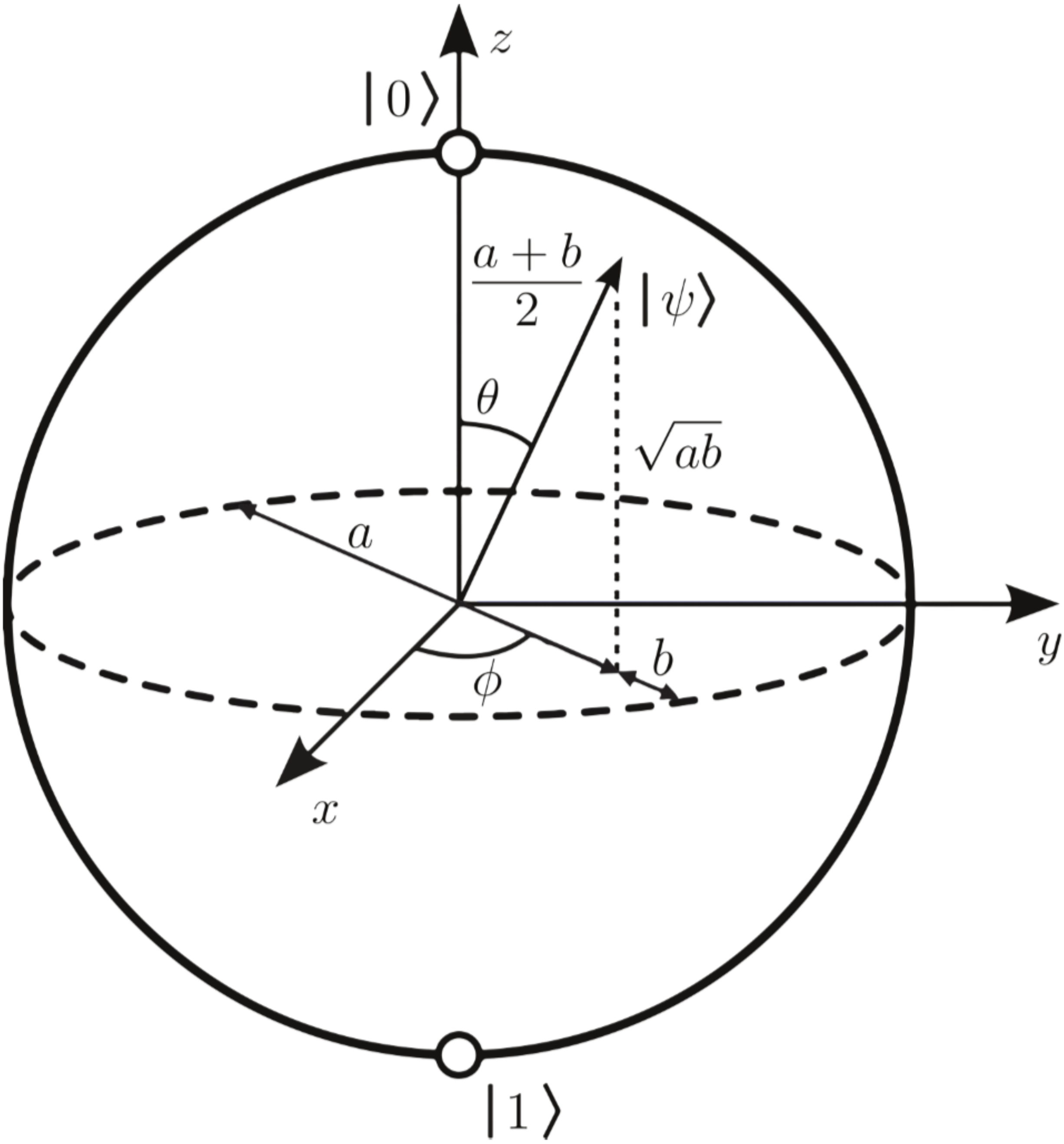
Inequalities and Quantum Entanglement

From the Cauchy-Schwarz Inequality to Non-linear Entanglement Witnesses

S.M. Loor

Delft University of Technology



TU Delft Delft University of Technology

Challenge the future

INEQUALITIES AND QUANTUM ENTANGLEMENT

FROM THE CAUCHY-SCHWARZ INEQUALITY TO NON-LINEAR ENTANGLEMENT WITNESSES

by

S.M. Loor
4574117

in partial fulfilment of the requirements for the degree of
Bachelor of Science
in Applied Physics and Applied Mathematics
at Delft University of Technology.

Supervisors:	Dr. D. Elkouss Coronas
	Dr. J. Vermeer
Committee Members:	Prof. Dr. C. Vuik
	Dr. T. Taminiau

TU Delft Delft University of Technology

Voor mama, mijn bron van toewijding, motivatie en inspiratie.

Abstract

In this thesis, two topics are studied: mathematical inequalities and non-linear quantum entanglement witnesses.
First, various inequalities, like the Cauchy-Schwarz inequality (on finite dimensional vector spaces) and Jensen's inequality, along with their extensions and generalisations, are proved and discussed. The intimate relationship between these inequalities is studied. Because this thesis was restricted to finite dimensional vector spaces, the consequences of generalising the results to infinite dimensional vector spaces are finally determined.
Secondly, the topic of entanglement detection is discussed - specifically, non-linear entanglement witnesses are considered. A bipartite and multipartite entanglement criterion based on the previously discussed inequalities are introduced and assessed extensively by considering their optimality, how they relate to other criteria as well as their limitations.

Contents

INTRODUCTION

The rich but peculiar structure of the realm of quantum phenomena is exemplified by quantum entanglement, which Albert Einstein famously described as *spukhafte Fernwirkung*: spooky action at a distance. In spite of - or rather due to - its very strange nature, entanglement has been the topic of a plethora of research since it was first described by Einstein, Podolsky and Rosen in 1935 (Einstein et al. (1935)).
Entanglement is not just interesting from a theoretical point of view, however: within the field of quantum information theory, entanglement is of great importance for applications like quantum teleportation (Bennett et al. (1993)), quantum cryptography and quantum communication; more specifically, entanglement is essential for superdense coding (Mattle et al. (1995)). In order to make use of entanglement as a resource, one needs to be certain that the states at hand are indeed entangled or indeed not entangled (i.e. separable). Up until today, however, no all-encompassing entanglement detection method of any practical use has been found. On top of that, it has even been proved that solving the problem of determining whether a state is separable or not is NP-hard (Ioannou (2007)). For practical purposes, however, this problem is often addressed using so-called *entanglement criteria*, which are often given in the form of an inequality. One of the most frequently-used inequality-based entanglement detection methods is based on so-called *entanglement witnesses* (Gühne and Tóth (2008)).

Inequalities, however, are not merely a mathematical tool for exploring the foundation of theoretical physics, but are also very essential for all branches of mathematics - in statistics, one encounters the famous Cramer-Rao bound (Rice (2007)), in probability theory, the Chebychev inequality (Grimmitt and Welsh (2014)), and so on. Furthermore, inequalities are key tools for investigating deeper mathematical structure, and as such, have proved their worth time and time again.

In this thesis, both the topic of entanglement as well as inequalities will be covered extensively. As such, this thesis will be split up into two parts. First, in Part I, an elaborate study of several well-known mathematical inequalities will be conducted, including, but not limited to, the Cauchy-Schwarz inequality, Jensen's inequality and the AM-GM inequality. They will be studied on two levels: on the one hand, they will be studied individually by delving into some of their generalisations, extensions and applications. On the other hand, they will be studied as a collective, by enquiring into the connection between these inequalities.

After this mathematical exploration, entanglement will be studied in Part II. First, the underlying mathematical structure of quantum mechanics will be described, after which the topic of entanglement detection will be delved into.

Lastly, these two fields of study will be brought together by extensively studying several entanglement criteria introduced by Wölk et al. (2014), which are based on the inequalities that will pass by in the mathematical part.

Part I

1

Mathematical Introduction: Defining the Setting

In this chapter, we introduce some preliminary definitions and concepts that will be relevant throughout the mathematical explorations that lie before us, as well as set the concrete setting in which we will be working, and explain this choice.
The rest of the mathematical part of this thesis is structured as follows.
In Chapter 2, we will start off by studying the Cauchy-Schwarz Inequality.
In Chapter 3, we will continue by studying the AM-GM inequality and will consider how it relates to the results derived in Chapter 2.
Then, we will study convexity and Jensen's inequality in Chapter 4.
In Chapter 5, we will introduce a generalisation of the AM-GM inequality.
In Chapter 6, we will study Hölder's inequality and the inequality of Minkowski, and show how these relate to *all* the inequalities derived up to that point.
In Chapter 7, we will finish off by studying p-spaces, and we will consider what happens when we extend the framework to the infinite dimensional case.

Note that Steele (2004) is used as the primary source for the mathematical journey that lies ahead of us.

1.1. Preliminary Definitions and Concepts

In this section, we recall some definitions. These are mostly based on Vermeer (2017) and Carothers (2000)). These definitions will be used and addressed throughout this thesis.

Before talking about inequalities, we need to introduce some preliminary concepts that allow for the description of these inequalities. As inequalities like that of Cauchy and Schwarz are fundamentally geometrical, we start off by introducing the concept of a vector space.

Definition 1.1. V is called a *vector space* over the field $\mathbb{L}$ if it is a non-empty set on which two operations are defined:

1. Addition such that $\mathbf{v}+\mathbf{w} \in V \quad \forall \mathbf{v}, \mathbf{w} \in V$
2. Scalar multiplication such that $\forall c \in \mathbb{L} \wedge \forall \mathbf{v} \in V : c\mathbf{v} \in V$

The operations with which we equip vector spaces should satisfy eight axioms.
For all $\mathbf{v}, \mathbf{w}, \mathbf{u} \in V$ and for all $c, d \in \mathbb{L}$, we have:

1. $\mathbf{v}+\mathbf{w}=\mathbf{w}+\mathbf{v}$ **commutativity**
2. $(\mathbf{v}+\mathbf{w})+\mathbf{u}=\mathbf{v}+(\mathbf{w}+\mathbf{u})$ **associativity**
3. $\exists \mathbf{0}\in V:\mathbf{v}+\mathbf{0}=\mathbf{v}$ **neutral element for addition**
4. $\forall \mathbf{v}\in V:\exists\,\hat{\mathbf{v}}\in V:\mathbf{v}+\hat{\mathbf{v}}=\mathbf{0}$
5. $1\mathbf{v}=\mathbf{v}$
6. $c\,(d\mathbf{v})=(cd)\,\mathbf{v}$
7. $(c+d)\,\mathbf{v}=c\mathbf{v}+d\mathbf{v}$ **distributivity I**
8. $c\,(\mathbf{v}+\mathbf{w})=c\mathbf{v}+c\mathbf{w}$ **distributivity II**

The structure provided by just a vector space is, however, not enough. We also want to be able to introduce additional structure in the form of - for example - distance, length and orthogonality. We will allow for this structure by presenting the concepts of an inner product, a norm and a metric.

Definition 1.2. Let V be a vector space. $\langle\cdot,\cdot\rangle : V\times V\to\mathbb{L}$[1] is called an *inner product* if it satisfies the following properties:

1. $\langle\mathbf{v},\mathbf{v}\rangle\geq 0 \quad \forall\mathbf{v}\in V$
2. $\langle\mathbf{v},\mathbf{v}\rangle=0 \iff \mathbf{v}=\mathbf{0}$
3. $\langle\alpha\mathbf{v},\mathbf{v}\rangle=\overline{\alpha}\,\langle\mathbf{v},\mathbf{v}\rangle \quad \forall\alpha\in\mathbb{L}\wedge\forall\mathbf{v},\mathbf{w}\in V$
4. $\langle\mathbf{u}+\mathbf{v},\mathbf{w}\rangle=\langle\mathbf{u},\mathbf{w}\rangle+\langle\mathbf{v},\mathbf{w}\rangle\,\forall\mathbf{v},\mathbf{w},\mathbf{u}\in V$
5. $\langle\mathbf{v},\mathbf{w}\rangle=\overline{\langle\mathbf{w},\mathbf{v}\rangle}\,\forall\,\mathbf{v},\mathbf{w}\in V$

We say that $(V,\langle\cdot,\cdot\rangle)$ is an *inner product space.*

Definition 1.3. Let V be a vector space over the field $\mathbb{L}$. Then $\|\cdot\|:V\to\mathbb{R}$ is called a *norm* if it satisfies the following properties:

1. $0\leq\|\mathbf{v}\|<\infty \quad \forall\mathbf{v}\in V$
2. $\|\mathbf{v}\|=0 \iff \mathbf{v}=0$
3. $\|\alpha\mathbf{v}\|=|\alpha|\|\mathbf{v}\| \quad \forall\alpha\in\mathbb{L}\wedge\forall\mathbf{v}\in V$
4. $\|\mathbf{v}+\mathbf{w}\|\leq\|\mathbf{v}\|+\|\mathbf{w}\| \quad \forall\mathbf{v},\mathbf{w}\in V$ **triangle Inequality**

We say that $(V,\|\cdot\|)$ is a *normed vector space.*

In this thesis, we will restrict ourselves to the real numbers (with an occasional extension to the complex field), and as such we will either set $\mathbb{L}=\mathbb{R}$ or $\mathbb{L}=\mathbb{C}$. Note that in the real case, property 5. of Definition 1.2 yields symmetry of the inner product.

Remark. Any inner product space can be extended to a normed inner product space by setting $\|\cdot\|:V\to R^{+}$ such that:

$$\|\mathbf{v}\|=\sqrt{\langle\mathbf{v},\mathbf{v}\rangle}$$

[1] Note that here $\mathbb{L}=\mathbb{R}$ or $\mathbb{L}=\mathbb{C}$.

Note that the norm of a vector represents the length of this vector. We also want to be able to measure distances. For this, we introduce the concept of a metric.

Definition 1.4. Given a set X, we call $d : X \times X \to \mathbb{R}$ a *metric* if it satisfies the following properties:

1. $d(x, y) \geq 0 \forall x, y \in X$
2. $d(x, y) = 0 \iff x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, z) + d(y, z) \leq d(x, y) \forall x, y, z \in X$ **triangle Inequality**

We call the pair (X, d) a *metric space*.

Remark. Any normed vector space can be extended to a metric space by setting:

$$d(x, y) = \|x - y\|$$

Note that this last remark is very important, since the notion of a metric is essential for performing analysis on normed vector spaces.

We will also be considering sequences and their convergence in this thesis. For this, it is necessary to introduce the concept of completeness.

Definition 1.5. Let (V, d) be a metric space, then it is called *complete* if every Cauchy sequence converges, that is, for every sequence $(x_n) \in V$ that is Cauchy, i.e.

$$\forall \epsilon > 0 : \exists N \in \mathbb{N} : \forall n, m \geq N : \|x_n - x_m\| < \epsilon$$

For the description of quantum mechanics, we need to combine all the concepts we have introduced up until now. This structure is called a Hilbert space.

Definition 1.6. H is called a *Hilbert Space* if it is both an inner product space and a complete metric space with respect to the metric induced by its inner product. We call H a *real* (or *complex*) Hilbert space if it is a real (or complex) inner product space.

Later on, we will discuss relation between inequalities on various vector fields. For this, we need to introduce the concepts of compactness and open sets.

Definition 1.7. Let (X, d) be a metric space. Then this metric space is called *compact* if it is *complete* and *totally bounded*, that is, if next to completeness, the following holds:

$$\forall \epsilon > 0 : \exists n \in \mathbb{N} : \exists \{x_1, \ldots, x_n \in X\} : X \subset \bigcup_{i=1}^{n} B_\epsilon(x_i)$$

We now introduce the concept of topologies.

Definition 1.8. Let X be a set and let τ denote a subset of X. Then τ is called a *topology* if:

1. $X \in \tau$ and $\emptyset \in \tau$
2. Given some index set $\mathscr{I}$ of sets U_i for $i \in \mathscr{I}$, then:
$$\bigcup_{i \in \mathscr{I}} U_i \in \tau$$
3. For any $n \in \mathbb{N}$:
$$\bigcap_{i=1}^{n} U_i \in \tau$$

We call the elements of τ the open sets of X. Furthermore, (X, τ) is called a *topological space*.

1.2. ESTABLISHING THE SETTING

In this thesis, we will mostly be concerned with studying inequalities on finite vector spaces - that is, on vector spaces with a basis of finite cardinality. This decision is justified, because restricting ourselves to finite fields allows for optimally exploring the topic of inequality, as the finitude of the vector spaces ensures preservation of many nice properties. This can be seen from the following theorem, which we state without proof.

Theorem 1.1. *let V be a finite dimensional vector space over the field $\mathbb{L}$ of dimension n. Then:*

$$V \cong \mathbb{L}^n$$

This theorem ensures that we can restrict ourselves to studying $\mathbb{R}^n$ (or in the case of complex extensions: $\mathbb{C}^n$) without loss of generality.
Lastly, it is also very relevant that our results are consistent in the sense that they remain true irrespective of our choice of basis. This will be established by a theorem and its corollary ,whose proof can be found in Appendix A.

