COVARIANCE, CORRELATION AND ENTANGLEMENT IN QUANTUM SYSTEMS

SZABOLCS FÖLDES

BSc thesis
Eötvös Loránd University, 2025
Supervisor: Szilárd Szalay, PhD., senior research fellow
Wigner Research Center for Physics

ELTE
EÖTVÖS LORÁND
TUDOMÁNYEGYETEM

In this BSc thesis, we consider different aspects of the correlation between the parts of classical and quantum systems. Motivated by the fact that independence is not necessary for the vanishing of the covariance in classical probability theory, we investigate the relation between the covariance of observables and correlation (nonproductness) of states in classical and quantum systems. We show that productness is necessary for the vanishing of covariance in two-bit classical systems, while in any larger classical systems and in any quantum systems (including those of two qubits) the covariance can vanish in correlated systems.

Contents

1. Introduction

Quantum theory offers us several interesting and counter-intuitive phenomena. To get an understanding at least of the questions, it is enlightening to compare the quantum and classical description.

Quantum theory is a probabilistic theory, it only gives the probabilities of the different outcomes in measurements of the observable quantities. Traditionally, we think that observables can take values in a continuous range; however, in nature, we also see examples of discrete and finite ranges. Interesting quantum phenomena occur even in these discrete, finite systems, which are much simpler to describe.

Many of the unusual quantum behaviors emerge when one examines the relations among parts of a larger, composite system. These are the different forms of quantum correlations, the most important one is entanglement, which is the main resource in quantum information theory and quantum computing, offering advantages over the classical ones. This is because quantum correlations differ in many ways from classical correlations, we will explore one of these ways. We can best understand and illustrate the differences if we consider the classical and quantum cases side by side and give a probabilistic description for the classical systems analogous to that of the quantum ones, which might be not that usual.

First, in section 2, we recall the mathematical tools needed, such as vector spaces and Hilbert spaces, linear operations, and the Hilbert–Schmidt inner product. Then, using these tools, we establish the classical and quantum probability theories in section 3 including also the important entropic quantities coming from classical and quantum information theory. In section 4, we finally turn our attention to bipartite systems and the ways of quantifying the different aspects of correlations. In these sections we rely on the excellent textbooks [1,3,6].

In section 5, we consider the relation between the covariance of observables, and the correlation (nonproductness) of the state, in both the classical and the quantum cases. It is already known in classical probability theory that if the covariance of two probabilistic variables (observables) is zero, this does not necessarily mean that the probability distribution (state) is a product. (In the latter case, the probabilistic variables are called independent). In the general theory, we do not stick to the observables that much; especially in the quantum case, when we have noncommuting observables in the same system. So the more important notion is the correlation of the state itself, that is, its nonproductness. So, from our point of view, the difference mentioned is that even if a state is highly correlated, this correlation cannot necessarily be observed by measuring the zero or low covariance of a pair of unfortunately chosen observables. We prove in general that such situation cannot occur in classical two-bit systems (there it is enough for the covariance to vanish on a fixed pair of nontrivial observables and even this implies that the state is uncorrelated), but can occur in any larger classical systems and in any quantum systems, including two-qubit systems (there are pairs of observables of vanishing covariance in states which are correlated). This result is just another difference between classical and quantum systems: two-bit systems are too small, but even two-qubit systems are large enough so that the vanishing of the covariance is not sufficient for the productness.

2. Vector spaces

Vector spaces are the main mathematical tools in describing classical and quantum systems. It this work we consider discrete finite systems, so we are only concerned with finite dimensional vector spaces, especially Hilbert spaces.

2.1. **Vector spaces in general.** A *vector space* $\mathcal{V}$ over the field $\mathbb{C}$ is closed with respect to *linear combination*, that is, if $v_i \in \mathcal{V}$ and $c_i \in \mathbb{C}$ then $\sum_{i=1}^{m} c_i v_i \in \mathcal{V}$. Linear combination is understood to be of finite terms, even if this is not written explicitly.

Another structure which also turns out to be important in classical and quantum probability theory is convexity. A *convex space* $\mathcal{C} \subseteq \mathcal{V}$ in a vector space $\mathcal{V}$ is one that is closed with respect to *convex combination*, that is, if $u_i \in \mathcal{C}$, $w_i \geq 0$ and $\sum_i w_i = 1$ then $\sum_{i=1}^m w_i u_i \in \mathcal{C}$. The convex combination of a finite set of vectors gives a *convex polytope*.

A set of vectors is *linearly independent* if only their trivial linear combination can give the null vector. The maximal number of linearly independent vectors in the vector space $\mathcal{V}$ is the *dimension* of that vector space, $\dim(\mathcal{V})$.

The paradigmatic example of finite dimensional vector spaces is the $\mathbb{C}^d$ vector space of complex d-tuples $\boldsymbol{v} = (v_1, v_2, \ldots, v_d)$. In $\mathbb{C}^d$ one can define the p-norm as $\|\boldsymbol{v}\|_p := \left(\sum_i |v_i|^p\right)^{1/p}$ for $p \in [1, \infty]$, where $\|\boldsymbol{v}\|_\infty := \lim_{p\to\infty} \|\boldsymbol{v}\|_p = \max(|v_i|)$.

Linear operators $A \in \mathrm{Lin}(\mathcal{V}, \mathcal{W})$ are functions $\mathcal{V} \to \mathcal{W}$ between vector spaces with the property $A\left(\sum_i c_i v_i\right) := \sum_i c_i A(v_i)$, that is, they preserve the linear structure of the vector space. In case of linear operators, the parentheses are often omitted. Linear operators form a vector space too, by the definition $\left(\sum_i c_i A_i\right)v := \sum_i c_i (A_i v)$. If $\mathcal{W} = \mathcal{V}$ then the $\mathrm{Lin}(\mathcal{V}) := \mathrm{Lin}(\mathcal{V}, \mathcal{V})$ notation is used, and $I \in \mathrm{Lin}(\mathcal{V})$ for the identity operator. Since this is an algebra, we also use the notation $\mathcal{A} := \mathrm{Lin}(\mathcal{V})$.

The paradigmatic example of finite dimensional linear operator spaces is the $\mathbb{C}^d \otimes \mathbb{C}^{d'}$ linear space of complex $d \times d'$ matrices. In $\mathbb{C}^d \otimes \mathbb{C}^{d'}$ one can define the Schatten p-norm as $\|M\|_p := \left(\mathrm{Tr}(|M|^p)\right)^{1/p}$, where $|M| = \sqrt{M^\dagger M}$, so $\|M\|_p$ is just the p-norm of the vector of the singular values of M. Algebras are usually noncommutative, as in the case of the matrix algebras $\mathrm{Lin}(\mathbb{C}^d) = \mathbb{C}^d \otimes \mathbb{C}^d$, however, there are also commutative ones. For example, the complex d-tuples $\mathbb{C}^d$ with the elementwise multiplication $\boldsymbol{vw} = (v_1, v_2, \ldots, v_d)(w_1, w_2, \ldots, w_d) := (v_1 w_1, v_2 w_2, \ldots, v_d w_d)$, form a commutative algebra, which we denote as $\mathbb{A} := \mathbb{C}^d$, with the identity element $\mathbf{1} = (1, 1, \ldots, 1)$.

In the special case when the range of a linear operator is the base field of the original vector space, $\mathcal{W} = \mathbb{C}$, the operator $f \in \mathrm{Lin}(\mathcal{V}, \mathbb{C})$ is called a *linear functional*. The space of linear functionals is called the *dual space* of the vector space, $\mathcal{V}^* := \mathrm{Lin}(\mathcal{V}, \mathbb{C})$, and $\dim(\mathcal{V}^*) = \dim(\mathcal{V})$ holds.

The *tensor product* of vectors $v \in \mathcal{V}$ and $w \in \mathcal{W}$ is denoted as $v \otimes w$, where $\otimes$ is bilinear,

$$\left(\sum_i a_i v_i\right) \otimes \left(\sum_j b_j w_j\right) := \sum_i \sum_j a_i b_j (v_i \otimes w_j). \tag{2.1}$$

The tensor product of the vector spaces $\mathcal{V}$ and $\mathcal{W}$ is the vector space of linear combinations of *elementary tensors* $v \otimes w$, that is, $\mathcal{V} \otimes \mathcal{W} := \mathrm{Span}\{v \otimes w | v \in \mathcal{V}, w \in \mathcal{W}\}$. The dimension of this is $\dim(\mathcal{V}) \dim(\mathcal{W})$.

For $f \in \mathcal{V}^*$ and $w \in \mathcal{W}$, we have the linear operator $w \otimes f \in \mathrm{Lin}(\mathcal{V}, \mathcal{W})$ with the definition

$$(w \otimes f)(v) := f(v) w, \tag{2.2}$$

by which we have the identification

$$\mathrm{Lin}(\mathcal{V}, \mathcal{W}) = \mathcal{W} \otimes \mathcal{V}^*. \tag{2.3}$$

That is, every operator $A \in \mathrm{Lin}(\mathcal{V}, \mathcal{W})$ can be decomposed as

$$A = \sum_i c_i w_i \otimes f_i, \tag{2.4}$$

or, equivalently, $A = \sum_{i,j} A_{ij} w_i \otimes f_j$.

For operators in $\mathcal{A} = \mathrm{Lin}(\mathcal{V})$ we have the *trace* map, which is a linear map, acting on elementary operators as

$$\mathrm{Tr}(v \otimes f) := f(v), \tag{2.5}$$

by which

$$\mathrm{Tr}(A) = \mathrm{Tr}\Big(\sum_i c_i v_i \otimes f_i\Big) = \sum_i c_i f_i(v_i), \tag{2.6}$$

or $\mathrm{Tr}(\sum_{i,j} A_{ij} v_i \otimes f_j) = \sum_{i,j} A_{ij} f_j(v_i)$.

The *tensor product* of operators $A \in \mathrm{Lin}(\mathcal{V}, \mathcal{V}')$ and $B \in \mathrm{Lin}(\mathcal{W}, \mathcal{W}')$ is the linear operator given on elementary tensors as

$$(A \otimes B)(v \otimes w) := (Av) \otimes (Bw). \tag{2.7}$$

The tensor product of linear maps (acting on operators) can be given similarly to that of operators. Having $\mathcal{A} := \mathrm{Lin}(\mathcal{V})$ and $\mathcal{B} := \mathrm{Lin}(\mathcal{W})$, the *tensor product* of linear maps $\Phi \in \mathrm{Lin}(\mathcal{A}, \mathcal{A}')$ and $\Upsilon \in \mathrm{Lin}(\mathcal{B}, \mathcal{B}')$ is the linear map given on elementary tensors of linear operators $A \in \mathcal{A}$ and $B \in \mathcal{B}$ as

$$(\Phi \otimes \Upsilon)(A \otimes B) := \Phi(A) \otimes \Upsilon(B). \tag{2.8}$$

For example, the *partial trace* is

$$\mathrm{Tr}_{\mathcal{A}} := \mathrm{Tr} \otimes \mathcal{I}, \tag{2.9}$$

where $\mathrm{Tr} \in \mathrm{Lin}(\mathcal{A}, \mathbb{C})$ is the trace map of $\mathcal{A}$, $\mathcal{I} \in \mathrm{Lin}(\mathcal{B})$ is the identity map on $\mathcal{B}$, so $\mathrm{Tr}_{\mathcal{A}}(A \otimes B) = \mathrm{Tr}(A)B$.

2.2. **Hilbert spaces.** A *Hilbert space* $\mathcal{H}$ is a vector space over the complex numbers with a binary operation called the *inner product*

$$\langle .,. \rangle : \mathcal{H} \times \mathcal{H} \longrightarrow \mathbb{C}, \tag{2.10}$$

which is, by definition *non-negative* ($\langle \psi, \psi \rangle \geq 0$), *non-degenerate* ($\langle \psi, \psi \rangle = 0$ if and only if $\psi = 0$), *Hermitian* or *conjugate symmetric* ($\langle \psi, \phi \rangle = \langle \phi, \psi \rangle^*$), and *linear in the second argument* ($\langle \phi, \sum_i c_i \psi_i \rangle = \sum_i c_i \langle \phi, \psi_i \rangle$). Because of the previous two properties it is also true that the inner product is conjugate linear in its first argument ($\langle \sum_i c_i \phi_i, \psi \rangle = \sum_i c_i^* \langle \phi_i, \psi \rangle$). In the case of d-tuples $\mathbb{C}^d$, we use the notation $(\boldsymbol{v}|\boldsymbol{w}) = \sum_{i=1}^d v_i^* w_i$ for the usual inner product.

The vectors $\psi, \phi \in \mathcal{H}$ are called *orthogonal*, if $\langle \psi, \phi \rangle = 0$. It follows from the properties of the inner product that $\langle \psi, \psi \rangle \geq 0$, and then a norm can be defined as $\|\psi\| := \sqrt{\langle \psi, \psi \rangle}$. The vector $\psi \in \mathcal{H}$ is *normalized*, if $\|\psi\| = 1$.

It can be shown that with this inner product the *Cauchy-Bunyakovsky-Schwarz* inequality holds, so for all $\phi, \psi \in \mathcal{H}$,

$$|\langle \psi, \phi \rangle|^2 \leq \langle \psi, \psi \rangle \langle \phi, \phi \rangle, \tag{2.11}$$

and with the previously defined norm can be rewritten

$$|\langle \psi, \phi \rangle| \leq \|\psi\| \|\phi\|. \tag{2.12}$$

We also have a more general inequality for p-norms, as the consequence of Hölder's inequality,

$$|\langle \psi, \phi \rangle| \leq \|\psi\|_p \|\phi\|_q, \tag{2.13}$$

for any p- and q-norms where $\frac{1}{p} + \frac{1}{q} = 1$. A special case of (2.13) is the Cauchy-Bunyakovsky-Schwarz inequality above when $p = q = 2$.

We again have the vector spaces of linear operators $\mathrm{Lin}(\mathcal{H}, \mathcal{K})$, the operator algebra $\mathcal{A} := \mathrm{Lin}(\mathcal{H})$, and the dual Hilbert space $\mathcal{H}^* = \mathrm{Lin}(\mathcal{H}, \mathbb{C})$. For an operator $A \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, we can define its *adjoint* $A^\dagger \in \mathrm{Lin}(\mathcal{K}, \mathcal{H})$ as

$$\langle A^\dagger \psi, \phi \rangle := \langle \psi, A\phi \rangle \tag{2.14}$$

for all $\psi \in \mathcal{K}$ and $\phi \in \mathcal{H}$.

