

[39] P. C. Fishburn and J. A. Reeds, "Bell inequalities, Grothendieck's constant, and root two.," *SIAM J. Discr. Math.* **7** (1994) 48–56.

[40] A. Acin, N. Gisin, and B. Toner, "Grothendieck's constant and local models for noisy entangled quantum states," *Phys. Rev. A* **73** (2006) 062105, `quant-ph/0606138`.

[41] W. van Dam, P. Grunwald, and R. Gill, "The statistical strength of nonlocality proofs," *IEEE-Transactions on Information Theory* **51** (2005) 2812–2835, `quant-ph/0307125`.

[42] A. Peres, "Bayesian analysis of Bell inequalities," *Fortsch. Phys.* **48** (2000) 531–535, `quant-ph/9905084`.

[43] S. Kullback and R. A. Leibler, "On information and sufficiency," *Annals of Mathematical Statistics* **22** (1951), no. 76-86,.

[44] T. Durt, D. Kaszlikowski, and M. Żukowski, "Violations of local realism with quantum systems described by n-dimensional Hilbert spaces up to $n = 16$," *Phys. Rev. A* **64** (2001) 024101, `quant-ph/0101084`.

[45] A. Acin, T. Durt, N. Gisin, and J. Latorre, "Quantum non-locality in two three-level systems," *Phys. Rev. A* **65** (2002) 052325, `quant-ph/0111143`.

[46] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2. Wiley, third ed., 1969.

[47] S. Zohren and R. D. Gill, "Maximal violation of the Collins-Gisin-Linden-Massar-Popescu inequality for infinite dimensional states," quant-ph/0612020.

[48] M. Keyl, D. Schlingemann, and R. F. Werner, "Infinitely entangled states," quant-ph/0212014.

[49] N. Korolkova and G. Leuchs, "Multimode quantum correlations," in *Coherence and statistics of photons and atoms*, J. Perina, ed. Wiley, 2001.

[50] J.-L. Chen, C. Wu, L. C. Kwek, C. H. Oh, and M.-L. Ge, "Violating Bell inequalities maximally for two d-dimensional systems," *unpublished* (2005) quant-ph/0507227.

[51] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.

# Index