Theorem 1.2. *Let V be a finite dimensional inner product space over the field $\mathbb{L}$ and let $B = \mathbf{b_1}, \dots, \mathbf{b_n}$ be a basis of V. Then $\forall \mathbf{v}, \mathbf{w} \in V$, we have that:*

$$\langle \mathbf{v}, \mathbf{w} \rangle = [\mathbf{v}]_B^\dagger G_B [\mathbf{w}]_B$$

where G_B is the Gramian matrix of B, so $(G_B)_{ij} = \langle \mathbf{b_i}, \mathbf{b_j} \rangle$ and $[\mathbf{v}]_B$ the coordinates of $\mathbf{v}$ expressed in terms of the new basis B, so if $\mathbf{v} = \sum_{k=1}^n c_k \mathbf{b_k}$, then $[\mathbf{v}]_{B,k} = c_k$.

Corollary 1.2.1. *Given a finite dimensional inner product space V with orthonormal basis $B = \{\mathbf{b_1}, \dots, \mathbf{b_n}\}$. Then:*

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle [\mathbf{v}]_B, [\mathbf{w}]_B \rangle$$

We will generously make use of these two facts throughout this thesis, and we will address the infinite dimensional case at the very end of our enquiry when discussing the so-called *p-spaces*.

2

THE STARTING POINT: CAUCHY AND SCHWARZ

In this chapter, we set off on our study of mathematical inequalities by introducing a very famous inequality, with many applications in various domains of mathematics and in theoretical physics: the Cauchy-Schwarz inequality. We will present several proofs of this inequality, first in a general real vector space and then specifically on $\mathbb{R}^n$, after which we study various generalisations, extensions and applications of this inequality.

2.1. CAUCHY-SCHWARZ ON GENERAL REAL INNER PRODUCT SPACES

We start off by presenting two forms of the Cauchy-Schwarz inequality along with their proofs, and we will show that these two forms are interchangeable on an arbitrary, real vector space. Let us start with the first of these two forms, and let us prove this statement using elementary properties of quadratic functions.

Theorem 2.1. (Cauchy-Schwarz Inequality). *Let* $(V, \langle\cdot,\cdot\rangle)$ *be a real inner product space and let* $\mathbf{v},\mathbf{w} \in V$. *Then*

$$|\langle \mathbf{v},\mathbf{w}\rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|,$$

Proof. Let $t \in \mathbb{R}$ and consider $\mathbf{v} + t\mathbf{w}$. Now:

$$<\mathbf{v}+t\mathbf{w},\mathbf{v}+t\mathbf{w}>\geq 0$$

$$\text{But also: } <\mathbf{v}+t\mathbf{w},\mathbf{v}+t\mathbf{w}>=<\mathbf{v},\mathbf{v}>+2t<\mathbf{v},\mathbf{w}>+t^2<\mathbf{w},\mathbf{w}>$$

But this is a parabola of t and is larger than or equal to zero, and can thus intersect the x-axis at most once, and thus has a discriminant of less than or equal to zero! So:

$$(2<\mathbf{v},\mathbf{w}>)^2-4<\mathbf{v},\mathbf{v}><\mathbf{w},\mathbf{w}>\leq 0$$

$$\Longleftrightarrow 4<\mathbf{v},\mathbf{w}>^2\leq 4<\mathbf{v},\mathbf{v}><\mathbf{w},\mathbf{w}>=4\|\mathbf{v}\|^2\|\mathbf{w}\|^2$$

Dividing both sides of the inequality by four and taking the square root of both sides yields the expected result. □

We can also express the Cauchy-Schwarz inequality by not taking the absolute value over the inner product and prove this statement separately.

Theorem 2.2. (Cauchy-Schwarz Inequality). *Let* $(V, \langle\cdot,\cdot\rangle)$ *be a real inner product space and let* $\mathbf{v},\mathbf{w} \in V$. *Then*

$$\langle \mathbf{v},\mathbf{w}\rangle \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|,$$

Proof. II. We now consider the case for t = -1, and let $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$ denote *normalised* vectors, i.e. $\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$ and $\hat{\mathbf{w}} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$. Then:

$$<\hat{\mathbf{v}}-\hat{\mathbf{w}}, \hat{\mathbf{v}}-\hat{\mathbf{w}}> \geq 0$$

Expanding the inner product yields:

$$<\hat{\mathbf{v}}, \hat{\mathbf{w}}> \leq \frac{1}{2} <\hat{\mathbf{v}}, \hat{\mathbf{v}}> + \frac{1}{2} <\hat{\mathbf{w}}, \hat{\mathbf{w}}> = 1$$

This yields:

$$\begin{aligned} &< \frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{\mathbf{w}}{\|\mathbf{w}\|} > \leq 1 \\ \iff &\frac{1}{\|\mathbf{v}\|}\frac{1}{\|\mathbf{w}\|} <\mathbf{v}, \mathbf{w}> \leq 1 \\ \iff &<\mathbf{v}, \mathbf{w}> \leq \|\mathbf{v}\|\|\mathbf{w}\| \end{aligned}$$

□

Remark. Note that the case of equality can very easily be derived from this proof. We find equality if and only if $<\hat{\mathbf{v}}-\hat{\mathbf{w}}, \hat{\mathbf{v}}-\hat{\mathbf{w}}> = 0$, which is equivalent to saying $\mathbf{v} = \lambda\mathbf{w}$ for some $\lambda \in \mathbb{R}$.

While one's mathematical intuition might lead us to think that the first version of this inequality is stronger, it can be proved that these two variants are fully equivalent. Let us now proceed to proving this statement.

Theorem 2.3. *On any real vector space V, the following statements are equivalent:*

1. $\langle \mathbf{v}, \mathbf{w} \rangle \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\| \ \forall\, \mathbf{v}, \mathbf{w} \in V$
2. $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\| \ \forall\, \mathbf{v}, \mathbf{w} \in V$

Proof. *2* $\Longrightarrow$ *1.* Since $\forall x \in \mathbb{R} : |x| \geq x$, the implication follows trivially.
1 $\Longrightarrow$ *2.* Note that $|x| = \max\{x, -x\}$. Thus $|\langle \mathbf{v}, \mathbf{w} \rangle| = \max\{\langle \mathbf{v}, \mathbf{w} \rangle, -\langle \mathbf{v}, \mathbf{w} \rangle\}$. note however that:

$$-\langle \mathbf{v}, \mathbf{w} \rangle = \langle -\mathbf{v}, \mathbf{w} \rangle \leq \|-\mathbf{v}\| \cdot \|\mathbf{w}\| = |-1|\|\mathbf{v}\| \cdot \|\mathbf{w}\| = \|\mathbf{v}\| \cdot \|\mathbf{w}\|$$

Thus surely $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|$ □

As such, we can restrict ourselves to either one of these inequalities without loss of generality.

We now turn to the Cauchy-Schwarz inequality on $\mathbb{R}^n$. Recall that this restriction can be implemented without loss of generality.

2.2. CAUCHY-SCHWARZ ON $\mathbb{R}^n$

In this section, we will study the inequality of Cauchy and Schwarz on the spaces of real vectors of finite dimension. We first present the theorem and give two proofs for it: the first proof uses the very simple fact that $(a-b)^2 \geq 0$, a fact that we will encounter again in the upcoming few chapters.

Theorem 2.4. Cauchy-Schwarz on $\mathbb{R}^n$. *Let $(a_k)_{k=1}^n, (b_k)_{k=1}^n$ be sequences such that $a_k, b_k \in \mathbb{R} \quad \forall k = 1, \ldots, n$. Then:*

$$\sum_{k=1}^n a_k b_k \leq \sqrt{\sum_{k=1}^n a_k^2}\sqrt{\sum_{k=1}^n b_k^2}$$

Proof. First note that $(a-b)^2 \geq 0 \,\forall\, a, b \in \mathbb{R}$. Therefore surely $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$. Thus:

$$\sum_{k=1}^{n} a_k b_k \leq \frac{1}{2}\sum_{k=1}^{n} a_k^2 + \frac{1}{2}\sum_{k=1}^{n} b_k^2$$

Now set $(a_n), (b_n)$ to normalised sequences. Then:

$$\sum_{k=1}^{n} \frac{a_k}{\sqrt{\sum_{k=1}^{n} a_k^2}} \frac{b_k}{\sqrt{\sum_{k=1}^{n} b_k^2}} \leq 1$$

$$\iff \sum_{k=1}^{n} a_k b_k \leq \sqrt{\sum_{k=1}^{n} a_k^2}\sqrt{\sum_{k=1}^{n} b_k^2}$$

□

We now present an alternative proof for the Cauchy-Schwarz Inequality. In this proof, we make use of the fact the inner product of two vectors as well as the norm of every vector are conserved under a change of bases.

Proof. Given two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. We first use the Gram-Schmidt process (see: Vermeer (2017)) to find an orthogonal basis:

$$\mathbf{w_1} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$
$$\mathbf{w_2} = \mathbf{y} - <\mathbf{y}, \mathbf{w_1}> \mathbf{w_1}, \text{ etc.}$$

Now $\mathbf{x}, \mathbf{y}$ can be expressed in terms of the new basis $\{\mathbf{w_1}, \mathbf{w_2}, \ldots, \mathbf{w_n}\}$ in the following manner:

$$[\mathbf{x}]_W = \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ and } [\mathbf{y}]_W = \begin{bmatrix} y_1 \\ y_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

But then:

$$< [\mathbf{x}]_W, [\mathbf{y}]_W >= x_1 y_1 \leq \sqrt{x_1^2}\sqrt{y_1^2 + y_2^2} = \|[\mathbf{x}]_W\| \|[\mathbf{y}]_W\|$$

Note, however, that since W is an orthogonal basis for $\mathbb{R}^n$, Corollary 1.2.1 yields that this statement is valid irrespective of the basis of choice. □

Now that we have introduced and proved the Cauchy-Schwarz inequality, let us consider whether we can generalise this inequality.

2.2.1. GENERALISATIONS OF THE CAUCHY-SCHWARZ INEQUALITY ON $\mathbb{R}^n$

As a first step towards a generalisation, let us consider whether or not we can extend the Cauchy-Schwarz inequality to three sequences. A logical first step to take is to just apply the inequality twice. The following theorem shows us that this yields a different upper bound.

Theorem 2.5. *Let $(a_k))_{k=1}^n, (b_k))_{k=1}^n$ and $(c_k))_{k=1}^n$ be sequences such that $\forall k = 1, \ldots, n : a_k, b_k, c_k \in \mathbb{R}$. Then:*

$$\left(\sum_{k=1}^{n} a_k b_k c_k\right)^4 \leq \left(\sum_{k=1}^{n} a_k^2\right)^2 \sum_{k=1}^{n} b_k^4 \sum_{k=1}^{n} c_k^4$$

Proof. Note that $(d_k))_{k=1}^n$ with $d_k = b_k c_k$ denotes a sequence. We can thus apply Theorem 2.4 on the sequences a_k and d_k to obtain:

$$\left(\sum_{k=1}^n a_k d_k\right)^4 \leq \left(\sum_{k=1}^n a_k^2\right)^2 \left(\sum_{k=1}^n d_k^2\right)^2$$

Substituting $d_k = b_k c_k$ and applying Theorem 2.4 to d_n yields:

$$\left(\sum_{k=1}^n a_k b_k c_k\right)^4 \leq \left(\sum_{k=1}^n a_k^2\right)^2 \left(\sum_{k=1}^n (b_k c_k)^2\right)^2 \leq \left(\sum_{k=1}^n a_k^2\right)^2 \sum_{k=1}^n b_k^4 \sum_{k=1}^n c_k^4$$

□

This is, however, not a form-preserving generalisation of the inequality at hand. Such a form-preserving generalisation is nevertheless still valid. Let us try to derive this result, first by tediously applying the principle of induction.

Theorem 2.6. *Let $(a_k))_{k=1}^n$, $(b_k))_{k=1}^n$ and $(c_k))_{k=1}^n$ be sequences such that $\forall\, k = 1,\ldots,n : a_k, b_k, c_k \in \mathbb{R}$. Then:*

$$\left(\sum_{k=1}^n a_k b_k c_k\right)^2 \leq \sum_{k=1}^n a_k^2 \sum_{k=1}^n b_k^2 \sum_{k=1}^n c_k^2$$

Proof. We will first prove this theorem using induction.
n = 1. We obtain equality.
n = 2. We have to prove that:

$$(a_1b_1c_1 + a_2b_2c_2)^2 \leq \left(a_1^2 + a_2^2\right)\left(b_1^2 + b_2^2\right)\left(c_1^2 + c_2^2\right)$$

Writing out this statement yields:

$$\begin{aligned}(a_1b_1c_1)^2 + 2(a_1b_1c_1)(a_2b_2c_2) + (a_2b_2c_2)^2 \leq (a_1b_1c_1)^2 + (a_1b_1c_2)^2 + (a_1b_2c_1)^2 + (a_1b_2c_2)^2 +\\ (a_2b_1c_1)^2 + (a_2b_1c_2)^2 + (a_2b_2c_1)^2 + (a_2b_2c_2)^2\end{aligned}$$

Working this out proves the basis step:

$$\begin{aligned}2(a_1b_1c_1)(a_2b_2c_2) &\leq (a_1b_1c_2)^2 + (a_1b_2c_1)^2 + (a_1b_2c_2)^2\\ &\quad + (a_2b_1c_1)^2 + (a_2b_1c_2)^2 + (a_2b_2c_1)^2\\ 0 &\leq (a_1b_2c_1)^2 + (a_1b_2c_2)^2 + (a_2b_1c_1)^2 + (a_2b_1c_2)^2\\ &\quad + (a_1b_1c_2)^2 - 2(a_1b_1c_2)(a_2b_2c_1) + (a_2b_2c_1)^2\end{aligned}$$

As such, we find:

$$\begin{aligned}(a_1b_2c_1)^2 + (a_1b_2c_2)^2 + (a_2b_1c_1)^2 + (a_2b_1c_2)^2 + (a_1b_1c_2)^2 - 2(a_1b_1c_2)(a_2b_2c_1) + (a_2b_2c_1)^2\\ \geq (a_1b_1c_2)^2 - 2(a_1b_1c_2)(a_2b_2c_1) + (a_2b_2c_1)^2\\ = ((a_1b_1c_2) - (a_2b_2c_1))^2 \geq 0\end{aligned}$$

which finishes the proof for the case $n = 2$.
n $\implies$ n + 1. Let us now assume that the induction hypothesis is valid for a certain $n \in \mathbb{N}$. We then find:

$$\left(\sum_{k=1}^{n+1} a_k b_k c_k\right)^2 = \left(\sum_{k=1}^n a_k b_k c_k + a_{n+1}b_{n+1}c_{n+1}\right)^2 \leq \left(\sqrt{\sum_{k=1}^n a_k^2 \sum_{k=1}^n b_k^2 \sum_{k=1}^n c_k^2} + a_{n+1}b_{n+1}c_{n+1}\right)^2$$

due to the induction hypothesis for n. Setting $\alpha = \sqrt{\sum_{k=1}^n a_k^2}$, $\beta = \sqrt{\sum_{k=1}^n b_k^2}$ and $\gamma = \sqrt{\sum_{k=1}^n c_k^2}$, we find:

$$\left(\alpha\beta\gamma + a_{n+1}b_{n+1}c_{n+1}\right)^2 \leq \left(\alpha^2 + a_{n+1}^2\right)\left(\beta^2 + b_{n+1}^2\right)\left(\gamma^2 + c_{n+1}^2\right)$$

due to the induction hypothesis for the case of $n = 2$, proving the hypothesis for $n+1$. □

This proof was not very elegant. There is a much more elegant (and surprisingly simple!) proof, however, which only makes use of the Cauchy-Schwarz inequality for two sequences and the triangle inequality.

Proof. We will now prove this using Theorem 2.4. We find, using the Triangle Inequality:

$$\begin{aligned}
\left(\sum_{k=1}^n a_k b_k c_k\right)^2 = \left|\sum_{k=1}^n a_k b_k c_k\right|^2 \leq \left(\sum_{k=1}^n |a_k b_k c_k|\right)^2 &= \left(\sum_{k=1}^n |a_k||b_k||c_k|\right)^2 \\
&\leq \left(\sum_{k=1}^n |a_k||b_k| \sqrt{\sum_{k=1}^n |c_k|^2}\right)^2 \\
&= \left(\sum_{k=1}^n |a_k||b_k|\right)^2 \sum_{k=1}^n c_k^2 \leq \sum_{k=1}^n a_k^2 \sum_{k=1}^n b_k^2 \sum_{k=1}^n c_k^2
\end{aligned}$$

where we use that $\forall i : |c_i| = \sqrt{|c_i|^2} \leq \sqrt{\sum_{k=1}^n |c_k|^2}$. □

From this, we see that a generalisation of the statement to three sequences holds. This naturally leads to the generalisation of Theorem 2.4 for $m \in \mathbb{N}\backslash\{1\}$ sequences. We now state this generalisation and proof it in two ways: first by generalising the induction proof, which ends up being reasonably straight-forward.