Having the inner product, we have a natural way to assign to every element in the space $\mathcal{H}$ an element in the dual space $\mathcal{H}^*$. That is, for all $\psi \in \mathcal{H}$ let $\psi^\dagger \in \mathcal{H}^*$ be the linear functional acting on all $\phi \in \mathcal{H}$ as $\psi^\dagger(\phi) := \langle \psi, \phi \rangle$. This map is antilinear, $\big(\sum_i c_i \psi_i\big)^\dagger = \sum_i c_i^* \psi_i^\dagger$, and bijective

in this finite dimensional case. This leads to the very convenient *Dirac notation*. In the Dirac notation vectors $\phi \in \mathcal{H}$ are denoted as $|\phi\rangle := \phi$, called "ket" and linear functionals $\psi^\dagger \in \mathcal{H}^*$ as $\langle\psi| := \psi^\dagger$, called "bra". Then the functional acting on the vector is $\psi^\dagger(\phi) = \langle\psi|(|\phi\rangle) = \langle\psi, \phi\rangle$, which is then denoted as $\langle\psi|\phi\rangle := \langle\psi, \phi\rangle$.

Dirac notation is also convenient for operators. In case of elementary operators $|\psi\rangle \otimes \langle\phi| \in \mathrm{Lin}(\mathcal{H}, \mathcal{K}) = \mathcal{K} \otimes \mathcal{H}^*$, we omit the $\otimes$ symbol, and write simply $|\psi\rangle\langle\phi|$, as usual in the literature. Linear operators $A \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ are in general of the form

$$A = \sum_i c_i |\alpha_i\rangle\langle\beta_i| \tag{2.15}$$

because of equation (2.4). Also, the trace map is $\mathrm{Tr}(|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle$ for elementary operators $|\psi\rangle\langle\phi| \in \mathcal{A} = \mathrm{Lin}(\mathcal{H})$, or $\mathrm{Tr}\big(\sum_i c_i|\alpha_i\rangle\langle\beta_i|\big) = \sum_i c_i\langle\beta_i|\alpha_i\rangle$ for $A = \sum_i c_i|\alpha_i\rangle\langle\beta_i| \in \mathcal{A}$ in general. The *adjoint* turns out to be $(|\psi\rangle\langle\phi|)^\dagger = |\phi\rangle\langle\psi|$ for elementary operators $|\psi\rangle\langle\phi| \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$, or $\big(\sum_i c_i|\alpha_i\rangle\langle\beta_i|\big)^\dagger = \sum_i c_i^*|\beta_i\rangle\langle\alpha_i|$ for $A = \sum_i c_i|\alpha_i\rangle\langle\beta_i| \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ in general.

All operators could be decomposed as in equation (2.4), but this decomposition is not unique in general. In the case of Hilbert spaces, all operators $A \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ can be written as

$$A = \sum_i a_i |\alpha_i\rangle\langle\beta_i|, \tag{2.16}$$

where the *singular values* are $a_i \geq 0$, and the *singular vectors* $|\alpha_i\rangle \in \mathcal{K}$ and $|\beta_i\rangle \in \mathcal{H}$ form orthonormal sets. This decomposition is essentially unique (up to degeneracies), and called *singular value decomposition*. Let us recall also some special cases of operators important in quantum probability theory. An operator $A \in \mathcal{A}$ is *normal*, if $A^\dagger A = AA^\dagger$, which holds if and only if it can be written in the 'diagonal form'

$$A = \sum_i a_i |\alpha_i\rangle\langle\alpha_i|, \tag{2.17}$$

where the *eigenvalues* are $a_i \in \mathbb{C}$ and the eigenvectors $|\alpha_i\rangle$ are orthonormal. An operator $A \in \mathcal{A}$ is *unitary*, if $A^\dagger = A^{-1}$, which holds if and only if it is normal and the eigenvalues in equation (2.17) are $a_i \in \mathbb{C}$, $|a_i| = 1$. An operator $A \in \mathcal{A}$ is *self-adjoint*, if $A^\dagger = A$, which holds if and only if it is normal and the eigenvalues in equation (2.17) are $a_i \in \mathbb{R}$. An operator $A \in \mathcal{A}$ is *positive semidefinite*, denoted as $A \geq 0$, if $\langle\psi|A|\psi\rangle \geq 0$ for all $|\psi\rangle \in \mathcal{H}$, which holds if and only if it is normal and the eigenvalues in equation (2.17) are $a_i \geq 0$. An operator $A \in \mathcal{A}$ is a *projection*, if $P^2 = P = P^\dagger$, which holds if and only if it is normal and the eigenvalues in equation (2.17) are $a_i \in \{0, 1\}$.

For linear operators $A, B \in \mathrm{Lin}(\mathcal{H}, \mathcal{K})$ we can define the *Hilbert-Schmidt inner product*

$$(A|B) := \mathrm{Tr}(A^\dagger B). \tag{2.18}$$

With this, $\mathrm{Lin}(\mathcal{H}, \mathcal{K})$ is also a Hilbert space, called *Hilbert-Schmidt space*. Although Dirac notation is rarely used here, it still holds that linear functionals over the Hilbert-Schmidt space can be given by the Hilbert-Schmidt inner product with operators.

The tensor product of Hilbert spaces $\mathcal{H}_1$, $\mathcal{H}_2$ is also a Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$ with the inner product

$$\big\langle \psi_1 \otimes \psi_2 \big| \phi_1 \otimes \phi_2 \big\rangle := \langle\psi_1|\phi_1\rangle\langle\psi_2|\phi_2\rangle \tag{2.19}$$

from the inner products of $\mathcal{H}_1$ and $\mathcal{H}_2$.

3. Probability theory

Probability theory deals with probabilistic variables and the probabilities of their measurement outcomes. Those physical quantities which are described probabilistically are called *observables*, they are elements in the observable algebra, in both the classical and the quantum cases. Linear

functionals giving the expectation values of observables (equivalently, the probabilities of the outcomes) are called *states* in both the classical and the quantum cases. In the classical theory the observable algebra is commutative, however in the quantum case it is not. In this work we deal with *discrete finite systems*, this means that the observables give a finite number of discrete values when measured. Then the observable algebras and state spaces are finite dimensional.

3.1. **Classical probability theory.** Let us have observables taking numerical (possibly complex) values a_i in measurement. It is meaningful to have linear combination and (commutative) product of observables, so let us represent them as $\boldsymbol{a} := (a_1, a_2, \dots, a_d) \in \mathbb{A} := \mathbb{C}^d$ with the element-wise multiplication, $(\boldsymbol{ab})_{ij} := a_i b_j$. We can write

$$\boldsymbol{a} = \sum_{i=1}^{d} a_i \boldsymbol{\chi}_i, \tag{3.1}$$

where $\boldsymbol{\chi}_i$ is the characteristic observable of the i-th measurement outcome, $\boldsymbol{\chi}_1 = (1, 0, 0, \dots, 0)$, $\boldsymbol{\chi}_2 = (0, 1, 0, \dots, 0)$ and so on. As we can see $\boldsymbol{\chi}_i$ takes the value 1 if the i-th outcome occurs, and 0 otherwise.

Let us have the probability of the i-th outcome p_i. That is, $p_i \geq 0$ and $\sum_{i=1}^{d} p_i = 1$. We can form the expectation value of the observable $\boldsymbol{a}$

$$\mathrm{Exp}_{\boldsymbol{p}}(\boldsymbol{a}) = \sum_{i=1}^{d} p_i a_i = (\boldsymbol{p}|\boldsymbol{a}), \tag{3.2}$$

given by the inner product in $\mathbb{C}^d$. We call such d-tuple $\boldsymbol{p} := (p_1, p_2, \dots, p_d) \in \mathbb{C}^d$ a *state*. In general, a state is a linear functional, an element of the dual of the observable algebra, however, we use the identification of the space and its dual by the inner product. And so similarly to the observable a more illustrative form of a state is

$$\boldsymbol{p} = \sum_{i=1}^{d} p_i \boldsymbol{\delta}_i, \tag{3.3}$$

where $\boldsymbol{\delta}_i$ is a *pure state*. All the components of a pure state are 0 except the i-th which is 1, $\boldsymbol{\delta}_1 = (1, 0, 0, ..)$, $\boldsymbol{\delta}_2 = (0, 1, 0, \dots)$ and so on. As we can see, in case of $\boldsymbol{\delta}_i$ the i-th outcome occurs with certainty, and the others cannot occur. Another special state is the *white noise* which means a fully random state, that is $\mathbf{1}/d := (1/d, 1/d, \dots, 1/d)$, all the outcomes occur with equal probabilities. The expectation value (3.2) is the inner product of an observable and a state, which illustrates well that the expectation value depends just as much on the state as on the observable we are measuring.

Let us have the space of all the possible states $\boldsymbol{p}$ (nonnegative, normalized), the *state space*

$$\Delta := \Big\{ \boldsymbol{p} \in \mathbb{C}^d \;\Big|\; p_i \geq 0, \sum_i p_i = 1 \Big\}, \tag{3.4}$$

which is a subset of dimension $d-1$ in $\mathbb{R}^d$. The main structure in probability theory is convexity, which means that the state space is closed under convex combination

$$\sum_{j=1}^{m} w_j \boldsymbol{p}_j \in \Delta, \quad \text{where } w_j \geq 0, \sum_{j=1}^{m} w_j = 1. \tag{3.5}$$

Note that it is not closed under linear combination, and so the linear combination of arbitrary states is not a state, only their convex combination. By equation (3.3), the state space is the convex hull of the finite number (d) of pure states, so it is a convex polytope. It is of dimension $d-1$, so it is a simplex, which means that the pure convex decomposition of every state is unique. The white noise $\mathbf{1}/d$ is the center of the state space in some sense.

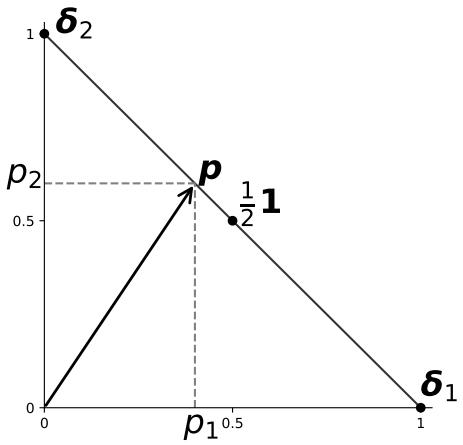

FIGURE 1. State space of a bit in the natural basis

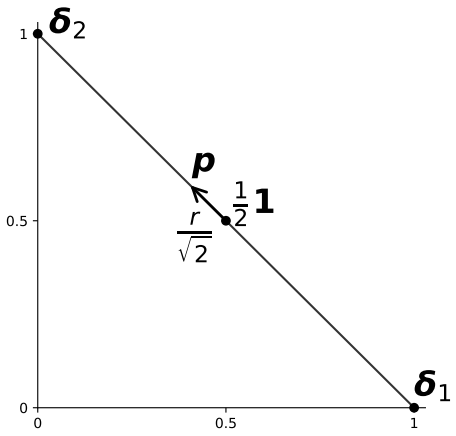

FIGURE 2. State space of a bit in the centered basis

Calculating the expectation value of a characteristic observable yields the probability of that outcome

$$\mathrm{Exp}_{\boldsymbol{p}}(\boldsymbol{\chi}_i) = \sum_j p_j(\boldsymbol{\delta}_j|\boldsymbol{\chi}_i) = p_i. \tag{3.6}$$

The $\boldsymbol{a}^2$ squared of an observable $\boldsymbol{a} \in \mathbb{A}$ (the square is taken elementwisely) is also an observable and so its expectation value is as equation (3.2) also. This way the variance of a given $\boldsymbol{a}$ observable is

$$\mathrm{Var}_{\boldsymbol{p}}(\boldsymbol{a}) := \mathrm{Exp}_{\boldsymbol{p}}((\boldsymbol{a} - \mathrm{Exp}_{\boldsymbol{p}}(\boldsymbol{a}))^2) = \mathrm{Exp}_{\boldsymbol{p}}(\boldsymbol{a}^2) - \mathrm{Exp}_{\boldsymbol{p}}(\boldsymbol{a})^2 = (\boldsymbol{p}|\boldsymbol{a}^2) - (\boldsymbol{p}|\boldsymbol{a})^2. \tag{3.7}$$

The most simple discrete finite system describes observables having two outcomes ($d = 2$), called a *bit*. And so the pure states are

$$\boldsymbol{\delta}_1 = (1, 0),\ \boldsymbol{\delta}_2 = (0, 1). \tag{3.8}$$

The state space is a 1-dimensional line segment. In the center of the line segment is the $(1/2, 1/2)$ white noise. Figure 1 illustrates the state space of the bit. We have the usual parametrization

$$\boldsymbol{p} = p_1\boldsymbol{\delta}_1 + p_2\boldsymbol{\delta}_2 = (p_1, p_2), \tag{3.9}$$

but we can also use a different one, expressing the difference from the white noise,

$$\boldsymbol{p} = \frac{1}{2}(\mathbf{1} + r\boldsymbol{\sigma}) = \frac{1}{2}(1 + r, 1 - r). \tag{3.10}$$

see in figure 2, where $\mathbf{1} = \boldsymbol{\delta}_1 + \boldsymbol{\delta}_2 = (1, 1)$, $\boldsymbol{\sigma} = \boldsymbol{\delta}_1 - \boldsymbol{\delta}_2 = (1, -1)$, so $r = 2p_1 - 1 = 1 - 2p_2 \in [-1, 1]$.

Considering an observable with 3 outcomes we get the trit ($d = 3$), with pure states

$$\boldsymbol{\delta}_1 = (1, 0, 0),\ \boldsymbol{\delta}_2 = (0, 1, 0),\ \boldsymbol{\delta}_3 = (0, 0, 1). \tag{3.11}$$

A general $\boldsymbol{p}$ state given by these pure states is

$$\boldsymbol{p} = p_1\boldsymbol{\delta}_1 + p_2\boldsymbol{\delta}_2 + p_3\boldsymbol{\delta}_3, \tag{3.12}$$

so the state space is an equilateral triangle with the white noise in the center of the triangle, this illustrated in figure 3

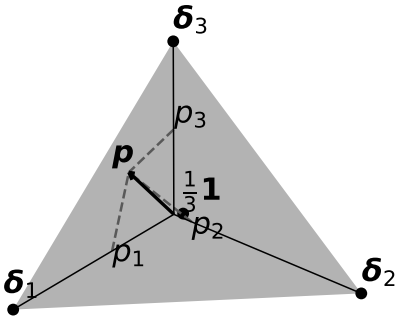

FIGURE 3. State space of a trit

3.2. **Quantum probability theory.** In quantum theory observables are represented by normal operators of a $\mathcal{H}$ Hilbert space

$$A = \sum_i a_i |\alpha_i\rangle\langle\alpha_i| \in \mathcal{A} = \mathrm{Lin}(\mathcal{H}), \tag{3.13}$$

where the eigenvalues a_i are the discrete values of the measurable quantity and $|\alpha_i\rangle\langle\alpha_i|$ are the respective projections of the eigenspaces. We can think of $P_i = |\alpha_i\rangle\langle\alpha_i|$ as a characteristic observable, taking a value 1 if the i-th outcome occurs in a given measurement and 0 otherwise.

The expectation value of an observable $A \in \mathcal{A}$ is given by the *density operator* $\rho \in \mathcal{A}^* = \mathrm{Lin}(\mathcal{H}, \mathbb{C})$, also called quantum state, as

$$\mathrm{Exp}_\rho(A) = \mathrm{Tr}(\rho A). \tag{3.14}$$

In general, a state is a linear functional, an element of the $\mathcal{A}^*$ dual of the observable algebra $\mathcal{A}$, however, we use the identification of the algebra and its dual by the Hilbert-Schmidt inner product. The Born rule is that the probability of a measurement outcome is given by the expectation value of the characteristic observable $P_i = |\alpha_i\rangle\langle\alpha_i|$,

$$p_i = \mathrm{Tr}(\rho P_i), \tag{3.15}$$

which is equivalent to the formula (3.14) for the expectation value. The collection of the p_i probabilities together is the measurement statistics of the observable $A \in \mathcal{A}$. The properties of the density operators follow from the Born rule. To get $p_i \in \mathbb{R}$, the density operator needs to be self adjoint, $\rho^\dagger = \rho$; to get $p_i \geq 0$, the density operator needs to be positive semidefinite, $\rho \geq 0$; and to get $\sum_{i=1}^d p_i = 1$, the density operator needs to be (trace-)normalized, $\mathrm{Tr}(\rho) = 1$. If we consider equation (3.15) for a state $\pi = |\psi\rangle\langle\psi|$, then for the probability we get

$$p_i = \langle\psi|\alpha_i\rangle\langle\alpha_i|\psi\rangle = |\langle\alpha_i|\psi\rangle|^2, \tag{3.16}$$

this is the usual form of the *Born rule*, however equation (3.15) is the general form. The special state of the form $\pi = |\psi\rangle\langle\psi| \in \mathcal{A}$ is called a *pure state*, and the vector $|\psi\rangle \in \mathcal{H}$ defining it is

a *state vector*, $\|\psi\| = 1$. If $|\psi\rangle = |\alpha_i\rangle$, then the pure state $\pi = |\alpha_i\rangle\langle\alpha_i| = P_i$ is the pure state describing when the i-th outcome of the observable $A = \sum_i a_i |\alpha_i\rangle\langle\alpha_i|$ occurs with certainty, and the others cannot occur. That is, the measurement statistics of the observable A is a classical pure state. However even if the state is pure there exists observables for which the measurement statistics is not pure, that is there will be no outcomes with certainty. In a classical system no such state exists, it is thanks to the commutativity of the underlying algebra, that is not there in quantum theory. A state is pure if and only if its squared is itself $\rho^2 = \rho$, that is it is a rank-1 projection. A state that is not pure, a *mixed state*, consists of a convex combination of pure states, for such a state there exists no observable for which the measurement statistics is pure.

Let us have the space of all the possible quantum states ρ (positive semidefinite, normalized), the state space

$$\mathcal{D} := \Big\{ \rho \in \mathrm{Lin}(\mathcal{H}) \;\Big|\; \rho \geq 0, \mathrm{Tr}(\rho) = 1 \Big\}, \tag{3.17}$$

which is a subset of (real) dimension $d^2 - 1$ in $\mathrm{Lin}(\mathcal{H})$. The main structure in probability theory is convexity, which means that the state space is closed under convex combination,

$$\sum_{j=1}^{m} w_j \rho_j \in \mathcal{D}, \quad \text{where } w_j \geq 0, \sum_{j=1}^{m} w_j = 1. \tag{3.18}$$

Note that it is not closed under linear combination, and so the linear combination of arbitrary states is not a state, only their convex combination. The state space is the convex hull of the continuously many pure states,

$$\mathcal{D} = \Big\{ \rho \in \mathrm{Lin}(\mathcal{H}) \;\Big|\; \rho = \sum_j w_j |\psi_j\rangle\langle\psi_j|, |\psi_j\rangle \in \mathcal{H}, \|\psi_j\| = 1, w_j \in \mathbb{R}, w_j \geq 0, \sum_j w_j = 1 \Big\}. \tag{3.19}$$

Let us also have the space of pure states

$$\mathcal{P} := \Big\{ \pi \in \mathrm{Lin}(\mathcal{H}) \;\Big|\; \pi = |\psi\rangle\langle\psi|, |\psi\rangle \in \mathcal{H}, \|\psi\| = 1 \Big\}, \tag{3.20}$$

this is a $2d - 2$ (real) dimensional subset of Δ_{12}. So a state can be generally written as

$$\rho = \sum_i w_i |\psi_i\rangle\langle\psi_i|. \tag{3.21}$$

Contrary to the classical case, the pure convex decomposition of a (non-pure) state is *not* unique, there are infinitely many different decompositions. However there is a minimal number of pure states that is needed for the decomposition, which is the rank of the state. A special state is the *white noise* $I/d \in \mathcal{D}$ for which the measurement statistics is the (classical) white noise for any observable. Again the white noise is the center of the state space in some sense.

The $A^2 \in \mathcal{A}$ square of an observable $A \in \mathcal{A}$ is $A^2 = \sum_i a_i^2 |\alpha_i\rangle\langle\alpha_i|$, and its expectation value is $\mathrm{Exp}_\rho(A^2) = \mathrm{Tr}(\rho A^2)$. The variance $\mathrm{Var}_{\boldsymbol{p}}(\boldsymbol{a})$ of the measured values $\boldsymbol{a} = (a_1, a_2, \ldots, a_d) \in \mathbb{C}^d$ by the measurement statistics $\boldsymbol{p} = (p_1, p_2, \ldots, p_d) \in \mathbb{C}^d$ can also be formulated in terms of operators,

$$\mathrm{Var}_\rho(A) := \mathrm{Exp}_\rho((A - \mathrm{Exp}_\rho(A)I)^2) = \mathrm{Exp}_\rho(A^2) - \mathrm{Exp}_\rho(A)^2 = \mathrm{Tr}(\rho A^2) - \mathrm{Tr}(\rho A)^2, \tag{3.22}$$

which equals to $\mathrm{Var}_{\boldsymbol{p}}(\boldsymbol{a})$. For a pure state there exist observables of 0 variance and for mixed states there does not.

Note that if we consider only observables which are commuting (for example, when only those are accessible), then these can be simultaneously diagonalized, and the quantum probability theory boils down to the classical one.

The simplest quantum system is the qubit, with $d = 2$. Thus $\mathcal{D}$ is a $\dim(\mathcal{D}) = 3$ dimensional convex body. In $\mathrm{Lin}(\mathcal{H})$ the basis (orthonormal with respect to the Hilbert-Schmidt inner product) is given by I and the *Pauli operators* $\left(\frac{I}{\sqrt{2}}, \frac{\sigma_1}{\sqrt{2}}, \frac{\sigma_2}{\sqrt{2}}, \frac{\sigma_3}{\sqrt{2}}\right)$ with the following multiplication

rule

$$\sigma_a\sigma_b = \delta_{ab}I + i\sum_{c=1}^{3}\varepsilon_{abc}\sigma_c. \tag{3.23}$$

The Pauli operators in matrix form are

$$\sigma_1 = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}, \sigma_2 = \begin{bmatrix}0 & -i\\ i & 0\end{bmatrix}, \sigma_3 = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}. \tag{3.24}$$

We can express the commutator and anticommutator for the above basis as

$$[\sigma_a, \sigma_b] = \sigma_a\sigma_b - \sigma_b\sigma_a = 2i\sum_{c=1}^{3}\varepsilon_{abc}\sigma_c, \tag{3.25}$$

$$\{\sigma_a, \sigma_b\} = \sigma_a\sigma_b + \sigma_b\sigma_a = 2\delta_{ab}I. \tag{3.26}$$

With this basis a given state is parametrized by $\boldsymbol{r} = (r_1, r_2, r_3) \in \mathbb{R}^3$ the *Bloch vector* as $\rho = \frac{1}{2}(I + \boldsymbol{r}\boldsymbol{\sigma})$, where $\boldsymbol{r}\boldsymbol{\sigma} = r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3$. This affine map preserves the convex combination, that is $\rho(\sum_j w_j\boldsymbol{r}_j) = \sum_j w_j\rho(\boldsymbol{r}_j)$ where $w \geq 0, \sum_j w_j = 1$, so we have the same geometry in the space of Bloch vectors as in the space of operators. To find out the shape of the state space $\mathcal{D}$, let us consider the pure states first, for which the $\rho^2 = \rho$ equality must hold,

$$\frac{1}{4}(I + \boldsymbol{r}\boldsymbol{\sigma})(I + \boldsymbol{r}\boldsymbol{\sigma}) \overset{!}{=} \frac{1}{2}(I + \boldsymbol{r}\boldsymbol{\sigma}). \tag{3.27}$$

The left-hand side is $\frac{1}{4}(I + \|\boldsymbol{r}\|^2 + 2\boldsymbol{r}\boldsymbol{\sigma})$, expanding it we see that this equation can hold if and only if $\|\boldsymbol{r}\|^2 = 1$, so the space of pure states $\mathcal{P}$ is a sphere, and the state space $\mathcal{D}$ is the convex hull of this, which is a (solid) ball. These are called Bloch sphere and Bloch ball, respectively. The white noise $\frac{1}{2}I$ is the $\|\boldsymbol{r}\|^2 = 0$ center of the Bloch ball.

We want to compute the eigenvalues of the state ρ. For this we have to know the eigenvalues of $\boldsymbol{r}\boldsymbol{\sigma}$. To get these we need to realize two things. First $(\boldsymbol{r}\boldsymbol{\sigma})^2 = \|\boldsymbol{r}\|^2 I$, because of this the squared of all the eigenvalues is $\|\boldsymbol{r}\|^2$. Secondly $\mathrm{Tr}(\boldsymbol{r}\boldsymbol{\sigma}) = 0$, so the eigenvalues can only be the opposites of each other. This way the eigenvalues of $\boldsymbol{r}\boldsymbol{\sigma}$ are $\pm\|\boldsymbol{r}\|$. So the spectral decomposition of $\boldsymbol{r}\boldsymbol{\sigma}$ is

$$\boldsymbol{r}\boldsymbol{\sigma} = \|\boldsymbol{r}\||\phi_+\rangle\langle\phi_+| - \|\boldsymbol{r}\||\phi_-\rangle\langle\phi_-|. \tag{3.28}$$

Since $I = |\phi_+\rangle\langle\phi_+| + |\phi_-\rangle\langle\phi_-|$, we can express the spectral projections as $|\phi_\pm\rangle\langle\phi_\pm| = (I \pm \frac{\boldsymbol{r}\boldsymbol{\sigma}}{\|\boldsymbol{r}\|})/2$. Then we also have the spectral decomposition of ρ as

$$\rho = \frac{1}{2}(I + \boldsymbol{r}\boldsymbol{\sigma}) = \frac{1}{2}\Big(\big(1 + \|\boldsymbol{r}\|\big)|\phi_+\rangle\langle\phi_+| + \big(1 - \|\boldsymbol{r}\|\big)|\phi_-\rangle\langle\phi_-|\Big). \tag{3.29}$$

In this expression we see that the ρ density operator has the eigenvalues $\big(1 \pm \|\boldsymbol{r}\|\big)/2$, and the same spectral projections as $\boldsymbol{r}\boldsymbol{\sigma}$.

Observables can be expressed via the $\left(\frac{I}{\sqrt{2}}, \frac{\sigma_1}{\sqrt{2}}, \frac{\sigma_2}{\sqrt{2}}, \frac{\sigma_3}{\sqrt{2}}\right)$ basis also. An observable $A = \sum_{i=1}^{2} a_i|\alpha_i\rangle\langle\alpha_i| \in \mathcal{A}$ can be expressed as $A = A_0 I + \boldsymbol{A}\boldsymbol{\sigma}$, where $\boldsymbol{A}\boldsymbol{\sigma} = A_1\sigma_1 + A_2\sigma_2 + A_3\sigma_3$. If $\boldsymbol{A} \in \mathbb{R}^3$ then the spectral decomposition of the observable A can be obtained similarly to that of the density operator, for example, the eigenvalues are $A_0 \pm \|\boldsymbol{A}\|$.