Theorem 2.7. *Let $m \in \mathbb{N}\backslash\{1\}$ and let $\left(x_k^1\right), \left(x_k^2\right), \ldots, \left(x_k^m\right)$ be sequences such that $\forall p = 1, \ldots, m :$ $\forall k = 1, \ldots, n : x_k^p \in \mathbb{R}$. Then:*

$$\left(\sum_{k=1}^n \prod_{p=1}^m x_k^p\right)^2 \leq \prod_{p=1}^m \sum_{k=1}^n \left(x_k^p\right)^2$$

Proof. Again, we first prove the inequality using induction. For $m = 2$, we retrieve Theorem 2.4, and for $m = 3$, we retrieve Theorem 2.6. We thus set $m \in \mathbb{N}\backslash\{1,2,3\}$.
We proceed in a similar fashion as we did for the case of $m = 2$ and $m = 3$ by proving this theorem using the principle of induction.
n = 1. We, again, find equality.
n = 2. We want to prove that

$$\left(\sum_{k=1}^2 \prod_{p=1}^m x_k^p\right)^2 \leq \prod_{p=1}^m \sum_{k=1}^2 \left(x_k^p\right)^2$$

$$\left(\prod_{p=1}^m x_1^p + \prod_{p=1}^m x_2^p\right)^2 \leq \prod_{p=1}^m \left(\left(x_1^p\right)^2 + \left(x_2^p\right)^2\right)$$

$$\left(\prod_{p=1}^m x_1^p\right)^2 + 2\prod_{p=1}^m x_1^p \prod_{p=1}^m x_2^p + \left(\prod_{p=1}^m x_2^p\right)^2 \leq \left(\prod_{p=1}^m x_1^p\right)^2 + \left(\prod_{p=1}^m x_2^p\right)^2 + \left(x_2^m \prod_{p=1}^{m-1} x_1^p\right)^2 + \left(x_1^m \prod_{p=1}^{m-1} x_2^p\right)^2 + \mathcal{O}\left(\left(x_1^p\right)^2, \left(x_2^p\right)^2\right)$$

where $\mathcal{O}\left((x_1^p)^2,(x_2^p)^2\right)$ is some polynomial sum of $(x_1^1)^2,(x_2^1)^2,(x_1^2)^2,\dots,(x_2^m)^2$ (and is therefore surely non-negative). We can rewrite this as:

$$2\prod_{p=1}^{m}x_1^p\prod_{p=1}^{m}x_2^p \le \left(x_2^m\prod_{p=1}^{m-1}x_1^p\right)^2+\left(x_1^m\prod_{p=1}^{m-1}x_2^p\right)^2+\mathcal{O}\left((x_1^p)^2,(x_2^p)^2\right)$$

$$0 \le \left(x_2^m\prod_{p=1}^{m-1}x_1^p\right)^2+\left(x_1^m\prod_{p=1}^{m-1}x_2^p\right)^2-2\prod_{p=1}^{m}x_1^p\prod_{p=1}^{m}x_2^p+\mathcal{O}\left((x_1^p)^2,(x_2^p)^2\right)$$

We find:

$$\begin{aligned}&\left(x_2^m\prod_{p=1}^{m-1}x_1^p\right)^2+\left(x_1^m\prod_{p=1}^{m-1}x_2^p\right)^2-2\prod_{p=1}^{m}x_1^p\prod_{p=1}^{m}x_2^p+\mathcal{O}\left((x_1^p)^2,(x_2^p)^2\right)\ge\\&\left(x_2^m\prod_{p=1}^{m-1}x_1^p\right)^2+\left(x_1^m\prod_{p=1}^{m-1}x_2^p\right)^2-2\prod_{p=1}^{m}x_1^p\prod_{p=1}^{m}x_2^p\\&=\left(x_2^m\prod_{p=1}^{m-1}x_1^p\right)^2+\left(x_1^m\prod_{p=1}^{m-1}x_2^p\right)^2-2\left(x_2^m\prod_{p=1}^{m-1}x_1^p\right)\left(x_1^m\prod_{p=1}^{m-1}x_2^p\right)\\&=\left(x_2^m\prod_{p=1}^{m-1}x_1^p-x_1^m\prod_{p=1}^{m-1}x_2^p\right)^2\ge 0\end{aligned}$$

which proves the hypothesis for $n=2$.
n $\implies$ n + 1. We first assume that the induction hypothesis holds for n. Then:

$$\left(\sum_{k=1}^{n+1}\prod_{p=1}^{m}x_k^p\right)^2=\left(\sum_{k=1}^{n}\prod_{p=1}^{m}x_k^p+\prod_{p=1}^{m}x_{n+1}^p\right)^2\le\left(\sqrt{\prod_{p=1}^{m}\sum_{k=1}^{n}(x_k^p)^2}+\prod_{p=1}^{m}x_{n+1}^p\right)^2=\left(\prod_{p=1}^{m}\sqrt{\sum_{k=1}^{n}(x_k^p)^2}+\prod_{p=1}^{m}x_{n+1}^p\right)^2$$

Setting $\alpha_p=\sqrt{\sum_{k=1}^{n}(x_k^p)^2}\quad\forall p=1,\dots,m$, we can use the induction hypothesis for the case of $n=2$ to find:

$$\left(\prod_{p=1}^{m}\sqrt{\sum_{k=1}^{n}(x_k^p)^2}+\prod_{p=1}^{m}x_{n+1}^p\right)^2=\left(\prod_{p=1}^{m}\alpha_p+\prod_{p=1}^{m}x_{n+1}^p\right)^2\le\prod_{p=1}^{m}\left(\alpha_p^2+(x_{n+1}^p)^2\right)=\prod_{p=1}^{m}\sum_{k=1}^{n+1}(x_k^p)^2$$

which proves the induction hypothesis for $n+1$. □

Even though the induction proof is relatively straight-forward, it would be substantially easier to prove the generalisation at hand in some generalised form of the second proof of the case of three sequences, which used the Triangle Inequality. Fortunately, this can be done. Let us do so below.

Proof. First note that for any sequence $(a_k)_{k=1}^n$, we have:

$$\forall i:|a_i|=\sqrt{a_i^2}\le\sqrt{\sum_{k=1}^{n}a_k}\tag{2.1}$$

We thus find:

$$\left(\sum_{k=1}^{n}\prod_{p=1}^{m}x_k^p\right)^2\le\left(\sum_{k=1}^{n}x_k^1x_k^2\prod_{p=3}^{m}\sqrt{\sum_{k=1}^{n}(x_k^p)^2}\right)^2=\left(\sum_{k=1}^{n}x_k^1x_k^2\right)^2\prod_{p=3}^{m}\sum_{k=1}^{n}(x_k^p)^2\le\prod_{p=1}^{m}\sum_{k=1}^{n}(x_k^p)^2$$

□

An interesting question to consider is whether the upper bound found in Theorem 2.5 (by taking the square root on both sides) yields a tighter upper bound than the one found in Theorem 2.6. We can prove that this is indeed the case. Before doing so, however, we first need to introduce a lemma. We first consider a lemma. Note that this lemma is a corollary of a more general theorem that will be covered in Chapter 7, and as such, we omit the proof.

Lemma 2.8. *Let $(x_k)_{k=1}^n$ be a sequence of non-negative real numbers. Then surely:*

$$\sum_{k=1}^n x_k \leq \sqrt{\sum_{k=1}^n x_k^2}$$

Theorem 2.9. *Let $(a_k))_{k=1}^n$, $(b_k))_{k=1}^n$ and $(c_k))_{k=1}^n$ be sequences such that $\forall k = 1,\ldots,n : a_k, b_k, c_k \in \mathbb{R}$*

$$\sum_{k=1}^n a_k^2 \sqrt{\sum_{k=1}^n b_k^4} \sqrt{\sum_{k=1}^n c_k^4} \leq \sum_{k=1}^n a_k^2 \sum_{k=1}^n b_k^2 \sum_{k=1}^n c_k^2$$

Proof. Since $\forall k : a_k^2 \geq 0$, it suffices to prove that:

$$\sqrt{\sum_{k=1}^n b_k^4} \sqrt{\sum_{k=1}^n c_k^4} \leq \sum_{k=1}^n b_k^2 \sum_{k=1}^n c_k^2$$

This inequality is definitely satisfied if:

$$\sqrt{\sum_{k=1}^n b_k^4} \leq \sum_{k=1}^n b_k^2 \wedge \sqrt{\sum_{k=1}^n c_k^4} \leq \sum_{k=1}^n c_k^2$$

Setting $x_k = b_k^2$ and $y_k = c_k^2$ and using Lemma 2.8 proves the above statement. □

The question at hand of course is whether the generalisations that we have derived are also independent of the basis at hand. Obviously , the right-hand side of Theorem 2.6 is also independent of our choice of orthonormal basis, as this term only contains valid inner products. The left-hand side, however, is only proved to be independent of our choice of orthonormal basis for the case of two sequences - for higher dimensions, it ceases to represent an inner product. Let us now tackle this question, specifically for the case of three sequences. First, however, it would be convenient to introduce some more compact notation.

Definition 2.1. Let $\langle\cdot,\cdot,\cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ such that $\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle = \sum_{k=1}^n u_k v_k w_k$. Then $\langle\cdot,\cdot,\cdot\rangle$ satisfies the following three properties $\forall \mathbf{u},\mathbf{v},\mathbf{w} \in \mathbb{R}^n$:

1. $\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle = \langle \mathbf{v},\mathbf{u},\mathbf{w}\rangle = \langle \mathbf{u},\mathbf{w},\mathbf{v}\rangle$ **symmetry**
2. $\langle \alpha\mathbf{u},\mathbf{v},\mathbf{w}\rangle = \alpha\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle$
3. $\langle \mathbf{u_1}+\mathbf{u_2},\mathbf{v},\mathbf{w}\rangle = \langle \mathbf{u_1},\mathbf{v},\mathbf{w}\rangle + \langle \mathbf{u_2},\mathbf{v},\mathbf{w}\rangle$

Now if we let $B = \{\mathbf{b_1},\ldots,\mathbf{b_n}\}$ and $\mathbf{u} = \sum_{i=1}^n u_i \mathbf{b_i}$, $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{b_i}$ and $\mathbf{w} = \sum_{i=1}^n w_i \mathbf{b_i}$, we can derive that

$$\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle = \sum_{i=1}^n u_i \sum_{j=1}^n v_j \sum_{k=1}^n w_k \langle \mathbf{b_k},\mathbf{b_j},\mathbf{b_i}\rangle \tag{2.2}$$

Furthermore, we can derive that

$$\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle_B = \sum_{i=1}^{n} u_i v_i w_i = \sum_{i=1}^{n} u_i \sum_{j=1}^{n} v_j \sum_{k=1}^{n} w_k \delta_{ijk} \tag{2.3}$$

We thus see that these two can only be equal if we have $\langle \mathbf{b_k}, \mathbf{b_j}, \mathbf{b_i}\rangle = \delta_{ijk}$. This equation, is in general, not satisfied[1] and therefore, we see that the generalisation of the Cauchy-Schwarz inequality is, contrary to the original inequality, unfortunately not independent of our choice of basis.

Let us now move on to consider various applications of the Cauchy-Schwarz Inequality. We prove various theorems mentioned as exercises in Chapter 1 of Steele (2004).

2.3. APPLICATIONS OF THE CAUCHY-SCHWARZ INEQUALITY

We can apply the Cauchy-Schwarz inequality to derive some other results: we first derive bounds on the sum of the elements of a vector and then consider the use of the inequality when considering weighted sums.

We start off by deriving two bounds on the sum of the elements of a vector.

Theorem 2.10. *Let $(a_k)_{k=1}^n$ be a sequence such that $b_k \in \mathbb{R} \quad \forall k = 1,\dots,n$. Then:*

$$\sum_{k=1}^{n} a_k \leq \sqrt{n}\sqrt{\sum_{k=1}^{n} a_k^2}$$

Proof. Let $(a_k)_{k=1}^n$ be as defined and let $(b_k)_{k=1}^n$ such that $a_k = 1 \quad \forall k = 1,\dots,n$. Using Theorem 2.1, we find:

$$\sum_{k=1}^{n} a_k b_k \leq \sqrt{\sum_{k=1}^{n} a_k^2}\sqrt{\sum_{k=1}^{n} b_k^2} = \sqrt{\sum_{k=1}^{n} 1^2}\sqrt{\sum_{k=1}^{n} a_k^2} = \sqrt{n}\sqrt{\sum_{k=1}^{n} a_k^2}$$

□

Note that this inequality, although very simple to derive, gives a stronger bound than the bound that will be derived in a much more complicated manner in Chapter 7.
Let us now consider what upper bounds a sequence puts on the sum of its elements.

Theorem 2.11. *Let $(a_k)_{k=1}^n$ be a sequence such that $a_k \in \mathbb{R} \quad \forall k = 1,\dots,n$. Then:*

$$\sum_{k=1}^{n} |a_k| \leq \sqrt{\sum_{k=1}^{n} |a_k|^{\frac{2}{3}}}\sqrt{\sum_{k=1}^{n} |a_k|^{\frac{4}{3}}}$$

Proof. Note that $\sum_{k=1}^{n} |a_k| = \sum_{k=1}^{n} |a_k|^{\frac{1}{3}} |a_k|^{\frac{2}{3}}$. Let $b_k = |a_k|^{\frac{1}{3}}, c_k = |a_k|^{\frac{2}{3}} \quad \forall k = 1,\dots,n$, then using the Triangle Inequality and Theorem 2.1:

$$\sum_{k=1}^{n} |a_k| = \sum_{k=1}^{n} b_k c_k \leq \sqrt{\sum_{k=1}^{n} b_k^2}\sqrt{\sum_{k=1}^{n} c_k^2} = \sqrt{\sum_{k=1}^{n} |a_k|^{\frac{2}{3}}}\sqrt{\sum_{k=1}^{n} |a_k|^{\frac{4}{3}}}$$

□

[1] Take for example, $\mathbf{u} = \frac{1}{\sqrt{14}}\begin{bmatrix}1\\2\\3\end{bmatrix}, \mathbf{v} = \frac{1}{\sqrt{3}}\begin{bmatrix}-1\\-1\\1\end{bmatrix}, \mathbf{u} = \frac{1}{\sqrt{42}}\begin{bmatrix}5\\-4\\1\end{bmatrix}$ as our basis. Then $\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle_B = 0$, while $\langle \mathbf{u},\mathbf{v},\mathbf{w}\rangle = 1$.

This result can obviously be extended very easily to prove that $\forall N > 1$ and $1 < k < N$:

$$\sum_{k=1}^{n} |a_k| \leq \sqrt{\sum_{k=1}^{n} |a_k|^{\frac{2k}{N}}} \sqrt{\sum_{k=1}^{n} |a_k|^{\frac{2(N-k)}{N}}}$$

Now let us consider what upper bounds we can derive when considering weighted sums of sequences.

Theorem 2.12. *Let $(p_k)_{k=1}^n$ be a sequence such that $p_k \geq 0 \quad \forall k = 1,\dots,n \wedge \sum_{k=1}^n p_k = 1$. Furthermore, let $(a_k)_{k=1}^n$ and $(b_k)_{k=1}^n$ such that $a_k \geq 0, b_k \geq 0 \wedge a_k b_k \geq 1 \quad \forall k = 1,\dots,n$. Then*

$$\left(\sum_{k=1}^{n} a_k p_k\right)\left(\sum_{k=1}^{n} b_k p_k\right) \geq 1$$

Proof. We can define $c_k = \sqrt{a_k p_k}$ and $d_k = \sqrt{b_k p_k}$ and use Theorem 2.1 to find:

$$\sum_{k=1}^{n} \sqrt{a_k p_k}\sqrt{b_k p_k} = \sum_{k=1}^{n} c_k d_k \leq \sqrt{\sum_{k=1}^{n} c_k^2}\sqrt{\sum_{k=1}^{n} d_k^2} = \sqrt{\left(\sum_{k=1}^{n} a_k p_k\right)\left(\sum_{k=1}^{n} b_k p_k\right)}$$

Furthermore, we find that:

$$\sum_{k=1}^{n} \sqrt{a_k p_k}\sqrt{b_k p_k} = \sum_{k=1}^{n} p_k \sqrt{a_k b_k} \geq \sum_{k=1}^{n} p_k \cdot 1 = 1.$$

We thus find

$$\left(\sum_{k=1}^{n} a_k p_k \sum_{k=1}^{n} b_k p_k\right)^{\frac{1}{2}} \geq 1 \implies \left(\sum_{k=1}^{n} a_k p_k\right)\left(\sum_{k=1}^{n} b_k p_k\right) \geq 1$$

□

An interesting application of the Cauchy-Schwarz inequality is in the proof of the so-called Harker-Kasper inequality, which has applications in crystallography (see Steele (2004)). We prove this inequality in two ways. Let us start with the first proof. This proof requires the following lemma:

Lemma 2.13. *Let $f(x) : D \to \mathbb{R}$ for some domain D and let $p_k \geq 0 \quad \forall k = 1,\dots,n$ such that $\sum_{k=1}^n p_k = 1$. Furthermore, let $x_1,\dots,x_n \in D$. Then:*

$$\left(\sum_{k=1}^{n} p_k f(x_k)^2\right) - \left(\sum_{k=1}^{n} p_k f(x_k)\right)^2 \geq 0$$

Proof. We rewrite the left side of the inequality:

$$\begin{aligned}
&\left(\sum_{k=1}^{n} p_k f(x_k)^2\right) - 1\cdot 2\left(\sum_{k=1}^{n} p_k f(x_k)\right)^2 + 1\cdot\left(\sum_{k=1}^{n} p_k f(x_k)\right)^2 \\
&= \sum_{k=1}^{n} p_k \left(f(x_k)\right)^2 - \sum_{k=1}^{n} p_k \cdot 2\left(\sum_{k=1}^{n} p_k f(x_k)\right)^2 + \sum_{k=1}^{n} p_k \left(\sum_{k=1}^{n} p_k f(x_k)\right)^2 \\
&= \sum_{k=1}^{n} p_k \left(f(x_k)^2 - 2\left(\sum_{k=1}^{n} p_k f(x_k)\right)^2 + \left(\sum_{k=1}^{n} p_k f(x_k)\right)^2\right) \\
&= \sum_{k=1}^{n} p_k \left(f(x_k) - \sum_{k=1}^{n} p_k f(x_k)\right)^2 \geq 0
\end{aligned}$$

Since all $p_k \geq 0$. □

Remark. Note that this proves that for the variance of a *discrete* random variable X we have $\sigma^2(X) \geq 0$. This can also be seen as a consequence of Jensen's Inequality, which will be covered extensively in Chapter 4.