An illustrative example of an observable is the $S_{\hat{\boldsymbol{v}}} = \frac{\hbar}{2}\hat{\boldsymbol{v}}\boldsymbol{\sigma}$ spin operator of direction $\hat{\boldsymbol{v}} \in \mathbb{R}^3$, where $\|\hat{\boldsymbol{v}}\| = 1$. The spectral decomposition of this is $S_{\hat{\boldsymbol{v}}} = \frac{\hbar}{2}|\alpha_+\rangle\langle\alpha_+| - \frac{\hbar}{2}|\alpha_-\rangle\langle\alpha_-|$, where the spectral projections are $|\alpha_\pm\rangle\langle\alpha_\pm| = (I \pm \hat{\boldsymbol{v}}\boldsymbol{\sigma})/2$. Measuring this in a state $\rho = \frac{1}{2}(I + \boldsymbol{r}\boldsymbol{\sigma})$ for the $+/-$ outcomes gives the measurement statistics

$$p_\pm = \mathrm{Tr}(\rho(\boldsymbol{r})|\alpha_\pm\rangle\langle\alpha_\pm|) = \mathrm{Tr}\left(\frac{1}{2}(I + \boldsymbol{r}\boldsymbol{\sigma})\frac{1}{2}(I \pm \hat{\boldsymbol{v}}\boldsymbol{\sigma})\right) = \frac{1}{4}\mathrm{Tr}(I \pm \hat{\boldsymbol{v}}\boldsymbol{r}I) = \frac{1}{2}(1 \pm \hat{\boldsymbol{v}}\boldsymbol{r}). \tag{3.30}$$

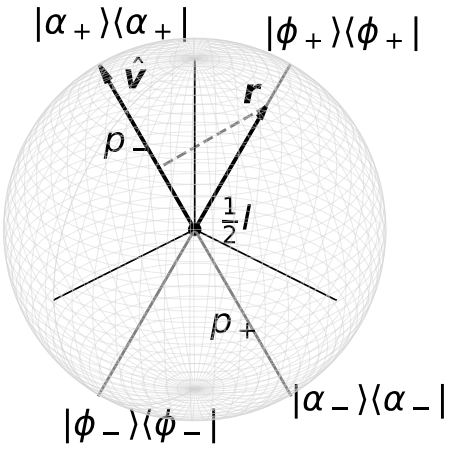

FIGURE 4. State space of a qubit, the Bloch ball and the measurement statistics

What this means is that one will get the outcome $\pm\frac{\hbar}{2}$ with probabilities $\frac{1}{2}(1\pm\hat{\boldsymbol{v}}\boldsymbol{r})$. So even if one measures in a pure state it is still possible to not get a certain outcome, for a certain outcome $\hat{\boldsymbol{v}}$ and $\boldsymbol{r}$ must point in the same (or complelety opposite) direction. One can see the Bloch ball and the measurement statistics on figure 4.

We can see that the classical bit is a special case of the quantum bit, when both $\hat{\boldsymbol{v}} = (0,0,1)$ and $\boldsymbol{r} = (0,0,r)$ points in the z direction. Then the matrices of A and ρ are diagonal, and the diagonal entries of ρ form the classical state (p_+, p_-), parametrized by r, see equation (3.10). More generally, if the considered observables are commuting, then these can be simultaneously diagonalized, and the quantum probability theory boils down to the classical one.

3.3. **Classical entropies.** In the classical theory we have the *Shannon-entropy* for quantifying the uncertainty of a state $\boldsymbol{p} \in \Delta$

$$S(\boldsymbol{p}) := -\sum_{i=1}^{d} p_i \ln p_i. \tag{3.31}$$

The Shannon-entropy takes values in the range $[0, \ln(d)]$, if we make the function $f(x) = -x\ln x$ right-continuous in $x = 0$, that is $f(0) := \lim_{x\to 0+} f(x) = 0$. It takes the value 0 if and only if $\boldsymbol{p}$ is a pure state and $\ln(d)$ if and only if the state is the white noise. The Shannon entropy quantifies the information content of a source with an outcome distribution $\boldsymbol{p}$, by *Shannon's noiseless coding theorem*, for the optimal length L of an output (bitstring) code in terms of the Shannon entropy as

$$\frac{S(\boldsymbol{p})}{\ln(2)} \leq L \leq \frac{S(\boldsymbol{p})+1}{\ln(2)}. \tag{3.32}$$

The Shannon entropy is concave, that is

$$\sum_i w_i S(\boldsymbol{p}_i) \leq S\Big(\sum_i w_i \boldsymbol{p}_i\Big), \tag{3.33}$$

expressing that the entropy is increasing for information loss. More importantly the Shannon entropy is *Schur-concave*, that is,

$$\boldsymbol{p}_1 \preceq \boldsymbol{p}_2 \quad \Longrightarrow \quad S(\boldsymbol{p}_1) \geq S(\boldsymbol{p}_2), \tag{3.34}$$

where $\boldsymbol{p}_1 \preceq \boldsymbol{p}_2$ means that $\boldsymbol{p}_1$ majorizes $\boldsymbol{p}_2$, which is given as $\sum_{i=1}^k p_{1,i}^{\downarrow} \leq \sum_{i=1}^k p_{2,i}^{\downarrow}$ for all k, where $p_{1,i}^{\downarrow}, p_{2i}^{\downarrow}$ are the values of $\boldsymbol{p}_1, \boldsymbol{p}_2$ in descending order. Majorization compares the mixedness of the two states in a way, therefore the Shannon entropy also reflects the mixedness of the state. Moreover the Shannon entropy is also subadditive, strongly subadditive and extensive. It is increasing in those physical processes (channels) which leave the white noise invariant, and in particular it does not change in processes permuting the indices i.

To describe the distinguishability of two states we have the *relative entropy* (or *Kullback-Leibler divergence*) of the states $\boldsymbol{p}, \boldsymbol{q} \in \Delta$

$$D(\boldsymbol{p}||\boldsymbol{q}) := \sum_{i=1}^d p_i(\ln(p_i) - \ln(q_i)), \tag{3.35}$$

if $p_i = 0$ when $q_i = 0$, and ∞ otherwise. It can be proven that $D(\boldsymbol{p}||\boldsymbol{q}) \geq 0$ and $D(\boldsymbol{p}||\boldsymbol{q}) = 0$ if and only if $\boldsymbol{p} = \boldsymbol{q}$, so the relative entropy takes values in the range $[0, \infty]$. The relative entropy is not symmetric, that is the relative entropy of $\boldsymbol{p}$ with respect to $\boldsymbol{q}$, is not the same as the relative entropy of $\boldsymbol{q}$ with respect to $\boldsymbol{p}$, so the relative entropy is a divergence, not a metric. By *Sanov's theorem*, it describes the distinguishability of the states $\boldsymbol{p}$ and $\boldsymbol{q}$, in the sense that it is the rate of the decaying (with the sample size) of the probability of error in hypothesis testing scenarios. The relative entropy is jointly convex, that is

$$D\Big(\sum_i w_i \boldsymbol{p}_i \Big\| \sum_i w_i \boldsymbol{q}_i\Big) \leq \sum_i w_i D(\boldsymbol{p}_i||\boldsymbol{q}_i). \tag{3.36}$$

It is decreasing in all physical processes (channels). The relative entropy can be compared with the 1-norm distance with the following *Pinsker inequality*

$$D(\boldsymbol{p}||\boldsymbol{q}) \geq \frac{1}{2}\|\boldsymbol{p} - \boldsymbol{q}\|_1^2. \tag{3.37}$$

3.4. **Quantum entropies.** In a quantum system the uncertainty of a state $\rho \in \mathcal{D}$ is given by the *von Neumann entropy*

$$S(\rho) := -\operatorname{Tr}(\rho \ln(\rho)). \tag{3.38}$$

If the eigendecomposition of the state is $\rho = \sum_i \eta_i |\phi_i\rangle\langle\phi_i|$ then the von Neumann entropy is

$$S(\rho) = -\sum_i \eta_i \ln(\eta_i) = S(\boldsymbol{\eta}), \tag{3.39}$$

which is the classical Shannon entropy of the $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_d)$ eigenvalues. It follows then that the von Neumann entropy can take values in the range $[0, \ln(d)]$. It can be shown that the von Neumann entropy takes zero if and only if ρ is a pure state and $\ln(d)$ if and only if it is the white noise. The von Neumann entropy quantifies the quantum information content of a source described by ρ, as a coding theorem can be formulated also in the quantum case, called *Schumacher's noiseless coding theorem*. The von Neumann entropy is *concave*, that is

$$\sum_i w_i S(\rho_i) \leq S\Big(\sum_i w_i \rho_i\Big), \tag{3.40}$$

expressing that the entropy is increasing for information loss. The most important property of the von Neumann entropy is that it is *Schur-concave*, that is

$$\rho_1 \preceq \rho_2 \quad \Longrightarrow \quad S(\rho_1) \geq S(\rho_2), \tag{3.41}$$

where $\rho_1 \preceq \rho_2$ means that ρ_2 majorizes ρ_1, which is defined by the majorization of the spectra, $\boldsymbol{\eta}_1 \preceq \boldsymbol{\eta}_2$. Majorization compares the mixedness of the two states in a way, therefore the von Neumann entropy also reflects the mixedness of the state. Moreover the von Neumann entropy is also subadditive, strongly subadditive and extensive. It is increasing in those physical processes (quantum channels) which leave the white noise invariant, and in particular it does not change in unitary channels, $S(U\rho U^\dagger) = S(\rho)$.

To describe the distinguishability of two states we have the *Umegaki quantum relative entropy (or quantum Kullback-Leibner divergence)* of the states $\rho, \sigma \in \mathcal{D}$

$$D(\rho||\sigma) := \mathrm{Tr}\left(\rho(\ln(\rho) - \ln(\sigma))\right) \tag{3.42}$$

if $\mathrm{Supp}(\rho) \subseteq \mathrm{Supp}(\sigma)$ and ∞ otherwise. It can be proven that $D(\rho||\sigma) \geq 0$ and $D(\rho||\sigma) = 0$ if and only if $\rho = \sigma$, so the quantum relative entropy can take values in the range $[0, \infty]$. The quantum relative entropy is not symmetric, therefore it is not a metric rather a divergence. By the *quantum Sanov's theorem*, it describes the distinguishability of the states ρ and σ, in the sense that it is the rate of the decaying (with the sample size) of the probability of error in hypothesis testing scenarios. The quantum relative entropy is jointly convex, that is

$$D\Big(\sum_i w_i\rho_i \Big\| \sum_i w_i\sigma_i\Big) \leq \sum_i w_i D(\rho_i||\sigma_i). \tag{3.43}$$

It is decreasing in all physical processes (channels) and in particular it does not change in unitary channels, $D(U\rho U^\dagger||U\sigma U^\dagger) = D(\rho||\sigma)$. The quantum relative entropy can be compared to the 1-norm distance of the states with *Pinsker's inequality*, which states that,

$$D(\rho||\sigma) \geq \frac{1}{2}\|\rho - \sigma\|_1^2. \tag{3.44}$$

4. Composite systems and correlations

Composite systems consist of two or more subsystems, which can be of arbitrary size individually. We restrict our attention to the case of two subsystems. Composite systems are described by the use of the tensor product of the spaces describing the subsystems. In this way we can describe the pairs of outcomes of simultaneously measurable observables, also having the linear structure on the whole system.

4.1. **Classical composite systems.**

Let us have two systems, described by the observable algebras $\mathbb{A}_1 := \mathbb{C}^{d_1}$ and $\mathbb{A}_2 := \mathbb{C}^{d_2}$, with observables of the form

$$\boldsymbol{a} = \sum_{i=1}^{d_1} a_i \boldsymbol{\chi}_{1,i} \in \mathbb{A}_1, \quad \boldsymbol{b} = \sum_{j=1}^{d_2} b_j \boldsymbol{\chi}_{2,j} \in \mathbb{A}_2, \tag{4.1}$$

where the characteristic observables satisfy $(\boldsymbol{\chi}_{1,i})_j = \delta_{ij}$ and $(\boldsymbol{\chi}_{2,j})_k = \delta_{jk}$, as before. The corresponding states of the systems are of the form

$$\boldsymbol{p}_1 = \sum_{i=1}^{d_1} p_{1,i}\boldsymbol{\delta}_{1,i} \in \Delta_1 \subset \mathbb{R}^{d_1} \quad \boldsymbol{p}_2 = \sum_{j=1}^{d_2} p_{2,j}\boldsymbol{\delta}_{2,j} \in \Delta_2 \subset \mathbb{R}^{d_2}, \tag{4.2}$$

where the pure states $(\boldsymbol{\delta}_{1,i})_j = \delta_{ij}$, $(\boldsymbol{\delta}_{2,j})_k = \delta_{jk}$, as before. This way the expectation values of the observables of the systems in the particular states are

$$\mathrm{Exp}_{\boldsymbol{p}_1}(\boldsymbol{a}) = (\boldsymbol{p}_1|\boldsymbol{a}) = \sum_{i=1}^{d_1} p_{1,i}a_i, \quad \mathrm{Exp}_{\boldsymbol{p}_2}(\boldsymbol{b}) = (\boldsymbol{p}_2|\boldsymbol{b}) = \sum_{j=1}^{d_2} p_{2,j}b_j, \tag{4.3}$$

as before.

The equations above describe the case when we measure the observables separately in two systems. But what if we measure the product of the variables, which is a joint measurement in the joint system? Let us have the joint observable

$$\boldsymbol{a} \otimes \boldsymbol{b} = \sum_i \sum_j a_i b_j \boldsymbol{\chi}_{1,i} \otimes \boldsymbol{\chi}_{2,j} \in \mathbb{A}_1 \otimes \mathbb{A}_2, \tag{4.4}$$

which is the tensor product of the observables $\boldsymbol{a} \in \mathbb{A}_1$ and $\boldsymbol{b} \in \mathbb{A}_2$. This represents the product observable, taking the $a_i b_j$ values. The characteristic observable $\boldsymbol{\chi}_{12,ij} := \boldsymbol{\chi}_{1,i} \otimes \boldsymbol{\chi}_{2,j}$ is the observable taking the value 1 if the i-th outcome occurs in the measurement of $\boldsymbol{a}$ *and* the j-th outcome occurs in the measurement of $\boldsymbol{b}$ and 0 otherwise. If $p_{12,ij}$ is the *joint probability* of the i-th outcome of the measurement of $\boldsymbol{a}$ and the j-th outcome of the measurement of $\boldsymbol{b}$, then the state of the whole system is

$$\boldsymbol{p}_{12} = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} p_{12,ij} \boldsymbol{\delta}_{12,ij} \in \Delta_{12} \subset \mathbb{R}^{d_1} \otimes \mathbb{R}^{d_2}. \tag{4.5}$$

Again, if the system is described by the pure state $\boldsymbol{\delta}_{12,ij}$, then the i-th outcome of the measurement of $\boldsymbol{a}$ and the j-th outcome of the measurement of $\boldsymbol{b}$ occur with certainty and the other combinations cannot occur. The expectation value of the product observable is

$$\mathrm{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{a} \otimes \boldsymbol{b}) = (\boldsymbol{p}_{12}|\boldsymbol{a} \otimes \boldsymbol{b}) = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} p_{12,ij} a_i b_j, \tag{4.6}$$

as before and the expectation value of the joint characteristic observable $\boldsymbol{\chi}_{12,ij}$ gives the probability of the joint outcome,

$$\mathrm{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{\chi}_{12,ij}) = \sum_{i,j} p_{12,ij} (\boldsymbol{p}_{12}|\boldsymbol{\chi}_{12,ij}) = p_{12,ij}, \tag{4.7}$$

as before.