Theorem 2.14. (Harker-Kasper) *Let* $p_k \geq 0 \quad \forall k = 1, \dots, n \wedge \sum_{k=1}^n p_k = 1$. *Then*

$$\left(\sum_{k=1}^n p_k \cos(\beta_k x)\right)^2 \leq \frac{1}{2}\left(1 + \sum_{k=1}^n p_k \cos(2\beta_k x)\right)$$

Proof. Note that:

$$\begin{aligned}\sum_{k=1}^n p_k \cos^2(\beta_k x) = \sum_{k=1}^n p_k \left(\frac{1}{2}(1 + \cos(2\beta_k x))\right) &= \frac{1}{2}\left(\sum_{k=1}^n p_k + \sum_{k=1}^n p_k \cos(2\beta_k x)\right) \\ &= \frac{1}{2}\left(1 + \sum_{k=1}^n p_k \cos(2\beta_k x)\right)\end{aligned}$$

It remains to be proved that

$$\left(\sum_{k=1}^n p_k \cos(\beta_k x)\right)^2 \leq \sum_{k=1}^n p_k \cos^2(\beta_k x)$$

Setting $f(x) = \cos(\beta_k x)$ in Lemma 2.13 yields the expected result. □

This proof was rather long and required quite a detour which was presented in the form of a lemma. A second proof, however, becomes very easy, all thanks to Cauchy and Schwarz.

Proof. Set $a_k = \sqrt{p_k}$ and $b_k = \sqrt{p_k}\cos(\beta_x x)$. Then using the Cauchy-Schwarz inequality, we find:

$$\begin{aligned}\left(\sum_k p_k \cos(\beta_x x)\right)^2 \leq \sqrt{\sum_{k=1}^n p_k \sum_{k=1}^n p_k \cos^2(\beta_x x)} &= \sqrt{\sum_{k=1}^n p_k \sum_{k=1}^n p_k \left(\frac{1+\cos(2\beta_x x)}{2}\right)} \\ &= \frac{1}{2}\left(1 + \sum_{k=1}^n p_k \cos(2\beta_k x)\right)\end{aligned}$$

□

2.4. EXTENDING CAUCHY-SCHWARZ COMPLEX INNER PRODUCT SPACES

Lastly, we show that the Cauchy-Schwarz inequality is still valid on complex inner product spaces. We generalise two of the proofs for the real case: one for the case of general complex vector spaces, and one for $\mathbb{C}^n$.

Theorem 2.15. Cauchy-Schwarz *Let* $\mathbf{v}, \mathbf{w} \in \mathbb{C}^n$. *Then:*

$$|\langle \mathbf{v}, \mathbf{w}\rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|$$

Proof. I. Note that:

$$\langle \hat{\mathbf{v}} - \mathbf{w}, \hat{\mathbf{v}} - \mathbf{w}\rangle \geq 0$$

But also:

$$\langle \hat{\mathbf{v}} - \mathbf{w}, \hat{\mathbf{v}} - \mathbf{w}\rangle = \langle \hat{\mathbf{v}}, \hat{\mathbf{v}}\rangle - \langle \hat{\mathbf{v}}, \mathbf{w}\rangle - \langle \mathbf{w}, \hat{\mathbf{v}}\rangle + \langle \mathbf{w}, \mathbf{w}\rangle = \|\hat{\mathbf{v}}\| - \langle \hat{\mathbf{v}}, \mathbf{w}\rangle - \overline{\langle \hat{\mathbf{v}}, \mathbf{w}\rangle} + \|\mathbf{w}\| = \|\hat{\mathbf{v}}\| - 2\,\mathrm{Re}\{\langle \hat{\mathbf{v}}, \mathbf{w}\rangle\} + \|\mathbf{w}\|$$

we thus find:

$$\|\hat{\mathbf{v}}\| - 2\,\mathrm{Re}\,\{\langle\hat{\mathbf{v}},\mathbf{w}\rangle\} + \|\mathbf{w}\| \geq 0 \iff \mathrm{Re}\,\{\langle\hat{\mathbf{v}},\mathbf{w}\rangle\} \leq \frac{1}{2}\|\hat{\mathbf{v}}\| + \frac{1}{2}\|\mathbf{w}\|$$

Choosing $\mathbf{v},\mathbf{w}$ as normalised vectors yields:

$$\mathrm{Re}\,\{\langle\hat{\mathbf{v}},\mathbf{w}\rangle\} \leq \|\hat{\mathbf{v}}\|\,\|\mathbf{w}\|$$

Since $\langle\hat{\mathbf{v}},\mathbf{w}\rangle \in \mathbb{C}$, we can assume that $\exists\rho,\theta \in \mathbb{R} : \langle\hat{\mathbf{v}},\mathbf{w}\rangle = \rho\exp(i\theta)$ without loss of generality. Choose $\mathbf{v} = \exp(-i\theta)$, then:

1. $\langle\mathbf{v},\mathbf{w}\rangle = \mathrm{Re}\,\{\langle\mathbf{v},\mathbf{w}\rangle\} = |\langle\hat{\mathbf{v}},\mathbf{w}\rangle|$
2. $\langle\hat{\mathbf{v}},\hat{\mathbf{v}}\rangle = \langle\mathbf{v},\mathbf{v}\rangle \implies \|\hat{\mathbf{v}}\| = \|\mathbf{v}\|$

But then:

$$|\langle\hat{\mathbf{v}},\mathbf{w}\rangle| = \mathrm{Re}\,\{\langle\mathbf{v},\mathbf{w}\rangle\} \leq \|\mathbf{v}\|\,\|\mathbf{w}\| = \|\hat{\mathbf{v}}\|\,\|\mathbf{w}\|$$

□

Note that Theorem 1.2 is also valid when $V = \mathbb{C}^n$, and that the Gram Schmidt process can still be applied. It is thus trivial to prove this theorem in a completely analogous manner as the second proof of Theorem 2.4:

Proof. Given two vectors $\mathbf{x},\mathbf{y} \in \mathbb{C}^n$. We first use the Gram-Schmidt process to find an orthogonal basis:

$$\mathbf{w_1} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$
$$\mathbf{w_2} = \mathbf{y} - <\mathbf{y},\mathbf{w_1}>\mathbf{w_1}, \text{ etc.}$$

Now $\mathbf{x},\mathbf{y}$ can be expressed in terms of the new basis $\{\mathbf{w_1},\mathbf{w_2},\ldots,\mathbf{w_n}\}$ as follows:

$$[\mathbf{x}]_W = \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \text{ and } [\mathbf{y}]_W = \begin{bmatrix} y_1 \\ y_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

But then:

$$<\mathbf{x},\mathbf{y}>_W = |[\mathbf{x}]_W^\dagger\,[\mathbf{y}]_W| = |\bar{x}_1 y_1| \leq \sqrt{x_1^2}\sqrt{y_1^2 + y_2^2} = \|\mathbf{x}\|^2\|\mathbf{y}\|^2$$

Note however that since W is an orthogonal basis for $\mathbb{C}^n$, Corollary 1.2.1 yields that this statement is valid irrespective of the basis of choice. □

3

The AM-GM Inequality

In Chapter 3, the proof by induction of The Cauchy-Schwarz inequality on $\mathbb{R}^n$ and its generalisations relied heavily on the trivial fact that for all $x, y \in \mathbb{R}$, we have $(x-y)^2 \geq 0$. By choosing $x = \sqrt{a}$ and $y = \sqrt{b}$, we can rewrite this statement to $\sqrt{ab} \leq \frac{a+b}{2}$.
This hints to another family of inequalities that is closely related to the Cauchy-Schwarz inequality: inequalities relating to averages. This chapter will deal with one well-known inequality of averages: the inequality of the arithmetic and the geometric average, shortened to the AM-GM inequality.

3.1. The AM-GM inequality

Let us first define the averages that will be at the core of this chapter.

Definition 3.1. The *arithmetic mean* $\bar{a}$ of numbers $a_1, a_2, \ldots, a_n$ is defined as

$$\bar{a} \equiv \frac{1}{n} \sum_{k=1}^{n} a_k$$

Definition 3.2. The *geometric mean* $\hat{a}$ of numbers $a_1, a_2, \ldots, a_n$ is defined as

$$\hat{a} \equiv \sqrt[n]{\prod_{k=1}^{n} a_k}$$

Having defined the AM and GM, we can introduce the AM-GM inequality. We will state this inequality and show how this inequality can be proven due to the self-generalising nature of $\sqrt{xy} \leq \frac{x+y}{2}$

Theorem 3.1. ***AM-GM inequality.*** *Let $(a_k)_{k=1}^n$ be a sequence of non-negative real numbers. Then:*

$$\sqrt[n]{\prod_{k=1}^{n} a_k} \leq \frac{1}{n} \sum_{k=1}^{n} a_k$$

Proof. We have already seen that $\sqrt{a_1 a_2} \leq \frac{a_1+a_2}{2}$. We claim that $\forall n \in \mathbb{N}: \prod_{k=1}^{2^n} a_k^{\frac{1}{2^n}} \leq \frac{\sum_{k=1}^{2^n} a_k}{2^n}$. We can prove this claim using induction. The first induction step has already been proved, thus:

$\mathbf{n} \implies \mathbf{n+1}$. Assume $\prod_{k=1}^{2^n} a_k^{\frac{1}{2^n}} \leq \frac{1}{2^n}\sum_{k=1}^{2^n} a_k$ for some $n \in \mathbb{N}$. Then:

$$\prod_{k=1}^{2^{n+1}} a_k^{\frac{1}{2^{n+1}}} = \sqrt{\left(\prod_{k=1}^{2^n} a_k^{\frac{1}{2^n}}\right)\left(\prod_{k=2^n+1}^{2^{n+1}} a_k^{\frac{1}{2^n}}\right)} \leq \frac{\prod_{k=1}^{2^n} a_k^{\frac{1}{2^n}} + \prod_{k=2^n+1}^{2^{n+1}} a_k^{\frac{1}{2^n}}}{2}$$

$$\leq \frac{\frac{1}{2^n}\sum_{k=1}^{2^n} a_k + \frac{1}{2^n}\sum_{k=2^n+1}^{2^{n+1}} a_k}{2} = \frac{1}{2^{n+1}}\sum_{k=1}^{2^{n+1}} a_k$$

Where we used the basis step to obtain the first inequality and the induction hypothesis to prove the second inequality. This thus proves the statement for all 2^k with $k \in \mathbb{N}$.
Now let $n \in \mathbb{N}$. We use the construction presented above to prove the theorem for a general value of $n \neq 2^k \forall k \in \mathbb{N}_0$. Take the first $k \in \mathbb{N}$ such that $n < 2^k$. Denoting the arithmetic mean as $\bar{a} \equiv \frac{1}{n}\sum_{i=1}^{n} a_i$ a new sequence α_i as follows:

$$\alpha_i = \begin{cases} a_i & i \in \{1, 2, \ldots n-1, n\} \\ \bar{a} & i \in \left\{n+1, n+2, \ldots, 2^k-1, 2^k\right\} \end{cases}$$

Applying the AM-GM inequality on this new sequence yields:

$$\left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{2^k}} \bar{a}^{\frac{2^k-n}{2^k}} = \left(\bar{a}^{2^k-n}\prod_{i=1}^{n} a_i\right)^{\frac{1}{2^k}} = \left(\prod_{i=1}^{2^k} \alpha_i\right)^{\frac{1}{2^k}} \leq \frac{1}{2^k}\sum_{i=1}^{2^k} \alpha_i = \frac{\sum_{i=1}^{n} a_i + \left(2^k - n\right)\bar{a}}{2^k} = \frac{2^k\bar{a}}{2^k} = \bar{a}$$

$$\implies \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{2^k}} \bar{a}^{1-\frac{n}{2^k}} \leq \bar{a}$$

$$\iff \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{2^k}} \leq \bar{a}^{\frac{n}{2^k}}$$

$$\iff \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}} \leq \frac{1}{n}\sum_{i=1}^{n} a_i = \bar{a}$$

Which proves the theorem for all $n \in \mathbb{N}$. □

Remark. Note that because the AM-GM inequality strictly follows from the fact that $(\sqrt{a} - \sqrt{b})^2 \geq 0$, we see that we only have equality in the case in which $a_i = a_j \forall i, j$. We will prove this more rigidly for the generalised version.

We now immediately show how the AM-GM inequality can be generalised. The proof of this statement, much to one's surprise, also follows from the self-generalising nature of the inequality we have proved above.

Theorem 3.2. ***Generalised AM-GM inequality.*** *Let $(a_k)_{k=1}^n$ be a sequence of non-negative real numbers and let $\left(p_k\right)_{k=1}^n$ be a non-negative sequence such that $\sum_{k=1}^n p_k = 1$ Then:*

$$\prod_{k=1}^{n} a_k^{p_k} \leq \sum_{k=1}^{n} p_k a_k$$

Proof. We want to prove this theorem for all $p_k \in \mathbb{R}$. Note that we have already constructed the proof for the cases when $p_k = \frac{1}{n} \forall k \in \{1, 2, \ldots, n\}$.
Now suppose $p_k \in \mathbb{Q} \forall k$. We can then find an $N \in \mathbb{N}$ such that each p_k can be written as

$p_k = \frac{c_k}{N}$ with $c_k \in \mathbb{N}$. Now define a new sequence $(\alpha_i)_{i=1}^N$ such that every a_k occurs c_k times. Then:

$$\left(\prod_{i=1}^N \alpha_i\right)^{\frac{1}{N}} \le \frac{1}{N}\sum_{i=1}^N \alpha_i \iff \left(\prod_{i=1}^n a_i^{c_i}\right)^{\frac{1}{N}} \le \frac{1}{N}\sum_{i=1}^n c_i a_i \iff = \prod_{i=1}^n a_i^{p_i} = \left(\prod_{i=1}^n a_i^{\frac{c_i}{N}}\right) \le \sum_{i=1}^n \frac{c_i}{N} a_i = \sum_{i=1}^n p_i a_i$$

Which proves the theorem for all $n \in \mathbb{Q}$.
We thus have to prove the theorem for $p_k \in \mathbb{R}^n$. Note, however, that we can always define a sequence of vectors $\big(p_1(t), p_2(t), \dots, p_n(t)\big)$, such that $\forall t \in \mathbb{N} : \forall k \in \{1, 2, \dots, n\} : p_k(t) \in \mathbb{Q}_{\ge 0}$, such that $\sum_{k=1}^n p_k(t) = 1$ and $\lim_{t\to\infty}\big(p_1(t), p_2(t), \dots, p_n(t)\big) = \big(p_1, p_2, \dots, p_n\big)$. But since:

$$\prod_{k=1}^n a_k^{p_k(t)} \le \sum_{k=1}^n p_k(t) a_k \forall t \in \mathbb{N}$$

We can take limits to find that:

$$\prod_{k=1}^n a_k^{p_k} = \lim_{t\to\infty}\prod_{k=1}^n a_k^{p_k(t)} \le \lim_{t\to\infty}\sum_{k=1}^n p_k(t) a_k = \sum_{k=1}^n p_k a_k$$

□

While the self-generalising nature of the AM-GM inequality follows beautifully from the induction proof, we can also prove this inequality in a different way: this proof was suggested by the mathematician Pólya, who claimed to have realised this proof in one of his dreams. (see: Steele (2004)) Before doing so, let us first prove the following lemma.