Note that we do not only have product observables $\boldsymbol{a} \otimes \boldsymbol{b} \in \mathbb{A}_{12} = \mathbb{A}_1 \otimes \mathbb{A}_2$, but any linear combinations of these are meaningful, so a joint observable in general is of the form

$$\boldsymbol{c} = \sum_i \sum_j c_{ij} \boldsymbol{\chi}_{12,ij} \in \mathbb{A}_{12}. \tag{4.8}$$

The expectation value of such an observable is

$$\mathrm{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{c}) = (\boldsymbol{p}_{12}|\boldsymbol{c}) = \sum_{i,j} p_{12,ij} c_{ij}. \tag{4.9}$$

We would like to obtain the state of the subsystems, or *reduced/marginal states* from the state of the whole system. This means that if we have the observables $\boldsymbol{a} \in \mathbb{A}_1$ of subsystem 1 as observables of form $\boldsymbol{a} \otimes \boldsymbol{1} \in \mathbb{A}_{12}$ of the joint system (where $\boldsymbol{1} = (1, 1, \ldots, 1)$), then the reduced state $\mathrm{Red}_2(\boldsymbol{p}_{12})$ is the state giving the expectation value of the observables $\boldsymbol{a} \in \mathbb{A}_1$ of subsystem 1, obtained from the joint state $\boldsymbol{p}_{12}$. That is,

$$\mathrm{Exp}_{\mathrm{Red}_2(\boldsymbol{p}_{12})}(\boldsymbol{a}) := \mathrm{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{a} \otimes \boldsymbol{1}). \tag{4.10}$$

Writing the right-hand side,

$$\mathrm{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{a}\otimes\mathbf{1}) = \sum_{i,j} p_{12,ij}a_i 1 = \sum_i \Big(\sum_j p_{12,ij}\Big)a_i, \tag{4.11}$$

so we have

$$\mathrm{Red}_2(\boldsymbol{p}_{12}) = \Big(\sum_j p_{12,ij}\Big)\boldsymbol{\delta}_{1,i} \in \Delta_1, \tag{4.12}$$

and similarly

$$\mathrm{Red}_1(\boldsymbol{p}_{12}) = \Big(\sum_i p_{12,ij}\Big)\boldsymbol{\delta}_{2,j} \in \Delta_2, \tag{4.13}$$

the components of which are the usual marginal probabilities. It is easy to check that this can also be formulated by the linear functional $\mathrm{Red} : \mathbb{R}^d \to \mathbb{R}$, acting as $\mathrm{Red}(\boldsymbol{p}) := (\mathbf{1}|\boldsymbol{p}) = \sum_i p_i$, then $\mathrm{Red}_2 := I \otimes \mathrm{Red} : \Delta_{12} \to \Delta_1$ and $\mathrm{Red}_1 := \mathrm{Red} \otimes I : \Delta_{12} \to \Delta_2$.

4.2. **Quantum composite systems.** In quantum theory the states and observables are operators acting on Hilbert spaces. Let us have two systems, described by the observable algebras $\mathcal{A}_1 = \mathrm{Lin}(\mathcal{H}_1)$ and $\mathcal{A}_2 = \mathrm{Lin}(\mathcal{H}_2)$, with observables of the form

$$A = \sum_i a_i|\alpha_i\rangle\langle\alpha_i| = \sum_i a_i P_{1,i} \in \mathcal{A}_1, \quad B = \sum_j b_j|\beta_j\rangle\langle\beta_j| = \sum_j b_j P_{2,j} \in \mathcal{A}_2, \tag{4.14}$$

where the characteristic observables are rank-1 projections $P_{1,i} = |\alpha_i\rangle\langle\alpha_i|$ and $P_{2,j} = |\beta_j\rangle\langle\beta_j|$, as before. The corresponding states of the systems are

$$\rho_1 = \sum_i w_i|\psi_i\rangle\langle\psi_i| \in \mathcal{D}_1, \quad \rho_2 = \sum_j w_j|\phi_j\rangle\langle\phi_j| \in \mathcal{D}_2, \tag{4.15}$$

where the pure states are rank-1 projections $|\psi_i\rangle\langle\psi_i|$ and $|\phi_j\rangle\langle\phi_j|$, as before. This way the expectation values of the observables of the systems in the particular states are

$$\mathrm{Exp}_{\rho_1}(A) = \mathrm{Tr}(\rho_1 A), \quad \mathrm{Exp}_{\rho_2}(B) = \mathrm{Tr}(\rho_2 B), \tag{4.16}$$

as before.

The equations above describe the case when we measure the observables separately in two systems. But what if we measure the product of the variables, which is a joint measurement in the joint system? Let us have the joint observable

$$A \otimes B = \sum_{i,j} a_i b_j P_{1,i} \otimes P_{2,j} \in \mathcal{A}_1 \otimes \mathcal{A}_2, \tag{4.17}$$

which is the tensor product of the observables $A \in \mathcal{A}_1$ and $B \in \mathcal{A}_2$. This represents the product observable, taking the $a_i b_j$ values. The characteristic observable $P_{12,ij} := P_{1,i} \otimes P_{2,j}$ is the observable taking the value 1 if the i-th outcome occurs in the measurement of A *and* the j-th outcome occurs in the measurement of B and 0 otherwise. The *state of the whole system* is

$$\rho_{12} = \sum_i w_i|\psi_{12,i}\rangle\langle\psi_{12,i}| \in \mathcal{D}_{12} \subset \mathcal{A}_1 \otimes \mathcal{A}_2. \tag{4.18}$$

Again, if the system is described by the pure state $|\psi\rangle\langle\psi| = P_{12,ij}$, then the i-th outcome of the measurement of A and the j-th outcome of the measurement of B occur with certainty, and the other combinations cannot occur. Note however that there are other pure states too, for which this does not hold. The expectation value of the product observable is

$$\mathrm{Exp}_{\rho_{12}}(A \otimes B) = (\rho_{12}|A \otimes B) = \sum_{k,i,j} w_k a_i b_j |\langle\psi_{12,k}|\alpha_i \otimes \beta_j\rangle|^2, \tag{4.19}$$

as before, and the expectation value of the joint characteristic observable $P_{12,ij}$ gives the probability of the joint outcome,

$$\mathrm{Exp}_{\rho_{12}}(P_{12,ij}) = \sum_k w_k |\langle \psi_{12,k} | \alpha_i \otimes \beta_j \rangle|^2 = p_{12,ij}, \tag{4.20}$$

as before.

Note that we have not only product observables $A \otimes B \in \mathcal{A}_{12} = \mathcal{A}_1 \otimes \mathcal{A}_2$, but any linear combinations of these are meaningful, so a joint observable in general is of the form

$$C = \sum_i c_i Q_{12,i} \in \mathbb{A}_{12}, \tag{4.21}$$

where $Q_{12,i} = |\gamma_{12,i}\rangle\langle\gamma_{12,i}|$ are joint characteristic observables. Note that these are not necessarily of a product form. The expectation value of such an observable is

$$\mathrm{Exp}_{\rho_{12}}(C) = (\rho_{12}|C) = \sum_i q_i c_i, \tag{4.22}$$

where $q_i = (\rho_{12}|Q_{12,i})$ is the measurement statistics.

We would like to obtain the state of the subsystems, or *reduced/marginal states* from the state of the whole system. This means that if we have the observables $A \in \mathcal{A}_1$ of subsystem 1 as observables of form $A \otimes I \in \mathcal{A}_{12}$ of the joint system (where I is the identity operator), then the reduced state $\mathrm{Red}_2(\rho_{12})$ is the state giving the expectation value of the observables $A \in \mathcal{A}_1$ of subsystem 1, obtained from the joint state ρ_{12}. That is,

$$\mathrm{Exp}_{\mathrm{Red}_2(\rho_{12})}(A) := \mathrm{Exp}_{\rho_{12}}(A \otimes I). \tag{4.23}$$

It turns out that the reduction map is given by the partial trace,

$$\mathrm{Red}_2(\rho_{12}) = \mathrm{Tr}_2(\rho_{12}) \in \mathcal{D}_1, \tag{4.24}$$

and similarly

$$\mathrm{Red}_1(\rho_{12}) = \mathrm{Tr}_1(\rho_{12}) \in \mathcal{D}_2, \tag{4.25}$$

see in equation (2.9).

Again see the example of the qubit, for 2 qubit systems we use the Pauli basis on each qubits. This way we can express a product state as

$$\rho_1 \otimes \rho_2 = \frac{1}{2}\big(I + \boldsymbol{r}\boldsymbol{\sigma}\big) \otimes \frac{1}{2}\big(I + \boldsymbol{s}\boldsymbol{\sigma}\big) = \frac{1}{4}\big(I \otimes I + \boldsymbol{r}\boldsymbol{\sigma} \otimes I + I \otimes \boldsymbol{s}\boldsymbol{\sigma} + \boldsymbol{r}\boldsymbol{\sigma} \otimes \boldsymbol{s}\boldsymbol{\sigma}\big), \tag{4.26}$$

where $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^3$ are the Bloch vectors on the subsystems. Now consider a general state, it can be written as

$$\rho_{12} = \frac{1}{4}\big(I \otimes I + \boldsymbol{r}\boldsymbol{\sigma} \otimes I + I \otimes \boldsymbol{s}\boldsymbol{\sigma} + \boldsymbol{t}(\boldsymbol{\sigma} \otimes \boldsymbol{\sigma})\big), \tag{4.27}$$

where $\boldsymbol{t} \in \mathbb{R}^3 \otimes \mathbb{R}^3$ a matrix, and we use the shorthand notation $\boldsymbol{t}(\boldsymbol{\sigma} \otimes \boldsymbol{\sigma}) = \sum_{ij} t_{ij} \sigma_i \otimes \sigma_j$. Expressing ρ_{12} in this form is useful because one can clearly see the parameters $\boldsymbol{r}$, $\boldsymbol{s}$ of the two subsystems, this way the reduced states are seen instantly, as $\mathrm{Tr}_2(\rho_{12}) = \frac{1}{2}\big(I + \boldsymbol{r}\boldsymbol{\sigma}\big)$ and $\mathrm{Tr}_1(\rho_{12}) = \frac{1}{2}\big(I + \boldsymbol{s}\boldsymbol{\sigma}\big)$. The matrix $\boldsymbol{t}$ contains the nonlocal information of the quantum state.

4.3. **Classical correlation.** The *two observables* $\boldsymbol{a} \in \mathcal{A}_1$ and $\boldsymbol{b} \in \mathcal{A}_2$ are *uncorrelated* in the system described by the state $\boldsymbol{p}_{12} \in \Delta_{12}$ if

$$\mathrm{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{a} \otimes \boldsymbol{b}) = \mathrm{Exp}_{\boldsymbol{p}_1}(\boldsymbol{a}) \, \mathrm{Exp}_{\boldsymbol{p}_2}(\boldsymbol{b}) \tag{4.28}$$

holds.

The *state* $\boldsymbol{p}_{12} \in \Delta_{12}$ is *uncorrelated* if any pair of observables are uncorrelated, that is,

$$\mathrm{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{a} \otimes \boldsymbol{b}) = \mathrm{Exp}_{\boldsymbol{p}_1}(\boldsymbol{a}) \, \mathrm{Exp}_{\boldsymbol{p}_2}(\boldsymbol{b}) \qquad \text{for all } \boldsymbol{a} \in \mathbb{A}_1, \boldsymbol{b} \in \mathbb{A}_2 \tag{4.29}$$

This is equivalent to that the state is a product state

$$\boldsymbol{p}_{12} = \boldsymbol{p}_1 \otimes \boldsymbol{p}_2, \tag{4.30}$$

which we will show in section 5.1. That is the state must be the product of its reduced states. (In usual probability theory in this case it is said that the two probabilistic variables are independent. Here we take a more general point of view, as the observables are not fixed.) The components of the product state are the products of the components of the reduced states: $p_{12,ij} = p_{1,i} p_{2,j}$. It easily follows that all pure states are uncorrelated, because any pure state of the composite system is a product of pure states of the subsystems, $(\boldsymbol{\delta}_{12,ij})_{kl} = \delta_{ik}\delta_{jl} = (\boldsymbol{\delta}_{1,i})_k (\boldsymbol{\delta}_{2,j})_l$. Let us have the state space of uncorrelated states

$$\Delta_{\mathrm{unc}} := \left\{ \boldsymbol{p}_1 \otimes \boldsymbol{p}_2 \;\middle|\; \boldsymbol{p}_1 \in \Delta_1, \boldsymbol{p}_2 \in \Delta_2 \right\} \subset \Delta_{12}. \tag{4.31}$$

Moreover

$$\Delta_{12} = \mathrm{Conv}(\{\boldsymbol{\delta}_{12,ij} \mid i = 1, 2, \ldots, d_1; j = 1, 2, \ldots, d_2\}) = \mathrm{Conv}(\Delta_{\mathrm{unc}}), \tag{4.32}$$

so any state can be prepared by the statistical mixture (convex combination) of uncorrelated (tensor product) states.

4.4. **Quantum correlation.** The definition of *uncorrelated observables* is formally the same as in the classical case. The *two observables* $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$ are *uncorrelated* in the system described by the state $\rho_{12} \in \mathcal{D}_{12}$ if

$$\mathrm{Exp}_{\rho_{12}}(A \otimes B) = \mathrm{Exp}_{\rho_1}(A) \, \mathrm{Exp}_{\rho_2}(B), \tag{4.33}$$

holds.