Lemma 3.3. *For all $x \in \mathbb{R}$, the following inequality holds:*

$$1 + x \le e^x$$

Proof. We consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = e^x - x - 1$. We see that $\frac{\mathrm{d}}{\mathrm{d}x} f(x) = e^x - 1$. Thus $f'(x) = 0$ for $x = 0$. Now since $f''(x) = e^x > 0$, we see that $f(x)$ is a strictly decreasing function on $(\leftarrow, 0)$ and strictly increasing on $(0, \rightarrow)$, and thus attains a minimum at $x = 0$. But $f(0) = 0$. So $\forall x \ne 0 : f(x) > 0$. □

Proof. **Pólya.** We note that through a shift of coordinates $x \to x - 1$, Lemma 3.3 yields:

$$x \le e^{x-1}$$

using this translated version of Lemma 3.3, for any sequence of non-negative numbers $(\alpha_k)_{k=1}^n$, we obtain:

$$\alpha_k \le e^{\alpha_k - 1} \implies \alpha_k^{p_k} \le e^{p_k\alpha_k - p_k} \implies \prod_{k=1}^n \alpha_k^{p_k} \le e^{\sum_{k=1}^n p_k\alpha_k - \sum_{k=1}^n p_k} = e^{\sum_{k=1}^n p_k\alpha_k - 1}$$

Taking $\alpha_k = \frac{a_k}{\sum_{i=1}^n p_i a_i}$, so a normalised version of the sequence a_k, we find:

$$\prod_{k=1}^n \left(\frac{a_k}{\sum_{i=1}^n p_i a_i}\right)^{p_k} = \prod_{k=1}^n \alpha_k^{p_k} \le e^{\sum_{k=1}^n p_k\alpha_k - 1} = e^0 = 1 \implies \prod_{k=1}^n a_k^{p_k} \le \sum_{k=1}^n p_k a_k$$

□

Remark. Note that we only have equality if we have $\alpha_k = e^{\alpha_k - 1}$, which only holds if $\alpha_k = 1$, so if $a_k = \sum_{i=1}^n p_i\alpha_i$. But this should hold for all k! So then we have equality if and only if $\alpha_i = \alpha_j \forall i, j$.

Lastly, we present a proof that at first instance looks different, but can nevertheless still be interpreted as a variation on the proof by Pólya.

Proof. Note that by taking the logarithm of both sides, Lemma 3.3 yields for $x > 0$:

$$x - 1 \leq \ln(x)$$

Furthermore, note that:

$$\prod_{k=1}^{n} (\lambda a_k)^{p_k} = \lambda \prod_{k=1}^{n} a_k^{p_k}$$
$$\sum_{k=1}^{n} p_k (\lambda a_k) = \lambda \sum_{k=1}^{n} p_k a_k$$

Taking $\lambda \geq 0$, it suffices to prove that

$$\lambda \prod_{k=1}^{n} a_k^{p_k} \leq \sum_{k=1}^{n} p_k(\lambda a_k)$$

Taking $\lambda = \frac{1}{\prod_{k=1}^{n} a_k^{p_k}}$, this reduces to:

$$\sum_{k=1}^{n} p_k(\lambda a_k) \geq 1$$

But now:

$$\begin{aligned}\sum_{k=1}^{n} p_k(\lambda a_k) - 1 = \sum_{k=1}^{n} p_k(\lambda a_k - 1) \geq \sum_{k=1}^{n} p_k \ln(\lambda a_k) = \sum_{k=1}^{n} \ln\left((\lambda a_k)^{p_k}\right) \\ = \ln\left(\prod_{k=1}^{n} \left(\frac{a_k}{\prod_{i=1}^{n} a_i^{p_i}}\right)^{p_k}\right) = \ln\left(\prod_{k=1}^{n} \left(\frac{a_k^{p_k}}{(\prod_{i=1}^{n} a_i^{p_i})^{p_k}}\right)\right) \\ = \ln\left(\frac{\prod_{k=1}^{n} a_k^{p_k}}{\prod_{i=1}^{n} a_i^{p_i}}\right) = 0\end{aligned}$$

□

3.2. PROPERTIES AND A REFINEMENT OF THE AM-GM INEQUALITY

In this section, we consider one interesting property of the geometric mean and then present a different, more tight bound on the arithmetic mean.

3.2.1. THE QUASI-ADDITIVITY OF THE GEOMETRIC MEAN

We now prove an important property of the geometric mean - its quasi-additivity. We first consider the case of two sequences.

Theorem 3.4. *Let $(a_k)_{k=1}^n, (b_k)_{k=1}^n$ be sequences of non-negative real numbers, then:*

$$\sqrt[n]{\prod_{k=1}^{n} a_k} + \sqrt[n]{\prod_{k=1}^{n} b_k} \leq \sqrt[n]{\prod_{k=1}^{n} (a_k + b_k)}$$

Proof. We first rewrite the theorem to:

$$\frac{\sqrt[n]{\prod_{k=1}^{n} a_k} + \sqrt[n]{\prod_{k=1}^{n} b_k}}{\sqrt[n]{\prod_{k=1}^{n} (a_k + b_k)}} \leq 1$$

We can rewrite the left side of the equation:

$$\begin{aligned}\frac{\sqrt[n]{\prod_{k=1}^{n} a_k}}{\sqrt[n]{\prod_{k=1}^{n} (a_k + b_k)}} + \frac{\sqrt[n]{\prod_{k=1}^{n} b_k}}{\sqrt[n]{\prod_{k=1}^{n} (a_k + b_k)}} &= \sqrt[n]{\frac{\prod_{k=1}^{n} a_k}{\prod_{k=1}^{n} (a_k + b_k)}} + \sqrt[n]{\frac{\prod_{k=1}^{n} b_k}{\prod_{k=1}^{n} (a_k + b_k)}} \\ &= \sqrt[n]{\prod_{k=1}^{n} \frac{a_k}{(a_k + b_k)}} + \sqrt[n]{\prod_{k=1}^{n} \frac{b_k}{(a_k + b_k)}}\end{aligned}$$

We can now use the original AM-GM inequality (Theorem 3.1) on both summands to obtain:

$$\sqrt[n]{\prod_{k=1}^{n} \frac{a_k}{(a_k + b_k)}} + \sqrt[n]{\prod_{k=1}^{n} \frac{b_k}{(a_k + b_k)}} \leq \frac{1}{n} \sum_{k=1}^{n} \left(\frac{a_k}{a_k + b_k}\right) + \frac{1}{n} \sum_{k=1}^{n} \left(\frac{b_k}{a_k + b_k}\right) = \frac{1}{n} \sum_{k=1}^{n} \left(\frac{a_k + b_k}{a_k + b_k}\right) = \frac{1}{n} \sum_{k=1}^{n} 1 = 1$$

□

We thus see that this property of the geometric mean (much to one's surprise) simply follows from its relation to the arithmetic mean!
This is not just restricted to the case of two sequences - this property is still valid when considering an arbitrary number of sequences. Let us proceed by proving this.

Theorem 3.5. *Let* $(x_k^1)_{k=1}^n, (x_k^2)_{k=1}^n, \ldots, (x_k^p)_{k=1}^n$ *be sequences of non-negative real numbers. Then:*

$$\sum_{p=1}^{m} \sqrt[n]{\prod_{k=1}^{n} x_k^p} \leq \sqrt[n]{\prod_{k=1}^{n} \sum_{p=1}^{m} x_k^p}$$

Proof. We first note that the theorem is equivalent to proving that:

$$\frac{\sum_{p=1}^{m} \sqrt[n]{\prod_{k=1}^{n} x_k^p}}{\sqrt[n]{\prod_{k=1}^{n} \left(\sum_{p=1}^{m} x_k^p\right)}} \leq 1$$

We consider the left-hand side of this inequality:

$$\frac{\sum_{p=1}^{m} \sqrt[n]{\prod_{k=1}^{n} x_k^p}}{\sqrt[n]{\prod_{k=1}^{n} \left(\sum_{p=1}^{m} x_k^p\right)}} = \sum_{p=1}^{m} \sqrt[n]{\frac{\prod_{k=1}^{n} x_k^p}{\prod_{k=1}^{n} \left(\sum_{p=1}^{m} x_k^p\right)}} = \sum_{p=1}^{m} \sqrt[n]{\prod_{k=1}^{n} \frac{x_k^p}{\sum_{p=1}^{m} x_k^p}}$$

Using Theorem 3.1, we find an upper bound for each of the summands:

$$\sum_{p=1}^{m} \sqrt[n]{\prod_{k=1}^{n} \frac{x_k^p}{\sum_{p=1}^{m} x_k^p}} \leq \sum_{p=1}^{m} \left(\frac{1}{n} \sum_{k=1}^{n} \frac{x_k^p}{\sum_{p=1}^{m} x_k^p}\right) = \frac{1}{n} \sum_{k=1}^{n} \left(\frac{\sum_{p=1}^{m} x_k^p}{\sum_{p=1}^{m} x_k^p}\right) = \frac{1}{n} \sum_{k=1}^{n} 1 = 1$$

□

3.2.2. A REFINEMENT OF THE INEQUALITY

Note that we can find a better lower bound for the arithmetic than the one found in Theorem 3.1. Before introducing this refinement, we first consider the following lemma, which will prove to be helpful in proving our refinement:

Lemma 3.6. *Let $x \geq 0$ and $n \in \mathbb{N}_{n\geq 2}$. Then:*

$$x\left(n - x^{n-1}\right) \leq n-1$$

Proof. We will prove the theorem by induction.
Consider the functions $f_n : \mathbb{R}_{\geq 0} \to \mathbb{R}$ such that $f_n(x) = x\left(n - x^{n-1}\right) - n + 1$. We thus have to prove that f_n is non-positive for all n.
n = 2. We obtain $f_2(x) = -x^2 + 2x - 1$. The discriminant is $D = 2^2 - 4 \cdot (-1) \cdot (-1) = 0$, such that the extremum of this parabola coincides with its only zero. From elementary mathematics, we know that this extremum is also its maximum. This thus proves the induction step.
n $\implies$ n + 1. Assume the induction hypothesis is valid for some n, so $f_n = x\left(n - x^{n-1}\right) + 1 - n \leq 0$. Then we can multiply both sides by x:

$$\begin{aligned}
& nx^2 - x^{n+1} + x - nx \leq 0 \\
\iff & nx^2 + nx - x^{n+1} + x - 2nx \leq 0 \\
\iff & nx^2 + nx - x^{n+1} + x - 2nx + n - n \leq 0 \\
\iff & x(n+1-x^n) - n + n(x-1)^2 \leq 0 \\
\implies & f_{n+1}(x) = x(n+1-x^n) - n \leq 0
\end{aligned}$$

$\square$

Theorem 3.7. ***Åkerberg's Refinement.*** *Let $(a_k)_{k=1}^n$ be a sequence of non-negative real numbers and let $n \geq 2$. Then:*

$$a_n \left(\frac{1}{n-1}\sum_{k=1}^{n-1} a_k\right)^{n-1} \leq \left(\frac{1}{n}\sum_{k=1}^{n} a_k\right)^n$$

Proof. We first note that

$$a_n = \sum_{k=1}^{n} a_k - \sum_{k=1}^{n-1} a_k = \frac{n}{n}\sum_{k=1}^{n} a_k - \frac{n-1}{n-1}\sum_{k=1}^{n-1} a_k$$

Thus the inequality reduces to:

$$\begin{aligned}
& \left(n \cdot \frac{1}{n}\sum_{k=1}^{n} a_k - \frac{n-1}{n-1}\sum_{k=1}^{n-1} a_k\right)\left(\frac{1}{n-1}\sum_{k=1}^{n-1} a_k\right)^{n-1} \leq \left(\frac{1}{n}\sum_{k=1}^{n} a_k\right)^n \\
\iff & n \cdot \frac{1}{n}\sum_{k=1}^{n} a_k \left(\frac{1}{n-1}\sum_{k=1}^{n-1} a_k\right)^{n-1} - (n-1)\left(\frac{1}{n-1}\sum_{k=1}^{n-1} a_k\right)^n \leq \left(\frac{1}{n}\sum_{k=1}^{n} a_k\right)^n \\
\iff & n \cdot \frac{1}{n}\sum_{k=1}^{n} a_k \left(\frac{1}{n-1}\sum_{k=1}^{n-1} a_k\right)^{n-1} - \left(\frac{1}{n}\sum_{k=1}^{n} a_k\right)^n \leq (n-1)\left(\frac{1}{n-1}\sum_{k=1}^{n-1} a_k\right)^n \\
\iff & n\left(\frac{\frac{1}{n}\sum_{k=1}^n a_k}{\frac{1}{n-1}\sum_{k=1}^{n-1} a_k}\right) - \left(\frac{\frac{1}{n}\sum_{k=1}^n a_k}{\frac{1}{n-1}\sum_{k=1}^{n-1} a_k}\right)^n \leq n-1 \\
\iff & \left(\frac{\frac{1}{n}\sum_{k=1}^n a_k}{\frac{1}{n-1}\sum_{k=1}^{n-1} a_k}\right)\left(n - \left(\left(\frac{\frac{1}{n}\sum_{k=1}^n a_k}{\frac{1}{n-1}\sum_{k=1}^{n-1} a_k}\right)\right)^{n-1}\right) \leq n-1
\end{aligned}$$

Setting $x \equiv \left(\frac{\frac{1}{n}\sum_{k=1}^{n} a_k}{\frac{1}{n-1}\sum_{k=1}^{n-1} a_k}\right)$, this reduces to the simple inequality

$$x\left(n - x^{n-1}\right) \leq n-1, \qquad x \geq 0, n \in \mathbb{N}_{\geq 2}$$

which was proved in Lemma 3.6. □

The proof that this bound is more tight can be established very easily through inductive reasoning - we can use the same inequality to prove that:

$$\prod_{k=1}^{n} a_k \leq \prod_{k=3}^{n} a_k \left(\frac{1}{2}(a_2 + a_1)\right)^2 \leq \ldots \leq a_n a_{n-1}\left(\frac{1}{n-2}\sum_{k=1}^{n-2} a_k\right)^{n-2} \leq a_n \left(\frac{1}{n-1}\sum_{k=1}^{n-1} a_k\right)^{n-1} \leq \frac{1}{n}\sum_{k=1}^{n} a_k$$

which yields our AM-GM inequality.

3.3. A COMPLEX EXTENSION

We can now extend the AM-GM to the complex numbers with some slight adaptions.

Theorem 3.8. *Given complex numbers $z_1, z_2, \ldots z_n$ and an angle $\psi < \frac{\pi}{2}$ such that $\forall \theta_k = Arg(z_k) : |\theta_k| < \psi$. Furthermore, assume $|z_n| < \infty\ \forall n$. Then:*

$$\cos(\psi)|\prod_{k=1}^{n} z_k|^{\frac{1}{n}} \leq \frac{1}{n}|\sum_{k=1}^{n} z_k|$$

Proof. Note that since $\psi \in \left(0, \frac{\pi}{2}\right)$, we have that $\cos(\psi) > 0$ and thus:

$$\cos(\psi)|\prod_{k=1}^{n} z_k|^{\frac{1}{n}} = |\prod_{k=1}^{n} z_k \cos^n(\psi)|^{\frac{1}{n}} \leq |\prod_{k=1}^{n} z_k \cos(\theta_k)|^{\frac{1}{n}} \stackrel{*}{=} |\prod_{k=1}^{n} \mathrm{Re}\{z_k\}|^{\frac{1}{n}}$$

where the inequality follows from the fact that the cos is an even function and decreasing on $\left[0, \frac{\pi}{2}\right]$, and equality $*$ follows from the fact that $\mathrm{Re}\{z_k\} = \mathrm{Re}\left\{|z_k|e^{i\theta_k}\right\} = \mathrm{Re}\{|z_k|(\cos(i\theta_k) + i\sin(\cos(i\theta_k)))\} = |z_k|\cos(i\theta_k)$. We thus find:

$$|\prod_{k=1}^{n} \mathrm{Re}\{z_k\}|^{\frac{1}{n}} \leq \frac{1}{n}\sum_{k=1}^{n} \mathrm{Re}\{z_k\} = \frac{1}{n}\mathrm{Re}\left\{\sum_{k=1}^{n} z_k\right\}$$

We now find:

$$\frac{1}{n}\mathrm{Re}\left\{\sum_{k=1}^{n} z_k\right\} \leq \frac{1}{n}\sqrt{\mathrm{Re}\left\{\sum_{k=1}^{n} z_k\right\}^2} \leq \frac{1}{n}\sqrt{\mathrm{Re}\left\{\sum_{k=1}^{n} z_k\right\}^2 + \mathrm{Im}\left\{\sum_{k=1}^{n} z_k\right\}^2} = \frac{1}{n}|\sum_{k=0}^{n} z_k|$$

□

3.4. AM-GM IN RELATION TO CAUCHY-SCHWARZ

Note that since both the Cauchy-Schwarz inequality as well as the AM-GM inequality follow from the simple fact that $(a-b)^2 \geq 0$, these inequalities are obviously very much related. In the first proof of the Cauchy-Schwarz inequality, the step:

$$\sum_{k=1}^{n} a_k b_k \leq \frac{1}{2}\sum_{k=1}^{n} a_k^2 + \frac{1}{2}b_k^2$$

follows simply from the AM-GM inequality. Note that this inequality is true if it holds for all components, so if:

$$a_k b_k \leq \frac{1}{2}a_k^2 + \frac{1}{2}b_k^2$$

holds for all k, which is equal to the requirement that:

$$\sqrt{a_k^2 b_k^2} \leq \frac{a_k^2 + b_k^2}{2}$$

which is just the AM-GM inequality. So we see by merely assuming the AM-GM inequality, we can derive the Cauchy-Schwarz inequality.