The *state* $\rho_{12} \in \mathcal{D}_{12}$ is *uncorrelated*, if any pair of observables are uncorrelated, that is,

$$\mathrm{Exp}_{\rho_{12}}(A \otimes B) = \mathrm{Exp}_{\rho_1}(A) \, \mathrm{Exp}_{\rho_2}(B) \qquad \text{for all } A \in \mathcal{A}_1, \, B \in \mathcal{A}_2. \tag{4.34}$$

This is equivalent to that the state is a product state

$$\rho_{12} = \rho_1 \otimes \rho_2, \tag{4.35}$$

which we will show in section 5.2. That is the state must be the product of its reduced states. The components of the product state are the products of the components of the reduced states: $\rho_{12,ii'jj'} = \rho_{1,ii'} \rho_{2,jj'}$. It is important to note that, contrary to the classical case, pure states are not always uncorrelated. This is the notion of entanglement, see in the next subsection. (An example for a non-product pure state is the Bell state, given by the vector $(|00\rangle + |11\rangle)/\sqrt{2}$.) Let us have the state space of uncorrelated states

$$\mathcal{D}_{\mathrm{unc}} := \left\{ \rho_1 \otimes \rho_2 \;\middle|\; \rho_1 \in \mathcal{D}_1, \rho_2 \in \mathcal{D}_2 \right\} \subset \mathcal{D}_{12}. \tag{4.36}$$

Note that, since we have pure states which are correlated,

$$\mathcal{D}_{12} = \mathrm{Conv}(\mathcal{P}_{12}) \supsetneq \mathrm{Conv}(\mathcal{D}_{\mathrm{unc}}), \tag{4.37}$$

which, again, points towards entanglement.

4.5. **Quantum entanglement.** In the quantum case there is another important category besides uncorrelated states. We have seen in equation (4.37) that, since there are correlated pure states, so there are (correlated) states which cannot be prepared by mixing uncorrelated states. This is the definition of entanglement [5]. First, let us have the space of *separable states*, being the convex combination of uncorrelated states,

$$\mathcal{D}_{\text{sep}} := \text{Conv}(\mathcal{D}_{\text{unc}}) = \Big\{ \sum_k w_k \rho_{1,k} \otimes \rho_{2,k} \;\Big|\; w_k \geq 0, \sum_k w_k = 1, \rho_{1,k} \in \mathcal{D}_1, \rho_{2,k} \in \mathcal{D}_2 \Big\} \subset \mathcal{D}_{12}. \tag{4.38}$$

These are the states which can be prepared locally with classical communication, where the latter establishes the statistical mixing. The states which are not separable are called *entangled*. To prepare those, some quantum communication is needed.

4.6. **Classical correlation measures.** The *covariance of the observables* $\boldsymbol{a} \in \mathbb{A}_1$ and $\boldsymbol{b} \in \mathbb{A}_2$ of a system described by the state $\boldsymbol{p}_{12} \in \Delta_{12}$ is

$$\text{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a}, \boldsymbol{b}) := \text{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{a} \otimes \boldsymbol{b}) - \text{Exp}_{\boldsymbol{p}_1}(\boldsymbol{a}) \, \text{Exp}_{\boldsymbol{p}_2}(\boldsymbol{b}). \tag{4.39}$$

This can also be expressed as

$$\text{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a}, \boldsymbol{b}) = (\boldsymbol{g}_{12} | \boldsymbol{a} \otimes \boldsymbol{b}) \tag{4.40}$$

where $\boldsymbol{g}_{12} := \boldsymbol{p}_{12} - \boldsymbol{p}_1 \otimes \boldsymbol{p}_2$ encodes the correlation in the state $\boldsymbol{p}_{12} \in \Delta_{12}$. Indeed,

$$\text{Exp}_{\boldsymbol{p}_{12}}(\boldsymbol{a} \otimes \boldsymbol{b}) - \text{Exp}_{\boldsymbol{p}_1}(\boldsymbol{a}) \, \text{Exp}_{\boldsymbol{p}_2}(\boldsymbol{b}) = (\boldsymbol{p}_{12} | \boldsymbol{a} \otimes \boldsymbol{b}) - (\boldsymbol{p}_1 | \boldsymbol{a})(\boldsymbol{p}_2 | \boldsymbol{b}) = (\boldsymbol{p}_{12} - \boldsymbol{p}_1 \otimes \boldsymbol{p}_2 | \boldsymbol{a} \otimes \boldsymbol{b}),$$

where we have used the properties of the inner product.

The *correlation of the observables* $\boldsymbol{a} \in \mathbb{A}_1$ and $\boldsymbol{b} \in \mathbb{A}_2$ of a system described by the state $\boldsymbol{p}_{12} \in \Delta_{12}$ is

$$\text{Corr}_{\boldsymbol{p}_{12}}(\boldsymbol{a}, \boldsymbol{b}) := \frac{\text{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a}, \boldsymbol{b})}{\sqrt{\text{Var}_{\boldsymbol{p}_1}(\boldsymbol{a}) \, \text{Var}_{\boldsymbol{p}_2}(\boldsymbol{b})}}. \tag{4.41}$$

This is a normalized variant of the covariance, $-1 \leq \text{Corr}_{\boldsymbol{p}_{12}}(\boldsymbol{a}, \boldsymbol{b}) \leq 1$. However in our case it is better to just stick to using covariance as it is linear in its variables while correlation is not. In addition, this kind of normalization is not useful at all.

We can also quantify the *correlation of the state* itself, without respect to the observables. The *relative entropy of correlation*, also called *correlation (of the state itself)* of a system described by the state $\boldsymbol{p}_{12} \in \Delta_{12}$ is

$$C(\boldsymbol{p}_{12}) := \min_{\boldsymbol{q}_{12} \in \Delta_{\text{unc}}} D(\boldsymbol{p}_{12} || \boldsymbol{q}_{12}), \tag{4.42}$$

where $D(\boldsymbol{p}_{12} || \boldsymbol{q}_{12})$ is the relative entropy, given in equation (3.35). It expresses how distinguishable the state is from the uncorrelated ones, in terms of the relative entropy. This is a *correlation measure* in the sense that it is nonincreasing with respect to local operations, reflecting that correlation cannot be created nor strengthened locally. (This follows from that the relative entropy is decreasing in all channels and the local operations map the set of uncorrelated states onto itself.) It is a *faithful correlation measure*, that is,

$$\boldsymbol{p}_{12} \in \Delta_{\text{unc}} \quad \Longleftrightarrow \quad C(\boldsymbol{p}_{12}) = 0. \tag{4.43}$$

(This follows from the properties of the relative entropy.) For pure states C is zero, since pure states are always uncorrelated in the classical case.

It turns out that the minimization in the formula (4.42) can be done explicitly. One can prove that the minimum is taken at $\boldsymbol{q}_{12} = \boldsymbol{p}_1 \otimes \boldsymbol{p}_2$, see [2], leading to that

$$C(\boldsymbol{p}_{12}) = D(\boldsymbol{p}_{12} || \boldsymbol{p}_1 \otimes \boldsymbol{p}_2) = S(\boldsymbol{p}_1) + S(\boldsymbol{p}_2) - S(\boldsymbol{p}_{12}). \tag{4.44}$$

Note that this is also called *mutual information* in a different context.

The correlation of the state is lower bounded by the covariance of any pair of (properly normalized) observables [7] as

$$\frac{1}{2}\Big|\operatorname{Cov}_{\boldsymbol{p}_{12}}\Big(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_\infty},\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_\infty}\Big)\Big|^2 \leq C(\boldsymbol{p}_{12}), \qquad \text{for all } \boldsymbol{a}\in\mathbb{A}_1,\boldsymbol{b}\in\mathbb{A}_2, \tag{4.45}$$

which we will show in section 5.1.

4.7. **Quantum correlation measures.** The *covariance* of two observables $A\in\mathcal{A}_1$, $B\in\mathcal{A}_2$ of a system described by the state $\rho_{12}\in\mathcal{D}_{12}$ is

$$\operatorname{Cov}_{\rho_{12}}(A,B) := \operatorname{Exp}_{\rho_{12}}(A\otimes B) - \operatorname{Exp}_{\rho_1}(A)\operatorname{Exp}_{\rho_2}(B). \tag{4.46}$$

This can also be expressed as

$$\operatorname{Cov}_{\rho_{12}}(A,B) = (\Gamma_{12}|A\otimes B), \tag{4.47}$$

where $\Gamma_{12}=\rho_{12}-\rho_1\otimes\rho_2$ encodes the correlation in the state $\rho_{12}\in\mathcal{D}_{12}$. Indeed,

$$\operatorname{Exp}_{\rho_{12}}(A\otimes B) - \operatorname{Exp}_{\rho_1}(A)\operatorname{Exp}_{\rho_2}(B) = (\rho_{12}|A\otimes B) - (\rho_1|A)(\rho_2|B) = (\rho_{12}-\rho_1\otimes\rho_2|A\otimes B),$$

where we have used the properties of the inner product.

The *correlation of the observables* $A\in\mathcal{A}_1$, $B\in\mathcal{A}_2$ of a system described by the state $\rho_{12}\in\mathcal{D}_{12}$ is

$$\operatorname{Corr}_{\rho_{12}}(A,B) := \frac{\operatorname{Cov}_{\rho_{12}}(A,B)}{\sqrt{\operatorname{Var}_{\rho_1}(A)\operatorname{Var}_{\rho_2}(B)}}. \tag{4.48}$$

This is a normalized variant of the covariance, $-1\leq\operatorname{Corr}_{\rho_{12}}(A,B)\leq 1$. However again sticking to covariance as a measure of correlation is more convenient, because it is linear in the observables $A\in\mathcal{A}_1$, $B\in\mathcal{A}_2$.

We can also quantify the *correlation of the state itself*, without respect to the observables. The *relative entropy of correlation*, also called *correlation (of the state itself)* of a system described by the state $\rho_{12}\in\mathcal{D}_{12}$ is

$$C(\rho_{12}) := \min_{\tau_{12}\in\mathcal{D}_{\text{unc}}} D(\rho_{12}||\tau_{12}), \tag{4.49}$$

where $D(\rho_{12}||\tau_{12})$ is the relative entropy as in equation (3.42). It expresses how distinguishable the state is from the uncorrelated ones, in terms of the relative entropy. This is a *correlation measure* in the sense that it is nonincreasing with respect to local operations, reflecting that correlation cannot be created nor strengthened locally. (This follows from that the quantum relative entropy is decreasing in all quantum channels and the local operations map the set of uncorrelated states onto itself.) It is a *faithful correlation measure*, that is,

$$\rho_{12}\in\mathcal{D}_{\text{unc}} \quad\Longleftrightarrow\quad C(\rho_{12})=0. \tag{4.50}$$

(This follows from the properties of the relative entropy.) Note that C is not necessarily zero for pure states, since pure states can also be correlated in the quantum case.

It turns out that the minimization in the formula (4.49) can be done explicitly. One can prove that the minimum is taken at $\rho_{12}=\rho_1\otimes\rho_2$, see [2],

$$C(\rho_{12}) = \min_{\tau_{12}\in\mathcal{D}_{unc}} D(\rho_{12}||\tau_{12}) = D(\rho_{12}||\rho_1\otimes\rho_2) = S(\rho_1)+S(\rho_2)-S(\rho_{12}). \tag{4.51}$$

Note that this is also called *quantum mutual information* in a different context. The correlation in pure states $\pi_{12}=|\psi_{12}\rangle\langle\psi_{12}|$ is

$$C(\pi_{12}) = 2S(\pi_1) = 2S(\pi_2), \tag{4.52}$$

two times the entropy of the reduced state, since $S(\pi)=0$, and the reduced states have the same nonzero eigenvalues.

The correlation of the state is lower bounded by the covariance of any pair of (properly normalized) observables [7] as

$$\frac{1}{2}\Big|\operatorname{Cov}_{\rho_{12}}\Big(\frac{A}{\|A\|_\infty},\frac{B}{\|B\|_\infty}\Big)\Big|^2 \le C(\rho_{12}), \qquad \text{for all } A\in\mathcal{A}_1, B\in\mathcal{A}_2, \tag{4.53}$$

which we will show in section 5.2.

4.8. **Quantum entanglement measures.** We may have a similar construction for measuring entanglement as in the previous subsection for correlation. The *relative entropy of entanglement* of a system described by the state $\rho_{12}\in\mathcal{D}_{12}$ is

$$E(\rho_{12}) := \min_{\tau_{12}\in\mathcal{D}_{\text{sep}}} D(\rho_{12}||\tau_{12}), \tag{4.54}$$

where $D(\rho_{12}||\tau_{12})$ is again the quantum relative entropy as in equation (3.42). It expresses how distinguishable the state is from the separable ones, in terms of the quantum relative entropy. This is an *entanglement measure* in the sense that it is nonincreasing with respect to local operations and classical communications, reflecting that entanglement cannot be created nor strengthened locally with classical communication. (This follows from that the quantum relative entropy is decreasing in all quantum channels and the local operations and classical communications map the set of separable states onto itself.) It is a *faithful entanglement measure*, that is,

$$\rho_{12}\in\mathcal{D}_{\text{sep}} \quad\Longleftrightarrow\quad E(\rho_{12})=0 \tag{4.55}$$

(This follows from the properties of the quantum relative entropy.)

Contrary to the correlations, the minimization in the formula (4.54) cannot be done explicitly. However, it is proven [4] that the relative entropy of entanglement in pure states $\pi_{12}=|\psi\rangle\langle\psi|$ is

$$E(\pi_{12}) = S(\pi_1) = S(\pi_2), \tag{4.56}$$

the entropy of the reduced state. This is also called *entanglement entropy* in a different context, being the asymptotic ratio n/m of the encoding (by local operation and classical communication) of m copies of the pure state into n copies of Bell-states.

5. Covariance versus correlation

We would like to explore the specific requirements for the covariance to be zero and how this relates to the state being uncorrelated. In two-bit and two-qubit systems this can be explicitly calculated.

5.1. **Covariance and correlation in classical systems.** Here we prove some properties stated previously in the classical case.