Alternatively, We can also derive the AM-GM inequality from the Cauchy Schwarz inequality. Let us consider the following two vectors in $\mathbb{R}^2$: $\mathbf{x} = \begin{bmatrix}\sqrt{a} \\ \sqrt{b}\end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix}\sqrt{b} \\ \sqrt{a}\end{bmatrix}$. Applying the Cauchy-Schwarz inequality yields:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sqrt{ab} + \sqrt{ab} = 2\sqrt{ab} \leq \left(\sqrt{\sqrt{a}^2 + \sqrt{b}^2}\right)^2 = a + b$$

which yields the AM-GM inequality for two sequences, and can be generalised to the full inequality using the self-generalising property of the AM-GM inequality.

We thus see that the AM-GM inequality and the Cauchy-Schwarz inequality can be derived from the other inequality.

4

Convexity and Jensen's Inequality

We now switch to a seemingly different topic which is also an important tool in proving inequalities and deriving upper bounds: convexity. Inseparable from convexity is another famous inequality, namely Jensen's inequality. In this chapter, we will prove this inequality, delve deeper into some interesting properties of convexity and then see how this seemingly unrelated topic can be connected to the Cauchy-Schwarz and AM-GM inequality.

4.1. Introducing Convexity and Jensen's Inequality

We start off by defining convexity.

Definition 4.1. A function $f:[a,b]\to\mathbb{R}$ is called *convex* if it satisfies:

$$f\left(px+(1-p)y\right)\le pf(x)+(1-p)f(y)\quad \forall x,y\in[a,b],\forall p\in[0,1]$$

Let us now consider Jensen's inequality and show how this inequality follows utmost elegantly from the definition of convexity.

Theorem 4.1. ***Jensen's Inequality.*** *Let $f:[a,b]\to\mathbb{R}$ be convex and let $\left(p_i\right)_{i=1}^{n}$ such that $\forall i:$ $p_i\ge 0$ and $\sum_{i=1}^{n}p_i=1$. Then $\forall x_i\in[a,b]$, f satisfies:*

$$f\left(\sum_{i=1}^{n}p_ix_i\right)\le\sum_{i=1}^{n}p_if\left(x_i\right)$$

Proof.

$$\begin{aligned}f\left(\sum_{i=1}^{n}p_ix_i\right)&=f\left(p_nx_n+(1-p_n)\sum_{i=1}^{n-1}\frac{p_i}{1-p_n}x_i\right)\\&\le p_nf(x_n)+(1-p_n)\sum_{i=1}^{n-1}\frac{p_i}{1-p_n}f\left(x_i\right)=\sum_{i=1}^{n}p_if\left(x_i\right)\end{aligned}$$

□

One may wonder what the conditions are for equality. We will prove that these conditions are relatively strict: one only has equality if all x_i are equal.

Theorem 4.2. *Theorem 4.1 only yields equality if and only if all variables are the same , i.e.:*

$$f\left(\sum_{i=1}^{n}p_ix_i\right)=\sum_{i=1}^{n}p_if\left(x_i\right)\iff x_1=x_2=\ldots=x_n$$

Proof.
$\Longrightarrow$. We prove this implication by contradiction. Assume we have equality for some $x_1, \dots, x_n$ which are not all equal. Set $S \equiv \left\{ i \mid x_i < \max_{k \in \{1,\dots,n\}} \{x_k\} \right\}$. Note that S is a *proper* subset of $\{1, 2, \dots, n\}$. Now define:

$$p = \sum_{i \in S} p_i$$
$$x = \frac{1}{p} \sum_{i \in S} p_i x_i$$
$$y = \frac{1}{1-p} \sum_{i \notin S} p_i x_i$$

Now since we have strict convexity, we find:

$$f(px + (1-p)y) < pf(x) + (1-p)f(y)$$

Now we can substitute x and y:

$$pf(x)+(1-p)f(y) = pf\left(\sum_{i \in S} \frac{p_i}{p} x_i\right)+(1-p)f\left(\sum_{i \notin S} \frac{p_i}{1-p} x_i\right) \leq \sum_{i \in S} p_i f(x_i)+\sum_{i \notin S} p_i f(x_i) = \sum_{k=1}^{n} p_k f(x_k)$$

Note that $f(px + (1-p)y) = f\left(\sum_{k=1}^{n} p_k x_k\right)$. But then:

$$f\left(\sum_{k=1}^{n} p_k x_k\right) < \sum_{k=1}^{n} p_k f(x_k) \tag{4.1}$$

But we assumed equality. $\Rightarrow\Leftarrow$.

$\Leftarrow$. Furthermore, note that if $x_1 = x_2 + \ldots = x_n = x$, then:

$$f\left(\sum_{k=1}^{n} p_k x_k\right) = f\left(x \sum_{k=1}^{n} p_k\right) = f(x) = \sum_{k=1}^{n} p_k f(x) = \sum_{k=1}^{n} p_k f(x_k)$$

□

4.2. CONNECTION TO THE AM-GM AND CAUCHY-SCHWARZ INEQUALITY

The concept of convexity will prove to be exceptionally useful in proving various inequalities. One such inequality is the AM-GM inequality (Theorem 3.2). We can easily proof this Theorem using the convexity of exponential functions.

Proof. First note that $f : x \to e^x$ is convex on $\mathbb{R}$. Thus:

$$e^{\sum_{k=1}^{n} p_k \hat{x}_k} \leq \sum_{k=1}^{n} p_k e^{\hat{x}_k}$$

Setting $x_k = e^{\hat{x}_k}$ yields:

$$\prod_{k=1}^{n} x_k^{p_k} \leq \sum_{k=1}^{n} p_k x_k$$

□

We see that the AM-GM gracefully follows from Jensen's inequality, which again is a direct implication of the definition of convexity! In retrospect, however, this should not be very surprising, since the most simple form of the AM-GM inequality, namely $\sqrt{xy} \le \frac{x+y}{2}$, is just a simple consequence of the convexity of $x \mapsto x^2$! Taking $f(x) = 4x^2$ and $x_1 = \sqrt{x}, x_2 = \sqrt{y}$ and $p_1 = p_2 = \frac{1}{2}$, we see that:

$$f\left(\frac{1}{2}(x_1 + x_2)\right) = (\sqrt{x} + \sqrt{y})^2 \le 2x + 2y = \frac{1}{2}\left(f(x_1) + f(x_2)\right)$$

working out the squares on the left hand side and rearranging the terms yields the AM-GM inequality.

Evidently, since the AM-GM inequality and the Cauchy-Schwarz inequality are so intimately related via exactly the relation which holds because of convexity, the Cauchy-Schwarz inequality can also be derived via Jensen's inequality by the exact argument we've presented above!

4.3. THE STRENGTH OF CONVEXITY

The beautiful structure that convexity brings is not just restricted to the main topic of this exploration - rather, it can be shown that convexity is a fairly *strong* property, in the sense that convexity implies quite some smoothness of a given function. We will prove this first by showing that convex functions are necessarily continuous, and then by showing that convex functions are even more smooth, in the sense that they even possess left and right derivatives. Before doing so, however, we will have to introduce a very important lemma first.

Lemma 4.3. *Let $f : [a,b] \to \mathbb{R}$ be convex. Furthermore, consider $x \in (a,b)$. Then:*

$$\frac{f(x) - f(a)}{x - a} \le \frac{f(b) - f(a)}{b - a} \le \frac{f(b) - f(x)}{b - x}$$

Proof. Note that $\exists p : x = pb + (1-p)a$. Then since $f(x)$ is convex:

$$f(x) = f(pb + (1-p)a) \le pf(b) + (1-p)f(a)$$

It can be checked that $p = \frac{x-a}{b-a}$ satisfies the first equality. Then:

$$\begin{aligned} f(x) &\le \left(\frac{x-a}{b-a}\right)f(b) + \left(1 - \left(\frac{x-a}{b-a}\right)\right)f(a) \\ \iff f(x) - f(a) &\le \left(\frac{x-a}{b-a}\right)\left(f(b) - f(a)\right) \\ \iff \frac{f(x) - f(a)}{x - a} &\le \frac{f(b) - f(a)}{b - a} \end{aligned}$$

Thus proving the first inequality.

Similarly, taking $p = \frac{b-x}{b-a}$, we find:

$$f(x) = f(pa + (1-p)b) \le pf(a) + (1-p)f(b)$$

after substitution, this yields:

$$\begin{aligned} f(x) - f(b) &\le \left(f(a) - f(b)\right)\left(\frac{b-x}{b-a}\right) \\ \iff \frac{f(b) - f(a)}{b - a} &\le \frac{f(b) - f(x)}{b - x} \end{aligned}$$

proving the second inequality. □

We can use this lemma to prove that a convex function is always continuous on (a,b).

Theorem 4.4. *Let $f:(a,b)\to\mathbb{R}$ be convex. Then $f(x)$ is continuous.*

Proof. We will prove that $f(x)$ is *locally Lipschitz continuous* in any point $c \in (a,b)$. Let $[a_1,b_1] \subseteq (a,b)$ and choose a_2, b_2 such that $a < a_2 < a_1, b_1 < b_2 < b$. Now let $x, y \in [a_1,b_1]$. Then, using Lemma 4.3, we find:

$$\frac{f(x)-f(y)}{x-y} \le \frac{f(b_2)-f(y)}{b_2-y} \le \frac{f(b_1)-f(b_2)}{b_1-b_2}$$

Similarly, using Lemma 4.3 once again, we find:

$$\frac{f(a_2)-f(a_1)}{a_2-a_1} \le \frac{f(y)-f(a_1)}{y-a_1} \le \frac{f(x)-f(y)}{x-y}$$

Setting $M \equiv \max\left\{\left|\frac{f(a_2)-f(a_1)}{a_2-a_1}\right|, \left|\frac{f(b_1)-f(b_2)}{b_1-b_2}\right|\right\}$, we see that:

$$\left|\frac{f(x)-f(y)}{x-y}\right| \le \frac{f(b_1)-f(b_2)}{b_1-b_2}$$

Thus $\left|f(x)-f(y)\right| \le M|x-y|$. □

We can prove that convex function possess an even stronger property: near-differentiability.

Theorem 4.5. *Let $f:[a,b]\to\mathbb{R}$ be convex. Then the left and right-hand derivative of f exist on (a,b).*

Proof. Let $x \in (a,b)$ fixed. We first use Lemma 4.3 to find that:

$$\frac{f(x+h)-f(x)}{h} \le \frac{f(x+h+\epsilon)-f(x)}{h+\epsilon}$$

with $\epsilon > 0$. Thus, we find that $\frac{f(x+h)-f(x)}{h}$ is a decreasing function of h. Furthermore, for any fixed value $y \in (a,x)$ Lemma 4.3 yields:

$$\frac{f(x)-f(y)}{x-y} \le \frac{f(x+h)-f(x)}{h}$$

for all $h > 0$. Note that the left-hand side of the inequality is definitely bounded. Thus, we see that $F(h) \equiv \frac{f(x+h)-f(x)}{h}$ is a monotone, decreasing function that is also bounded. The Monotone Convergence Theorem now ensures us that the limit for $h \downarrow 0$ exists, i.e.

$$\lim_{h\downarrow 0} \frac{f(x+h)-f(x)}{h}$$

exists and is finite.

In a similar manner, we can prove that:

$$\frac{f(x-h-\epsilon)-f(x)}{h+\epsilon} \le \frac{f(x-h)-f(x)}{h} \le \frac{f(y)-f(x)}{y-x} \tag{4.2}$$

with $y \in (x, b)$. This yields a monotone, increasing and bounded sequence, and thus, once again using the Monotone Convergence Theorem, we see that:

$$\lim_{h \uparrow 0} \frac{f(x-h) - f(x)}{h}$$

exists and is finite. $\square$

We thus see that convexity (through Jensen's inequality) allows for the emergence of a lot of rich structure, both in terms of smoothness of functions and in terms of the inequalities that can be derived by these functions. We now proceed to study three more famous inequalities, of which the inequalities studied heretofore are just a special case, and we will show how these inequalities are nevertheless closely related to those we have derived up until now.

5

The Ladder of Power Means

Hitherto, we have seen two means: the arithmetic and the geometric mean. These means, however, belong to a more general class of means, for which numerous important properties can be proved. This class is called the class of the power means. *power means.* In this chapter, we first define these power means and then prove an inequality on this class which can be considered a generalisation of the AM-GM inequality.

5.1. The Power Means

We start off by defining power means.

Definition 5.1. M_t with $t \in \mathbb{R}$ for $\left(p_k\right)_{k=1}^n$ such that $p_k \geq 0$ and $\sum_{k=1}^n p_k = 1$ and a sequence of non-negative numbers $(x_k)_{k=1}^n$ is defined as follows:

$$M_t = \left(\sum_{k=1}^n p_k x_k^t\right)^{\frac{1}{t}}$$

We call M_t a *power mean* of power t. Note that for $t = 0$, the definition of M_t is unclear. Furthermore, the definition can be extended to include the cases for $t \to \pm\infty$. The definition for these cases follows from the natural requirement that the mapping $t \to M_t$ should be continuous. Let us first consider the case for $t = 0$.

Theorem 5.1. *For $t \to 0$ M_t reduces to the geometric mean, i.e.*

$$\lim_{t\to 0} M_t = \prod_{k=1}^n x_k^{p_k}$$

Proof. Taking the logarithm of M_t yields:

$$\log(M_t) = \log\left(\sum_{k=1}^n p_k x_k^t\right)^{\frac{1}{t}} = \frac{1}{t}\log\left(\sum_{k=1}^n p_k x_k^t\right) = \frac{1}{t}\log\left(\sum_{k=1}^n p_k e^{t\log(x_k)}\right)$$

The exponential function can be expanded using a Taylor polynomial:

$$\log(M_t) = \frac{1}{t}\log\left(\sum_{k=1}^n p_k e^{t\log(x_k)}\right) = \frac{1}{t}\log\left(\sum_{k=1}^n p_k\left(1 + t\log(x_k) + \mathcal{O}(t^2)\right)\right) = \frac{1}{t}\log\left(1 + t\sum_{k=1}^n p_k \log(x_k) + \mathcal{O}(t^2)\right)$$

Now, using a Taylor expansion for the logarithm (i.e. $\log(1+x)$), we find:

$$\log(M_t) = \frac{1}{t}\log\left(1 + t\sum_{k=1}^n p_k \log(x_k) + \mathcal{O}(t^2)\right) = \frac{1}{t}\left(t\sum_{k=1}^n p_k \log(x_k) + \mathcal{O}(t^2)\right) = \sum_{k=1}^n p_k \log(x_k) + \mathcal{O}(t)$$

Taking the limit as $t \to 0$ on both sides and inverting the logarithm yields the expected result. □

This brings us to the most important inequality of this section: the *Power Mean Inequality*.

Theorem 5.2. ***Power Mean Inequality.*** *Consider a sequence* $(p_k)_{k=1}^n$ *with* $p_k \geq 0$ *and* $\sum_{k=1}^n p_k = 1$ *and a sequence* $(x_k)_{k=1}^n$ *of non-negative numbers. Then* $\forall s, t \in \mathbb{R} : -\infty < s < t < \infty$*:*

$$M_s = \left(\sum_{k=1}^n p_k x_k^s\right)^{\frac{1}{s}} \leq \left(\sum_{k=1}^n p_k x_k^t\right)^{\frac{1}{t}} = M_t$$

Proof. We will split this problem up into three cases: the case for $0 < s < t$, the case for $s < t < 0$ and the case for $s \leq 0 \leq t$.

Case I: $0 < s < t$. Consider the function $f : x \to x^{\frac{t}{s}}$. Note that since $t > s$, f is a convex function. Thus, Jensen's inequality (Theorem 4.1) yields:

$$f\left(\sum_{k=1}^n p_k \hat{x}_k\right) \leq \sum_{k=1}^n p_k f(\hat{x}_k)$$

Setting $\hat{x}_k = x_k^s$ now yields:

$$\left(\sum_{k=1}^n p_k x_k^s\right)^{\frac{t}{s}} \leq \sum_{k=1}^n p_k \left(x_k^s\right)^{\frac{t}{s}} = \sum_{k=1}^n p_k (x_k)^t$$

Raising both sides to the power of $\frac{1}{t}$ yields the expected result.

Case II: $s < t < 0$. Note that the result from case 1 yields that:

$$\frac{1}{\left(\sum_{k=1}^n p_k \hat{x}_k^{-t}\right)^{\frac{1}{t}}} = \left(\sum_{k=1}^n p_k \hat{x}_k^{-t}\right)^{-\frac{1}{t}} \leq \left(\sum_{k=1}^n p_k \hat{x}_k^{-s}\right)^{-\frac{1}{s}} = \frac{1}{\left(\sum_{k=1}^n p_k \hat{x}_k^{-s}\right)^{\frac{1}{s}}}$$

since $0 < -t < -s$. By taking the reciprocal, we obtain:

$$\left(\sum_{k=1}^n p_k \hat{x}_k^{-s}\right)^{\frac{1}{s}} \leq \left(\sum_{k=1}^n p_k \hat{x}_k^{-t}\right)^{\frac{1}{t}}$$

Setting $\hat{x}_k = \frac{1}{x_k}$ yields the expected result.

Case III: $s \leq 0 \leq t$. This problem can be split up into two parts: proving that $M_0 \leq M_t$ for $t > 0$ and $M_s \leq M_0$.

First note that case I yields that: $M_{\frac{1}{2^{n+1}}} \leq M_{\frac{1}{2^n}}$ for all $n \in \mathbb{N}$. Thus. Note, however, that for every $t \in \mathbb{R}_{>0} : \exists N \in \mathbb{N} : \frac{1}{2^N} \leq t$. Thus:

$$M_{\frac{1}{2^N}} \leq M_t$$

Taking the limit for $N \to \infty$ yields the expected result.

The second case is covered by considering $-s$:

$$\left(\sum_{k=1}^n p_k \hat{x}_k^{-s}\right)^{-\frac{1}{s}} = \frac{1}{\left(\sum_{k=1}^n p_k \hat{x}_k^{-s}\right)^{\frac{1}{s}}} \geq \prod_{k=1}^n \hat{x}_k^{p_k}$$

Taking reciprocals yields:

$$\left(\sum_{k=1}^{n} p_k \hat{x}_k^{-s}\right)^{\frac{1}{s}} \leq \prod_{k=1}^{n} \hat{x}_k^{-p_k}$$

Setting $\hat{x}_k = \frac{1}{x_k}$ finishes the proof. □

Note how this inequality between power means is also a simple consequence of Jensen's inequality! We can therefore easily see that we only find a case of equality if we have that $x_1 = x_2 = \dots x_n$ as this is the only possibility for us to have equality in the first case, and every other case follows from that.

Note, however, that an analytical extension of the concept of a power mean for $t \to \pm\infty$ is possible. This reduces to an unexpectedly simple result, which we will prove here.

Theorem 5.3. *If $p_k > 0$, the following holds:*

$$M_{\pm\infty} \equiv \lim_{t\to\pm\infty} M_t = \pm \max_k \{\pm x_k\}$$

Proof. We first note that:

$$p_k x_k^t \leq \sum_{k=1}^{n} p_k x_k^t \leq \sum_{k=1}^{n} p_k \left(\max_k \{x_k\}\right)^t = \left(\max_k \{x_k\}\right)^t$$

We can take the t^{th} root on all sides of the inequality. Now since limits preserve non-strict inequalities, we can take $t \to \infty$:

$$\lim_{t\to\infty} p_k^{\frac{1}{t}} x_k \leq \liminf_{t\to\infty} M_t \leq \limsup_{t\to\infty} M_t \leq \max_k \{x_k\}$$

Where we obtain a lim inf and a lim sup since it is uncertain whether $\lim_{t\to\infty} M_t$ exists. Now, since $p_k > 0$, we know that $\lim_{t\to\infty} p^{\frac{1}{t}} = 1$. We thus obtain:

$$x_k \leq \liminf_{t\to\infty} M_t \leq \limsup_{t\to\infty} M_t \leq \max_k \{x_k\}$$

Now the first inequality holds for any x_k, thus it also holds for $\max_k x_k$:

$$\max_k \{x_k\} \leq \liminf_{t\to\infty} M_t \leq \limsup_{t\to\infty} M_t \leq \max_k \{x_k\}$$

But then $\lim_{t\to\infty} M_t = \max_k x_k$.
Furthermore, using the following expression (which was used very often in the previous proof): $M_{-t}(x_1, \dots, x_n) = M_t^{-1}(\frac{1}{x_1}, \dots, \frac{1}{x_n})$, we find that $\lim_{t\to-\infty} M_t = \min_k \{x_k\}$. □

5.2. LOOSENING THE CONDITIONS

One might, of course, consider whether the condition of $\sum_{k=1}^{n} p_k = 1$ is truly necessary. Let us attempt to prove the necessity of this requirement for the power mean inequality, by looking at the so-called p-norms (which we will study extensively in Chapter 7), so by taking $p_k = 1\,\forall k$. Before doing so, however, we first prove a useful lemma.

Lemma 5.4. *Given the following function $f : \mathbb{R}^+ \to \mathbb{R}$ defined as $f(x) = x\sqrt{x}$. Then:*

$$f(x+y) \leq f(x) + f(y)$$

Proof.
$$f(x+y) = (x+y)\sqrt{x+y} = x\sqrt{x+y} + y\sqrt{x+y} \leq x\sqrt{x} + y\sqrt{y}$$
where the inequality follows from the fact that $\sqrt{\cdot}$ is an increasing function and $x, y \geq 0$. □

Theorem 5.5. *Let $(a_k)_{k=1}^n$ be a sequence of non-negative numbers. Then:*
$$\left(\sum_{k=1}^n a_k^{\frac{1}{2}}\right)^2 \leq \left(\sum_{k=1}^n a_k^{\frac{1}{3}}\right)^3$$

Proof. We will prove the following equivalent statement:
$$\left(\sum_{k=1}^n a_k^{\frac{1}{3}}\right)^{\frac{3}{2}} \geq \sum_{k=1}^n a_k^{\frac{1}{2}}$$
Note that:
$$\left(\sum_{k=1}^n a_k^{\frac{1}{3}}\right)^{\frac{3}{2}} \geq \sum_{k=1}^n \left(a_k^{\frac{1}{3}}\right)^{\frac{3}{2}} = \sum_{k=1}^n a_k^{\frac{1}{2}}$$
where the first equality follows from Lemma 5.4. □

We thus see that dropping the requirement flips the power mean inequality.

5.3. AN APPLICATION OF POWER MEANS

Lastly, we now turn to an interesting theorem that connects the mean value of sequences to their limit sequence and can be proved based on the theory of generalised means.

Theorem 5.6. ***Niven-Zuckerman Lemma.*** *Given a sequence of non-negative n-dimensional vectors $(\mathbf{a})_{k=1}^n$, such that $\mathbf{a}_k = \begin{bmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{bmatrix}$. Now let $\mu \in \mathbb{R}_{\geq 0}$ be a constant such that:*
$$\lim_{k\to\infty} \sum_{i=1}^n a_{ki} = n\mu$$
Furthermore, let $p \in (1, \infty)$ such that:

Then $\forall i \in \{1, \ldots, n\}$:
$$\lim_{k\to\infty} a_{ki} = \mu$$

Proof. Note that the two statements can be rewritten as:
$$\lim_{k\to\infty} \sum_{i=1}^n \frac{1}{n} \cdot a_{ki} = \mu$$
and
$$\lim_{k\to\infty} \sum_{i=1}^n \frac{1}{n} \cdot a_{ki}^p = \mu^p$$
Setting $p_k = \frac{1}{n}$, this comes down to:
$$\lim_{k\to\infty} M_1 = \mu$$
$$\lim_{k\to\infty} M_p^p = \mu^p \implies \lim_{k\to\infty} M_p = \mu$$

But then $\lim_{k\to\infty} M_1 = \lim_{k\to\infty} M_p$. In Theorem 5.2, we have proved that equality between means implies equality of the variables. We thus claim that:

$$\lim_{k\to\infty} a_{k1} = \lim_{k\to\infty} a_{k2} = \ldots = \lim_{k\to\infty} a_{kn}$$

But then $\lim_{k\to\infty} a_{ki} = \mu \, \forall \, i \in \{1,\ldots,n\}$.

Of course, one has to take great care when extending results to limit cases. Thus, to ensure that the step towards equality of variables is valid, we consider what happens when the limits of the variables are not equal. Without loss of generality, we restrict ourselves to the case of two unequal limits.
Let us assume that $\lim_{k\to\infty} a_{k1} = \mu + \epsilon$ for some $\epsilon \in \left(0,\mu\right)$[1] and that $\lim_{k\to\infty} a_{ki} = \mu \forall i \in \{2,\ldots,n-1\}$. Now in order to satisfy the first condition, we must surely have:

$$\lim_{k\to\infty} a_{kn} = n\mu - \sum_{i=1}^{n-1} \lim_{k\to\infty} a_{ki} = \mu - \epsilon$$

In order to satisfy the second condition, we must have that:

$$\left(\mu+\epsilon\right)^p + \left(\mu-\epsilon\right)^p = 2\mu^p$$

Rewriting this statement, however, we obtain:

$$\left(\left(\mu+\epsilon\right)^p + \left(\mu-\epsilon\right)^p\right)^{\frac{1}{p}} = 2\mu = \left(\mu+\epsilon\right)^1 + \left(\mu-\epsilon\right)^1$$

But the left hand side of this equation is the generalised power mean for power p, while the right hand side is the generalised mean for $p = 1$. Thus, we know that we only have equality if $\mu + \epsilon = \mu - \epsilon$. $\Rightarrow\Leftarrow$. $\square$

[1] Note that taking $\epsilon = \mu$ makes it impossible to satisfy both conditions, and thus, this case can be excluded.

6

Hölder's and Minkowski's Inequalities

The theory derived in the previous chapters can be used to prove another well-known inequality: *Hölder's inequality*. From Hölder's inequality, we can derive yet another famous inequality: Minkowsi's inequality. In this chapter, these inequalities, their generalisations and some of their applications will be studied. Furthermore, we will show how these inequalities relate to all the inequalities discussed up until now.

6.1. Hölder's Inequality

Before considering Hölder's Inequality, we first prove a lemma, which comes down to the direct application of the AM-GM inequality. The inequality introduced through this lemma carries its own name: Young's inequality.

Lemma 6.1. ***Young's Inequality*** *Let α, β, x and y be positive. Then:*

$$x^\alpha y^\beta \leq \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} y^{\alpha+\beta}$$

Proof. First, we denote $p_1 \equiv \frac{\alpha}{\alpha+\beta}$ and $p_2 \equiv \frac{\beta}{\alpha+\beta}$ and note that $p_1 + p_2 = 1$. Now, if we consider two positive numbers $\hat{x}$ and $\hat{y}$, we can apply the AM-GM inequality (Theorem 3.2) in order to find:

$$\hat{x}^{p_1} \hat{y}^{p_2} \leq p_1 \hat{x} + p_2 \hat{y}$$

Setting $\hat{x} \equiv x^{\alpha/p_1}$ and $\hat{y} \equiv y^{\beta/p_2}$ yields the expected result. □

Theorem 6.2. ***Hölder's Inequality*** *Let $p, q > 1$ be numbers such that $\frac{1}{p} + \frac{1}{q} = 1$, and let $(a_k)_{k=1}^n, (b_k)_{k=1}^n$ be sequences of non-negative real numbers. Then:*

$$\sum_{k=1}^n a_k b_k \leq \left(\sum_{k=1}^n a_k^p \right)^{\frac{1}{p}} \left(\sum_{k=1}^n b_k^q \right)^{\frac{1}{q}}$$

Proof. Let us consider one index, say $k = i$, first. Now let us consider the result of Lemma 6.1:

$$x^\alpha y^\beta \leq \frac{\alpha}{\alpha+\beta} x^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} y^{\alpha+\beta}$$

Setting $x = \sqrt[\alpha]{a_i}$ and $y = \sqrt[\beta]{b_i}$, we find:

$$a_i b_i \leq \frac{\alpha}{\alpha+\beta} a_i^{(\alpha+\beta)/\alpha} + \frac{\beta}{\alpha+\beta} b_i^{(\alpha+\beta)/\beta}$$

Setting $\alpha = \frac{1}{p}$ and $\beta = \frac{1}{q}$, we find:

$$a_i b_i \leq \frac{1}{p} a_i^p + \frac{1}{q} b_i^q$$

We now continue in a similar manner as was done in the proof of the Cauchy-Schwarz inequality. First, note that we thus find:

$$\sum_{k=1}^n a_k b_k \leq \frac{1}{p} \sum_{k=1}^n a_k^p + \frac{1}{q} \sum_{k=1}^n b_k^q$$

We now apply the *normalisation trick*: note that this equation also holds for the sequences $\hat{a}_k \equiv \frac{a_k}{\sqrt[p]{\sum_{i=1}^n a_i^p}}$ and $\hat{b}_k \equiv \frac{b_k}{\sqrt[q]{\sum_{i=1}^n b_i^q}}$. Substituting these into the right-hand side of the inequality yields:

$$\sum_{k=1}^n \hat{a}_k \hat{b}_k \leq \frac{1}{p} + \frac{1}{q} = 1$$

Thus:

$$\sum_{k=1}^n \frac{a_k}{\sqrt[p]{\sum_{i=1}^n a_i^p}} \frac{b_k}{\left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}} = 1 \iff \sum_{k=1}^n a_k b_k = \left(\sum_{i=1}^n a_i^p\right) \frac{1}{p} \left(\sum_{i=1}^n b_i^q\right)^{\frac{1}{q}}$$

□

Now note how Hölder's inequality simply follows from the AM-GM inequality. But since the Cauchy-Schwarz inequality is just Hölder's inequality for $p = \frac{1}{2}$, we can also easily derive the AM-GM inequality from Hölder's inequality!

Let us now derive the conditions for equality. From Lemma 6.1, we see that we only have equality when $(\hat{a}_i)^p = \left(\hat{b}_i\right)^q$ for all i. But this is satisfied only if $a_i = \lambda b_i^{\frac{w}{p}}$.

6.2. GENERALISATIONS, HISTORICAL FORMS AND A CONVERSE STATEMENT

In this section, we consider generalising Hölder's Inequality. After that, we study various historical forms of the inequality and show how they are equivalent to the form presented above. Lastly, we show how to find a bound on one of the terms on the right hand side.

Let us commence by generalising the inequality to N sequences. Before doing so, however, we introduce an important lemma, which can also be considered to be a generalisation of Hölder's inequality.

Lemma 6.3. *Let p, q, r be positive numbers such that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$ and let $(a_k)_{k=1}^n, (b_k)_{k=1}^n$ be sequences of non-negative real numbers. Then:*

$$\left(\sum_{k=1}^n |a_k b_k|^r\right)^{\frac{1}{r}} \leq \left(\sum_{k=1}^n a_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^n b_k^q\right)^{\frac{1}{q}}$$

Proof. Consider $t = \frac{p}{r}, s = \frac{q}{r}$. Then $\frac{1}{t} + \frac{1}{s} = 1$. We can now use 6.2 on the sequences a_k^r, b_k^r to find:

$$\sum_{k=1}^{n} |a_k b_k|^r \leq \left(\sum_{k=1}^{n} |a_k^r|^t\right)^{\frac{1}{t}} \left(\sum_{k=1}^{n} |b_k^r|^s\right)^{\frac{1}{s}} = \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{r}{p}} \left(\sum_{k=1}^{n} b_k^q\right)^{\frac{r}{q}} \tag{6.1}$$

Taking the rth root on both sides finishes the proof. □

Using this lemma, we first consider the case of $N = 3$ sequences.

Theorem 6.4. *Let p, q, r be positive numbers such that $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1$ and let $(a_k)_{k=1}^n, (b_k)_{k=1}^n, (c_k)_{k=1}^n$ be sequences of non-negative real numbers. Then:*

$$\sum_{k=1}^{n} a_k b_k c_k \leq \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} b_k^q\right)^{\frac{1}{q}} \left(\sum_{k=1}^{n} c_k^r\right)^{\frac{1}{r}}$$

Proof. Let us consider the sequence $d_k \equiv b_k c_k \forall k$. Then, given t such that $\frac{1}{p} + \frac{1}{t} = 1$, we find:

$$\sum_{k=1}^{n} a_k d_k \leq \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} d_k^t\right)^{\frac{1}{t}} = \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} (b_k c_k)^t\right)^{\frac{1}{t}}$$

Now since $\frac{1}{q} + \frac{1}{r} = \frac{1}{t}$, we can use Lemma 6.3 to find the desired result. □

We can now proceed to generalise this statement to an arbitrary number of sequences.