First we show that the uncorrelated states (equation 4.29) are exactly the product ones (equation 4.30), that is

$$\operatorname{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a},\boldsymbol{b})=0, \forall \boldsymbol{a}\in\mathbb{A}_1, \boldsymbol{b}\in\mathbb{A}_2 \quad\Longleftrightarrow\quad \boldsymbol{p}_{12}\in\Delta_{\text{unc}}. \tag{5.1}$$

To see the *"if" direction*, we have that if $\boldsymbol{p}_{12}\in\Delta_{\text{unc}}$ then $\boldsymbol{p}_{12}=\boldsymbol{p}_1\otimes\boldsymbol{p}_2$ and so

$$\operatorname{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a},\boldsymbol{b}) = (\boldsymbol{p}_{12}-\boldsymbol{p}_1\otimes\boldsymbol{p}_2|\boldsymbol{a}\otimes\boldsymbol{b}) = (\boldsymbol{0}|\boldsymbol{a}\otimes\boldsymbol{b}) = 0,$$

which is true for any pair of observables $\boldsymbol{a},\boldsymbol{b}$. To see the *"only if" direction*, we have that if $\operatorname{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a},\boldsymbol{b})=0$ then

$$\operatorname{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a},\boldsymbol{b}) = (\boldsymbol{p}_{12}-\boldsymbol{p}_1\otimes\boldsymbol{p}_2|\boldsymbol{a}\otimes\boldsymbol{b}) = 0, \forall \boldsymbol{a}\in\mathbb{A}_1, \boldsymbol{b}\in\mathbb{A}_2$$

means that $\boldsymbol{p}_{12}-\boldsymbol{p}_1\otimes\boldsymbol{p}_2=\boldsymbol{0}$, because the elementary tensors $\boldsymbol{a}\otimes\boldsymbol{b}$ span the observable algebra of the composite system, and a vector is zero if and only if its inner product with all vectors is zero.

Second, we show the inequality (4.45) [7]. It follows from

$$\begin{aligned} |\operatorname{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a},\boldsymbol{b})| &= |(\boldsymbol{p}_{12}-\boldsymbol{p}_1\otimes\boldsymbol{p}_2|\boldsymbol{a}\otimes\boldsymbol{b})| \\ &\leq \|\boldsymbol{p}_{12}-\boldsymbol{p}_1\otimes\boldsymbol{p}_2\|_1\|\boldsymbol{a}\otimes\boldsymbol{b}\|_\infty \\ &\leq \sqrt{2D(\boldsymbol{p}_{12}||\boldsymbol{p}_1\otimes\boldsymbol{p}_2)}\|\boldsymbol{a}\|_\infty\|\boldsymbol{b}\|_\infty, \end{aligned}$$

where the Hölder inequality (2.13) for $p=1$ and $q=\infty$ and the Pinsker inequality (3.37) were used. After rearranging the terms, and using the second form of the correlation in equation (4.44) we have $\frac{1}{2}\big|\operatorname{Cov}_{\boldsymbol{p}_{12}}\big(\frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_\infty},\frac{\boldsymbol{b}}{\|\boldsymbol{b}\|_\infty}\big)\big|^2 \leq C(\boldsymbol{p}_{12})$, with the $\frac{\boldsymbol{a}}{||\boldsymbol{a}||_\infty}$, $\frac{\boldsymbol{b}}{||\boldsymbol{b}||_\infty}$ normalized observables. The values of such observables lie inside the unit circle.

5.2. **Covariance and correlation in quantum systems.** Here we prove some properties stated previously in the quantum case.

We show that the uncorrelated states (equation (4.34)) are exactly the product ones (equation (4.35)), that is

$$\operatorname{Cov}_{\rho_{12}}(A,B)=0 \quad\Longleftrightarrow\quad \rho\in\mathcal{D}_{\text{unc}}. \tag{5.2}$$

To see the *"if" direction*, we have that if $\rho_{12}\in\mathcal{D}_{\text{unc}}$ then $\rho_{12}=\rho_1\otimes\rho_2$ and so

$$\operatorname{Cov}_{\rho_{12}}(A,B)=(\rho_{12}-\rho_1\otimes\rho_2|A\otimes B)=(0|A\otimes B)=0$$

which is true for any pair of A,B observables. To see the *"only if" direction*, we have that if $\operatorname{Cov}_{\rho_{12}}(A,B)=0$ then

$$\operatorname{Cov}_{\rho_{12}}(A,B)=(\rho_{12}-\rho_1\otimes\rho_2|A\otimes B)=0, \forall A\in\mathcal{A}_1, B\in\mathcal{A}_2$$

means that $\rho_{12}-\rho_1\otimes\rho_2=0$, because the elementary tensors $A\otimes B$ span the observable algebra of the composite system, and a vector is zero if and only if its inner product with all vectors is zero.

Next we show the inequality (4.53) [7]. We have

$$\begin{aligned} |\operatorname{Cov}_{\rho_{12}}(A,B)| &= |(\rho_{12}-\rho_1\otimes\rho_2|A\otimes B)| \\ &\leq \|\rho_{12}-\rho_1\otimes\rho_2\|_1\|A\otimes B\|_\infty \\ &\leq \sqrt{2D(\rho_{12}||\rho_1\otimes\rho_2)}\|A\|_\infty\|B\|_\infty, \end{aligned}$$

where the Hölder inequality (2.13) for $p=1$ and $q=\infty$ and the Pinsker inequality (3.44) were used. After rearranging the terms, and using the second form of the correlation in equation (4.51), we have $\big|\operatorname{Cov}_{\rho_{12}}\big(\frac{A}{\|A\|_\infty},\frac{B}{\|B\|_\infty}\big)\big|^2 \leq C(\rho_{12})$, with the $\frac{A}{\|A\|_\infty}$, $\frac{B}{\|B\|_\infty}$ normalized observables. The values of such observables lie inside the unit circle.

5.3. **Two-bit systems.** For 2 bits we will prove that zero covariance on one pair of nontrivial observables (taking two different values) guarantees that the state is uncorrelated. So let us restrict ourselves to observables $\boldsymbol{a}\in\mathbb{A}_1$, $\boldsymbol{b}\in\mathbb{A}_2$ which are *nontrivial*, which means $\boldsymbol{a},\boldsymbol{b}\not\sim\mathbf{1}$. For any pair of such observables, for $d_1,d_2=2$ we have

$$\operatorname{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a},\boldsymbol{b})=0 \quad\Longleftrightarrow\quad \boldsymbol{p}_{12}\in\Delta_{\text{unc}}. \tag{5.3}$$

Note that the $\Longleftarrow$ implication is obvious by the result (5.1), the extra point here is the $\Longrightarrow$ implication. We can prove this by the use of the classical Pauli basis given in equation (3.10). That is the states of the two bits take the form

$$\boldsymbol{p}_1=\frac{1}{2}(\mathbf{1}+r\boldsymbol{\sigma})=\frac{1}{2}(1+r,1-r), \quad \boldsymbol{p}_2=\frac{1}{2}(\mathbf{1}+s\boldsymbol{\sigma})=\frac{1}{2}(1+s,1-s). \tag{5.4}$$

Similarly, the general form of a state of the whole 2 bit system $\boldsymbol{p}_{12}$ is

$$\begin{aligned}\boldsymbol{p}_{12} &= \frac{1}{4}\big(\mathbf{1}\otimes\mathbf{1} + r\boldsymbol{\sigma}\otimes\mathbf{1} + s\mathbf{1}\otimes\boldsymbol{\sigma} + t\boldsymbol{\sigma}\otimes\boldsymbol{\sigma}\big)\\ &= \frac{1}{4}\big((1,1,1,1) + r(1,1,-1,-1) + s(1,-1,1,-1) + t(1,-1,-1,1)\big)\\ &= \frac{1}{4}\big(1+r+s+t, 1+r-s-t, 1-r+s-t, 1-r-s+t\big).\end{aligned} \tag{5.5}$$

In the end we want to get the covariance, so we have to compute $\boldsymbol{g}_{12} = \boldsymbol{p}_{12} - \boldsymbol{p}_1\otimes\boldsymbol{p}_2$ and $\boldsymbol{a}\otimes\boldsymbol{b}$. For $\boldsymbol{p}_1\otimes\boldsymbol{p}_2$ we have to take the product of $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ elementwise

$$\begin{aligned}\boldsymbol{p}_1\otimes\boldsymbol{p}_2 &= \frac{1}{4}\big(\mathbf{1}\otimes\mathbf{1} + r\boldsymbol{\sigma}\otimes\mathbf{1} + s\mathbf{1}\otimes\boldsymbol{\sigma} + rs\boldsymbol{\sigma}\otimes\boldsymbol{\sigma}\big)\\ &= \frac{1}{4}\big((1+r)(1+s), (1+r)(1-s), (1-r)(1+s), (1-r)(1-s)\big)\\ &= \frac{1}{4}(1+r+s+rs, 1+r-s-rs, 1-r+s-rs, 1-r-s+rs).\end{aligned} \tag{5.6}$$

Now to get $\boldsymbol{g}_{12}$ we substitute $\boldsymbol{p}_{12}$ and $\boldsymbol{p}_1\otimes\boldsymbol{p}_2$

$$\boldsymbol{g}_{12} = \boldsymbol{p}_{12} - \boldsymbol{p}_1\otimes\boldsymbol{p}_2 = \frac{1}{4}(t-rs)\boldsymbol{\sigma}\otimes\boldsymbol{\sigma} = \frac{t-rs}{4}(1,-1,-1,1). \tag{5.7}$$

To get $\boldsymbol{a}\otimes\boldsymbol{b}$ we again have to multiply elementwise

$$\boldsymbol{a}\otimes\boldsymbol{b} = (a_1,a_2)\otimes(b_1,b_2) = (a_1b_1, a_1b_2, a_2b_1, a_2b_2). \tag{5.8}$$

Now finally the covariance is given as

$$\mathrm{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a}\otimes\boldsymbol{b}) = (\boldsymbol{g}_{12}|\boldsymbol{a}\otimes\boldsymbol{b}) = \frac{t-rs}{4}(\boldsymbol{\sigma}|\boldsymbol{a})(\boldsymbol{\sigma}|\boldsymbol{b}) = \frac{t-rs}{4}(b_1-b_2)(a_1-a_2) \overset{!}{=} 0. \tag{5.9}$$

From this result we can see that either any of the observables is trivial (that is, $a_1 = a_2$ or $b_1 = b_2$) or $t - rs = 0$. So the (5.9) condition can only be satisfied by $t = rs$, which holds if and only if the state is uncorrelated (a product) (see equation (5.7)).

We can visualize the state space of a 2 bit system with the $\big(\frac{\mathbf{1}}{\sqrt{2}}, \frac{\boldsymbol{\sigma}}{\sqrt{2}}\big)$ basis. First we have to express r, s, t with the p_i probabilities. We have two expressions for $\boldsymbol{p}_{12}$, $\boldsymbol{p}_{12} = (p_0, p_1, p_2, p_3)$ and $\boldsymbol{p}_{12} = \frac{1}{4}(1+r+s+t, 1+r-s-t, 1-r+s-t, 1-r-s+t)$. This way the r, s, t parameters are

$$r = 2(p_0+p_1) - 1, \quad s = 2(p_0+p_2) - 1, \quad t = 2(p_0+p_3) - 1. \tag{5.10}$$

When considering the pure states in p_i ($\boldsymbol{\delta}_1 = (1,0,0,0)$ and so on) we get the extremal points of this 3-dimensional simplex in r, s, t.

5.4. **Bit-trit systems.** Next we want to check if (5.3) can be true for larger systems. The most simple case is when $d_1 = 3$, $d_2 = 2$. Can we find a correlated state with nontrivial observables of zero covariance? Let us have the ansatz for the state

$$\boldsymbol{p}_{12} = (0, x, y, 0, 0, z), \tag{5.11}$$

$$\boldsymbol{p}_1 = (x, y, z), \quad \boldsymbol{p}_2 = (y, x+z) \tag{5.12}$$

with the free variables $0 \le x, y, z \in \mathbb{R}$ such that $x + y + z = 1$. Taking the tensor product of the reduced states we get

$$\boldsymbol{p}_1\otimes\boldsymbol{p}_2 = \big(xy, x(x+z), y^2, y(x+z), yz, z(x+z)\big). \tag{5.13}$$

The observables and their tensor product in general are

$$\boldsymbol{a} = (a_1, a_2, a_3), \quad \boldsymbol{b} = (b_1, b_2), \tag{5.14}$$

$$\boldsymbol{a} \otimes \boldsymbol{b} = (a_1b_1, a_1b_2, a_2b_1, a_2b_2, a_3b_1, a_3b_2). \tag{5.15}$$

We need to calculate $\boldsymbol{g}_{12}$, that is

$$\begin{aligned} \boldsymbol{g}_{12} &= \boldsymbol{p}_{12} - \boldsymbol{p}_1 \otimes \boldsymbol{p}_2 \\ &= \big(-xy, x(1-x-z), y(1-y), -y(x+z), -yz, z(1-x-z)\big) \\ &= \big(-xy, xy, y(x+z), -y(x+z), -yz, yz\big). \end{aligned} \tag{5.16}$$

Here we also see that $\boldsymbol{p}_{12}$ can only be uncorrelated if either $y = 0$ or $x = z = 0$. The question is, wether the covariance could be zero for correlated states. The covariance is

$$\begin{aligned} \mathrm{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a}, \boldsymbol{b}) &= (\boldsymbol{g}_{12} | \boldsymbol{a} \otimes \boldsymbol{b}) \\ &= -a_1b_1xy - a_1b_2xy + a_2b_1y(x+z) - a_2b_2y(x+z) - a_3b_1yz + a_3b_2yz \\ &= y\big(x(-a_1b_1 + a_1b_2 + a_2b_1 - a_2b_2) + z(a_2b_1 - a_2b_2 - a_3b_1 + a_3b_2)\big) \\ &= y(b_2 - b_1)\big(x(a_1 - a_2) + z(a_3 - a_2)\big), \end{aligned} \tag{5.17}$$

we can again see that a nontrivial $\boldsymbol{b}$ can not lead to a zero result. However $x(a_1 - a_2) + z(a_3 - a_2)$ can be zero for nontrivial 3 valued observables. Thus even for the system slightly larger than two bits equation (5.3) does not hold anymore. To see this through a more concrete example, let us have $\boldsymbol{a} = (1, 0, -1)$ and $\boldsymbol{b} = (1, -1)$ similar to a spin measurement. In this case if the state is of the form $\boldsymbol{p}_{12} = (0, x, y, 0, 0, x)$, then the covariance always vanishes even though $\boldsymbol{p}_{12}$ is correlated (which is $x, y \neq 0$).

As smaller systems can be embedded into larger ones, we can conclude that (5.3) can hold only in the two-bit systems, that is, for nontrivial observables $\boldsymbol{a} \in \mathbb{A}_1$, $\boldsymbol{b} \in \mathbb{A}_2$ for $d_1, d_2 > 2$,

$$\mathrm{Cov}_{\boldsymbol{p}_{12}}(\boldsymbol{a}, \boldsymbol{b}) = 0 \quad \underset{(\Longleftarrow)}{\not\Longrightarrow} \quad \boldsymbol{p}_{12} \in \Delta_{\mathrm{unc}}. \tag{5.18}$$

5.5. **Two-qubit systems.** Here we pose the question whether a result analogue to (5.3) could hold in quantum systems. For this we consider two-qubit systems, which is the smallest possibility.

We can exploit the wider possibilities of the quantum systems by considering states which do not commute with the observables. We again write the particular ansatz for the state of the subsystems and the whole system

$$\rho_1 = \frac{1}{2}(I + \boldsymbol{r}\boldsymbol{\sigma}) := \frac{1}{2}(I + r\sigma_1) \tag{5.19}$$

$$\rho_2 = \frac{1}{2}(I + \boldsymbol{s}\boldsymbol{\sigma}) := \frac{1}{2}(I + s\sigma_1) \tag{5.20}$$

$$\begin{aligned} \rho_{12} &= \frac{1}{4}\big(I \otimes I + \boldsymbol{r}\boldsymbol{\sigma} \otimes I + I \otimes \boldsymbol{s}\boldsymbol{\sigma} + \boldsymbol{t}(\boldsymbol{\sigma} \otimes \boldsymbol{\sigma})\big) \\ &:= \frac{1}{4}\big(I \otimes I + r\sigma_1 \otimes I + I \otimes s\sigma_1 + t\sigma_1 \otimes \sigma_1\big). \end{aligned} \tag{5.21}$$

Now substituting the Pauli matrices $\sigma_1 = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ into the states we get

$$\rho_1 = \frac{1}{2}\begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}, \quad \rho_2 = \frac{1}{2}\begin{bmatrix} 1 & s \\ s & 1 \end{bmatrix}, \tag{5.22}$$

$$\rho_{12} = \frac{1}{4}\left(\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} + r\begin{bmatrix}0&0&1&0\\0&0&0&1\\1&0&0&0\\0&1&0&0\end{bmatrix} + s\begin{bmatrix}0&1&0&0\\1&0&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix} + t\begin{bmatrix}0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{bmatrix}\right) \tag{5.23}$$
$$= \frac{1}{4}\begin{bmatrix}1&s&r&t\\s&1&t&r\\r&t&1&s\\t&r&s&1\end{bmatrix}.$$

For $\Gamma_{12} = \rho_{12} - \rho_1 \otimes \rho_2$ we need $\rho_1 \otimes \rho_2$, that is

$$\rho_1 \otimes \rho_2 = \frac{1}{4}\begin{bmatrix}1&s&r&rs\\s&1&rs&r\\r&rs&1&s\\rs&r&s&1\end{bmatrix}. \tag{5.24}$$

This way

$$\Gamma_{12} = \frac{1}{4}\begin{bmatrix}1&s&r&t\\s&1&t&r\\r&t&1&s\\t&r&s&1\end{bmatrix} - \frac{1}{4}\begin{bmatrix}1&s&r&rs\\s&1&rs&r\\r&rs&1&s\\rs&r&s&1\end{bmatrix} = \begin{bmatrix}0&0&0&t-rs\\0&0&t-rs&0\\0&t-rs&0&0\\t-rs&0&0&0\end{bmatrix} \tag{5.25}$$

is the expression for Γ_{12}. The observables are of the same form as previously $A = \left[\begin{smallmatrix}a_1&0\\0&a_2\end{smallmatrix}\right]$ $B = \left[\begin{smallmatrix}b_1&0\\0&b_2\end{smallmatrix}\right]$, so we can finally get the covariance $\mathrm{Cov}_{\rho_{12}}(A,B) = \mathrm{Tr}(\Gamma_{12}(A\otimes B))$ as

$$\mathrm{Cov}_{\rho_{12}}(A,B) = \mathrm{Tr}\begin{bmatrix}0&0&0&(t-rs)a_2b_2\\0&0&(t-rs)a_2b_1&0\\0&(t-rs)a_1b_2&0&0\\(t-rs)a_1b_1&0&0&0\end{bmatrix} = 0. \tag{5.26}$$

We see that Γ_{12} is not the zero operator, so the state is correlated, however, $\mathrm{Cov}_{\rho_{12}}(A,B) = 0$. (A very similar derivation of covariance can be done for the case when $\boldsymbol{r\sigma} = r\sigma_2$ with the same result.) This example already rules out the quantum analogue of the result (5.3), however, we will also elaborate on this in a more general form, without using matrices explicitly.

For this again consider the states of the 2 qubit system

$$\rho_1 = \frac{1}{2}(I + \boldsymbol{r\sigma}), \quad \rho_2 = \frac{1}{2}(I + \boldsymbol{s\sigma}), \tag{5.27}$$

$$\rho_1 \otimes \rho_2 = \frac{1}{4}\big(I\otimes I + \boldsymbol{r\sigma}\otimes I + I\otimes \boldsymbol{s\sigma} + \boldsymbol{r\sigma}\otimes\boldsymbol{s\sigma}\big), \tag{5.28}$$

$$\rho_{12} = \frac{1}{4}\big(I\otimes I + \boldsymbol{r\sigma}\otimes I + I\otimes \boldsymbol{s\sigma} + \boldsymbol{t}(\boldsymbol{\sigma}\otimes\boldsymbol{\sigma})\big). \tag{5.29}$$

Now to use them we need to rewrite them in indexed form, this looks as

$$\rho_1 = \frac{1}{2}\big(I + \sum_i r_i\sigma_i\big), \quad \rho_2 = \frac{1}{2}\big(I + \sum_j s_j\sigma_j\big), \tag{5.30}$$

$$\rho_1 \otimes \rho_2 = \frac{1}{4}\big(I\otimes I + \sum_i r_i\sigma_i\otimes I + I\otimes\sum_j s_j\sigma_j + \sum_{ij} r_is_j(\sigma_i\otimes\sigma_j)\big), \tag{5.31}$$

$$\rho_{12} = \frac{1}{4}\big(I\otimes I + \sum_i r_i\sigma_i\otimes I + I\otimes\sum_j s_j\sigma_j + \sum_{ij} t_{ij}(\sigma_i\otimes\sigma_j)\big). \tag{5.32}$$

Now to get the operator Γ_{12} encoding the correlation as

$$\Gamma_{12} = \rho_{12} - \rho_1 \otimes \rho_2 = \frac{1}{4}\sum_{ij}(t_{ij} - r_i s_j)(\sigma_i \otimes \sigma_j), \tag{5.33}$$

and finally the covariance as

$$\begin{aligned} \mathrm{Cov}_{\rho_{12}}(A, B) &= \mathrm{Tr}(\Gamma_{12}(A \otimes B)) \\ &= \frac{1}{4}\,\mathrm{Tr}\Big(\sum_{ij}(t_{ij} - r_i s_j)(\sigma_i \otimes \sigma_j)A \otimes B\Big) \\ &= \frac{1}{4}\sum_{ij}(t_{ij} - r_i s_j)\,\mathrm{Tr}(\sigma_i A)\,\mathrm{Tr}(\sigma_j B). \end{aligned} \tag{5.34}$$

From this expression we see that to get zero covariance one has to measure such state-observable pairs in which the observable $A \otimes B$ and the correlation operator Γ_{12} has non-zero components in different Pauli directions. Meaning that at least one subsystem has to have a zero component in the direction that the state has a non-zero component. Therefore the covariance being zero on a pair of observables does not imply that the state ρ_{12} is uncorrelated even on two qubits.

We can illustrate these if we consider for example a Bell state $|\mathrm{Bell}\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ and another state $|\mathrm{Bell}'\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle - |11\rangle)$, where $|ij\rangle$ is the computational basis, by which all the matrices are expressed. The matrices of the density operators of the two states $\rho_B = |B_0\rangle\langle B_0|$ and $\rho_\psi = |\psi\rangle\langle\psi|$ are

$$\rho_{\mathrm{Bell}} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \quad \rho_{\mathrm{Bell}'} = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix} \tag{5.35}$$

We get the reduced states from both density operators by the partial trace. When calculating for both matrices we see that their reduced states are the same, these are

$$\rho_1 := \mathrm{Tr}_2(\rho_{\mathrm{Bell}}) = \mathrm{Tr}_2(\rho_{\mathrm{Bell}'}) = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \rho_2 := \mathrm{Tr}_1(\rho_{\mathrm{Bell}}) = \mathrm{Tr}_1(\rho_{\mathrm{Bell}'}) = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{5.36}$$

so even the reduced states are the same. Now if we calculate $\Gamma_{12} = \rho_{12} - \rho_1 \otimes \rho_2$ for both cases, we get

$$\Gamma_{\mathrm{Bell}} = \frac{1}{4}\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix}, \quad \Gamma_{\mathrm{Bell}'} = \frac{1}{4}\begin{bmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{bmatrix}. \tag{5.37}$$

Now we can identify what combination of Pauli products are these Γ_{Bell}, $\Gamma_{\mathrm{Bell}'}$ matrices. This way the Pauli forms of the matrices are

$$\Gamma_{\mathrm{Bell}} = \frac{1}{4}(\sigma_1 \otimes \sigma_1 - \sigma_2 \otimes \sigma_2 + \sigma_3 \otimes \sigma_3), \quad \Gamma_{\mathrm{Bell}'} = \frac{1}{4}(\sigma_1 \otimes \sigma_3 + \sigma_2 \otimes \sigma_2 + \sigma_3 \otimes \sigma_1). \tag{5.38}$$

These are nonzero, so the Bell states are correlated. On the other hand, the covariances for spin measurements $\boldsymbol{A\sigma}$, $\boldsymbol{B\sigma}$ are

$$\mathrm{Cov}_{\rho_{\mathrm{Bell}}}(\boldsymbol{A\sigma}, \boldsymbol{B\sigma}) = A_1B_1 - A_2B_2 + A_3B_3, \quad \mathrm{Cov}_{\rho_{\mathrm{Bell}}}(\boldsymbol{A\sigma}, \boldsymbol{B\sigma}) = A_1B_3 + A_2B_2 + A_3B_1, \tag{5.39}$$

which can be zero for carefully chosen measurement directions.

As smaller systems can be embedded into larger ones, we can conclude that a result analogue to (5.3) cannot hold in the quantum systems, that is, for nontrivial observables $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$

for $d_1, d_2 \geq 2$,

$$\mathrm{Cov}_{\rho_{12}}(A, B) = 0 \quad \underset{(\Longleftarrow)}{\not\Longrightarrow} \quad \rho_{12} \in \mathcal{D}_{\mathrm{unc}}. \tag{5.40}$$

6. Summary, remarks and open questions

We have reviewed the tools that are used in classical and quantum probability theory to describe discrete finite systems and the useful and expressive measures for quantifying different kinds of correlations. By the use of these tools we have examined the properties of the covariance of observables and the correlation of the states in both the classical and the quantum cases. A state is uncorrelated (product), by definition, if covariance vanishes for all pairs of observables. (In the classical probability theory this is also called independence.) The point is that it might happen that we may have zero covariance for a specific pair of observables, while the state is not uncorrelated. Or we may have a low value of covariance for a specific pair of observables, while the state is highly correlated. (See also the inequalities (4.45), (4.53).) In this case that pair of observables is not good enough to detect the correlation. We obtained the result that such situation cannot occur in the classical two-bit system, it is enough for the covariance to vanish on a fixed pair of nontrivial observables and even this implies that the state is uncorrelated (see equation (5.3)). This, however, holds only in the two-bit case. In any larger classical systems (see equation (5.18)) and in any quantum systems (see equation (5.40)) there are pairs of observables of vanishing covariance in states which are correlated. We have shown this by constructing explicit examples. The result on bit/qubit systems is just another difference between classical and quantum correlations: two-bit systems are too small, but even two-qubit systems are large enough for the vanishing of the covariance not to be sufficient for the productness.

There are many interesting directions to proceed in this research. For example, we plan to quantify the strength of these phenomena, and study its relation to the entanglement in the quantum state, at least in the two-qubit case.

Acknowledgments

I would like to thank my internal supervisor László Oroszlány for accepting this role and reviewing my thesis. Also I am grateful for the financial support of the Wigner Internship Program of the Wigner Research Center for Physics. Finally I am greatly thankful for my supervisor Szilárd Szalay for his enormous support and help during the project.

References

[1] Ingemar Bengtsson and Karol Życzkowski, *Geometry of quantum states: An introduction to quantum entanglement*, 2nd ed., Cambridge University Press, 2017.
[2] Kavan Modi, Tomasz Paterek, Wonmin Son, Vlatko Vedral, and Mark Williamson, *Unified view of quantum and classical correlations*, Physical Review Letters **104** (2010Feb), 080501.
[3] Dénes Petz, *Quantum information theory and quantum statistics*, Springer Berlin Heidelberg, 2008.
[4] Vlatko Vedral and Martin B. Plenio, *Entanglement measures and purification procedures*, Phys. Rev. A **57** (1998Mar), 1619–1633.
[5] Reinhard F. Werner, *Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model*, Physical Review A **40** (1989Oct), no. 8, 4277–4281.
[6] Mark M. Wilde, *Quantum information theory*, Cambridge University Press, 2013.
[7] Michael M. Wolf, Frank Verstraete, Matthew B. Hastings, and J. Ignacio Cirac, *Area laws in quantum systems: Mutual information and correlations*, Phys. Rev. Lett. **100** (2008Feb), 070502.