Theorem 6.5. *Let p_m for $m = 1, \ldots, N$ be a sequence of non-negative numbers such that $\sum_{m=1}^{N} \frac{1}{p_m} = 1$. Furthermore, let $\left(x_k^m\right)_{k=1}^n$ for $m = 1, \ldots, N$ be sequences of non-negative real numbers. Then:*

$$\sum_{k=1}^{n} \prod_{m=1}^{N} x_k^m \leq \prod_{m=1}^{N} \left(\sum_{k=1}^{n} |x_k^m|^{p_m}\right)^{\frac{1}{p_m}}$$

Proof. without loss of generality we assume that $N > 3$. Suppose N is even. We then obtain the desired result in a recursive manner as follows. Define q, r in such a manner that $\frac{1}{q} = \sum_{m=1}^{N/2} \frac{1}{p_m}$ and $\frac{1}{r} = \sum_{m=N/2+1}^{N} \frac{1}{p_m}$. Then, using Theorem 6.2 we find:

$$\sum_{k=1}^{n} \prod_{m=1}^{N} x_k^m \leq \left(\sum_{k=1}^{n} |\prod_{m=1}^{N/2} x_k^m|^q\right)^{\frac{1}{q}} \left(\sum_{k=1}^{n} |\prod_{m=N/2+1}^{N} x_k^m|^r\right)^{\frac{1}{r}}$$

We now have two possibilities. Either $N/2$ is even (and the products on the right hand side have an even number of terms). We then perform the same sequence of operations until we reach an odd number of terms.

If we obtain one term, we are done. So let us assume that the odd number of terms is larger than or equal to three. We then proceed in a similar fashion as we did in the proof of Theorem 6.4: We use Lemma 6.3 in order to split the expression up into a part over one sequence and a part over an even number of sequences. We then consider the part over the even number of sequences, and continue as we did above.

If we start off with an odd number of sequences, we use Lemma 6.3 and then proceed in a similar manner as presented above.
Note that this algorithm will always yield a result in at most N steps. □

Furthermore, one can show that Hölders inequality still holds when we include weights in both the inner product and the norms!

Theorem 6.6. *Let p, q be positive numbers such that $\frac{1}{p}+\frac{1}{q}=1$ and let $(a_k)_{k=1}^n, (b_k)_{k=1}^n, (w_k)_{k=1}^n$ be sequences of non-negative real numbers. Then:*

$$\sum_{k=1}^n w_k a_k b_k \leq \sqrt[p]{\sum_{k=1}^n w_k a_k^p}\sqrt[q]{\sum_{k=1}^n w_k |b_k|^q}$$

Proof. Note that $w_k = w_k^{\frac{1}{p}+\frac{1}{q}} = w_k^{\frac{1}{p}} w_k^{\frac{1}{q}}$. Applying Theorem 6.2 on the sequences $w_k^{\frac{1}{p}} a_k$ and $w_k^{\frac{1}{q}} b_k$ yields the expected result. □

Much to anyone's surprise, the original inequality that Hölder proved was actually a specific case of this weighted inequality! His original inequality follows by taking $a_k = 1 \forall k$.

Theorem 6.7. *Let p, q be positive numbers such that $\frac{1}{p}+\frac{1}{q}=1$ and let $(a_k)_{k=1}^n, (w_k)_{k=1}^n$ be sequences of non-negative real numbers. Then:*

$$\sum_{k=1}^n w_k b_k \leq \sqrt[p]{\sum_{k=1}^n w_k}\sqrt[q]{\sum_{k=1}^n w_k |b_k|^q}$$

We will prove this version of the inequality later in this chapter by making use of one of the inequalities we have derived in a previous chapter. Note that this statement is equivalent to Theorem 6.2 for two sequences c_k, d_k: seting $w_k = c_k^p$ and $b_k = \frac{d_k}{\sqrt[q]{c_k^p}}$ yields the modern version of Hölder's Inequality. The other implication follows from the proof of Theorem 6.6.

Interestingly enough, we use a statement that looks partially like Hölder's inequality in order to say something about the remaining aspects of the equality - we present this result, which can be considered to be a "converse" statement.

Theorem 6.8. *Let $p \in (1, \infty)$ and let $(a_k)_{k=1}^n$ be a sequence of real numbers. Suppose $\exists C \in \mathbb{R}$ such that $\forall x_k, k = 1, \ldots, n$:*

$$\sum_{k=1}^n |a_k x_k| \leq C \sqrt[p]{\sum_{k=1}^n |x_k|^p}$$

Then for $q = \frac{p}{p-1}$:

$$\sqrt[q]{\sum_{k=1}^n |a_k|^q} \leq C$$

Proof. Consider[1] $x_k \equiv \operatorname{sgn}(a_k)|a_k|^{\frac{q}{p}}$. Then surely:

$$\sum_{k=1}^n |a_k|^{\frac{q}{p}+1} \leq C \sqrt[p]{\sum_{k=1}^n |a_k|^q}$$

[1] Note that in Steele (2004) one finds an incorrect expression. The author of this thesis has corrected this and has implemented the correct result.

Note that $q = \frac{p}{p-1}$, such that $\frac{q}{p} = \frac{1}{p-1}$. Thus the power of the left hand side yields $\frac{q}{p} + 1 = \frac{1}{p-1} + \frac{p-1}{p-1} = \frac{p}{p-1} = q$. Therefore, we find:

$$\sum_{k=1}^{n} |a_k|^q \leq C \sqrt[p]{\sum_{k=1}^{n} |a_k|^q}$$

Dividing both sides by the right hand side yields the desired result, since:

$$\sqrt[q]{\sum_{k=1}^{n} |a_k|^q} = \left(\sum_{k=1}^{n} |a_k|^q\right)^{1-\frac{1}{p}} = \frac{\sum_{k=1}^{n} |a_k|^q}{\sqrt[p]{\sum_{k=1}^{n} |a_k|^q}} \leq C$$

□

6.3. ONE LAST INEQUALITY: MINKOWSKI'S INEQUALITY

We now introduce another inequality, which is often considered to be a consequence of Hölder's inequality, but can also be considered the generalisation of the Triangle inequality on p-spaces, which we will discuss extensively in the next chapter. Without further ado, let us prove Minkowski's inequality.

Theorem 6.9. ***Minkowski's Inequality.*** *Let $p \geq 1$ and let $(a_k)_{k=1}^{n}, (b_k)_{k=1}^{n}$. Then:*

$$\left(\sum_{k=1}^{n} |a_k b_k|^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^{n} |a_k|^p\right)^{\frac{1}{p}} + \left(\sum_{k=1}^{n} |b_k|^p\right)^{\frac{1}{p}}$$

Proof. Note that:

$$\sum_{k=1}^{n} |a_k + b_k|^p = \sum_{k=1}^{n} |a_k + b_k||a_k + b_k|^{p-1} \leq \sum_{k=1}^{n} |a_k||a_k + b_k|^{p-1} + \sum_{k=1}^{n} |b_k||a_k + b_k|^{p-1}$$

Now we can use Theorem 6.2 on both sums on the right hand side. We then find:

$$\sum_{k=1}^{n} |a_k||a_k + b_k|^{p-1} \leq \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^{n} |a_k + b_k|^p\right)^{\frac{p-1}{p}}$$

An analogous result can be derived for the second sum. Combining these results yields:

$$\sum_{k=1}^{n} |a_k + b_k|^p \leq \left(\left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}} + \left(\sum_{k=1}^{n} b_k^p\right)^{\frac{1}{p}}\right) \left(\sum_{k=1}^{n} |a_k + b_k|^p\right)^{\frac{p-1}{p}}$$

We can divide out the last term to find the desired results:

$$\left(\sum_{k=1}^{n} |a_k + b_k|^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^{n} a_k^p\right)^{\frac{1}{p}} + \left(\sum_{k=1}^{n} b_k^p\right)^{\frac{1}{p}}$$

□

6.4. CONNECTING THE DOTS: HÖLDER IN RELATION TO THE PRIOR INEQUALITIES

We have already seen how Hölder's inequality follows from the AM-GM inequality. Furthermore, Hölder's inequality is obviously a generalisation of the inequality of Cauchy and Schwarz, and therefore, intuitively, Hölder's inequality appears to be a much stronger statement. In this chapter, we will see how Hölder's inequality can be derived from all the other inequalities - much to one's surprise!

6.4.1. DERIVING HÖLDER FROM CAUCHY-SCHWARZ

We will now show that Hölder's inequality can be derived from the Cauchy-Schwarz inequality (2.4). One will see that the Cauchy-Schwarz Inequality also possesses a self-generalising property, and as such, the proof will be analogous to that of the AM-GM inequality to quite some extend. Let us first consider the case of 2^N sequences, so $(x_k^1)_{k=1}^n, \dots, \left(x_k^{2^N-1}\right)_{k=1}^n, \left(x_k^{2^N}\right)_{k=1}^n$.We can then use the Cauchy-Schwarz inequality to derive the following result:

$$\left(\sum_{k=1}^n \prod_{m=1}^{2^N} x_k^m\right)^{2^N} \leq \prod_{m=1}^{2^N} \sum_{k=1}^n |x_k^m|^{2^N}$$

This can easily be seen by iterating as follows:

$$\left(\sum_{k=1}^n \prod_{m=1}^{2^N} x_k^m\right)^{2^N} \overset{C-S}{\leq} \left(\sum_{k=1}^n \prod_{m=1}^{2^{N-1}} |x_k^m|^2\right)^{2^{N-1}} \left(\sum_{k=2^{N-1}+1}^n \prod_{m=1}^{2^N} |x_k^m|^2\right)^{2^{N-1}} \overset{C-S}{\leq} \dots \overset{C-S}{\leq} \prod_{m=1}^{2^N} \sum_{k=1}^n |x_k^m|^{2^N}$$

This results also follows immediately from the generalisation of the Cauchy-Schwarz inequality (Theorem 2.7). Now from this, it easily follows that Holder's inequality holds for two non-negative real sequences $(a_k)_{k=1}^n, (b_k)_{k=1}^n$ and for any number $p \in \left\{\frac{2^N}{j} : j = 1, \dots, 2^N - 1\right\}$ and its reciprocal by choosing X_k^m such that $a_k = \prod_{m=1}^j x_k^m$ and x_k^m equal for all $m = 1, \dots, j$ and choosing the other x_k^m equal such that $b_k = \prod_{m=j+1}^{2^N} x_k^m$, as we then obtain:

$$\begin{aligned}\left(\sum_{k=1}^n a_k b_k\right)^{2^N} = \left(\sum_{k=1}^n \left(\prod_{m=1}^j x_k^m\right)\left(\prod_{m=j+1}^{2^N} x_k^m\right)\right)^{2^N} &\leq \left(\prod_{m=1}^j \sum_{k=1}^n |x_k^m|^{2^N}\right)\left(\prod_{m=j+1}^{2^N} \sum_{k=1}^n |x_k^m|^{2^N}\right) \\ &= \left(\sum_{k=1}^n |a_k|^{\frac{2^N}{j}}\right)^j \left(\sum_{k=1}^n |x_k^m|^{\frac{2^N}{2^N-j}}\right)^{2^N-j}\end{aligned}$$

Taking the 2^Nth root of both sides yields:

$$\sum_{k=1}^n a_k b_k \leq \left(\sum_{k=1}^n |a_k|^{\frac{2^N}{j}}\right)^{\frac{j}{2^N}} \left(\sum_{k=1}^n |x_k^m|^{\frac{2^N}{2^N-j}}\right)^{\frac{2^N-j}{2^N}} = \left(\sum_{k=1}^n a_k^p\right)^{\frac{1}{p}} \left(\sum_{k=1}^n b_k^q\right)^{\frac{1}{q}}$$

Note that even though p can now only attain rational values, for any $p \in \mathbb{R}_{\geq 1} \setminus \mathbb{Q}$ we can define a sequence $p_m \in \mathbb{Q}$ such that $p_m \to p$. Setting $c_m = \left(\sum_{k=1}^n a_k^{p_m}\right)^{\frac{1}{p_m}} \left(\sum_{k=1}^n b_k^{q_m}\right)^{\frac{1}{q_m}}$, we find that:

$$\sum_{k=1}^n a_k b_k \leq c_m \quad \forall n \in \mathbb{N}$$

Thus, taking the limit $m \to \infty$ yields Hölder's inequality for any real p.

STARTING WITH JENSEN'S INEQUALITY

We can show that both Minkowski's as well as Hölder's inequality can be derived from Jensen's Inequality.
First, note that Jensen's Inequality can be rephrased as follows:

Theorem 6.10. *Let w_i be positive, real numbers and let ϕ be concave. Then:*

$$\frac{\sum_{i=1}^n w_i \phi(x_i)}{\sum_{k=1}^n w_i} \leq \phi\left(\frac{\sum_{i=1}^n w_i x_i}{\sum_{k=1}^n w_i}\right)$$

Let us consider the concave function $\phi(x) = \left(1 + x^{\frac{1}{p}}\right)^p$ for $x \in [0,\infty)$. We then obtain:

$$\frac{\sum_{i=1}^n w_i \left(1 + x_i^{\frac{1}{p}}\right)^p}{\sum_{k=1}^n w_k} \leq \left(1 + \left(\frac{\sum_{i=1}^n w_i x_i}{\sum_{k=1}^n w_i}\right)^{\frac{1}{p}}\right)^p$$

Which is equivalent to:

$$\left(\sum_{i=1}^n w_i \left(1 + x_i^{\frac{1}{p}}\right)^p\right)^{\frac{1}{p}} \leq \left(\sum_{k=1}^n w_i\right)^{\frac{1}{p}} + \left(\sum_{i=1}^n w_i x_i\right)^{\frac{1}{p}}$$

Substituting $w_i = a_i^p$ and $x_i = \left(\frac{b_i}{a_i}\right)^p$, we retrieve Minkowski's inequality.

Let us now derive the historical Hölder's Inequality in the same manner as Hölder did. We consider the *convex* function $\phi(x) = x^q$. We apply Jensen's inequality to find:

$$\left(\frac{\sum_{k=1}^n w_k b_k}{\sum_{i=1}^n w_i}\right)^q = \phi\left(\frac{\sum_{k=1}^n w_k b_k}{\sum_{i=1}^n w_i}\right) \leq \frac{\sum_{k=1}^n w_k \phi(b_k)}{\sum_{i=1}^n w_i} = \frac{\sum_{k=1}^n w_k b_k^q}{\sum_{i=1}^n w_i}$$

Which is equivalent to:

$$\sum_{k=1}^n w_k b_k \leq \frac{\sum_{i=1}^n w_i}{\left(\sum_{i=1}^n w_i\right)^{\frac{1}{q}}} \left(\sum_{k=1}^n w_k b_k^q\right)^{\frac{1}{q}} = \frac{\left(\sum_{i=1}^n w_i\right)^{\frac{1}{p}+\frac{1}{q}}}{\left(\sum_{i=1}^n w_i\right)^{\frac{1}{q}}} \left(\sum_{k=1}^n w_k b_k^q\right)^{\frac{1}{q}} = \left(\sum_{i=1}^n w_i\right)^{\frac{1}{p}} \left(\sum_{k=1}^n w_k b_k^q\right)^{\frac{1}{q}}$$

This is the historical inequality, from which the inequality in the form that we presented initially follows.

DERIVING HÖLDER'S INEQUALITY FROM MINKOWSKI'S INEQUALITY

Lastly, while we proved Minkowski's inequality from Hölder's inequality, we can also prove Hölder's Inequality from Minkowski's inequality by taking a small detour via the results that we obtained surrounding power means! We present these results in the form of a lemma.

Lemma 6.11. *Given $\theta \in [0,1]$ and $a, b \geq 0$. Then:*

$$\lim_{p\to\infty} \left(\theta a^{1/p} + (1-\theta) b^{1/p}\right)^p = a^\theta b^{1-\theta}$$

We now apply Minkowski's inequality to the sequences $\left(\theta a_k^{p/s}\right)_{k=1}^n, \left((1-\theta) b_k^{q/s}\right)_{k=1}^n$ using an s-norm:

$$\left(\sum_{k=1}^n \left(\theta a_k^{p/s} + (1-\theta) b_k^{q/s}\right)^s\right)^{1/s} \leq \left(\sum_{k=1}^n \left(\theta a_k^{p/s}\right)^s\right)^{1/s} + \left(\sum_{k=1}^n \left(1-\theta) b_k^{q/s}\right)^s\right)^{1/s} = \theta \left(\sum_{k=1}^n a_k^p\right)^{1/s} + (1-\theta) \left(\sum_{k=1}^n b_k^q\right)^{1/s}$$

Raising both sides to the power of s, we find:

$$\sum_{k=1}^{n}\left(\theta a_k^{p/s}+(1-\theta)b_k^{q/s}\right)^s \leq \left(\theta\left(\sum_{k=1}^{n} a_k^p\right)^{1/s}+(1-\theta)\left(\sum_{k=1}^{n} b_k^q\right)^{1/s}\right)^s$$

We can now consider the limit $s \to \infty$ and use Lemma 6.11 to find:

$$\sum_{k=1}^{n} a_k^{p\theta} b_k^{(1-\theta)q} \leq \left(\sum_{k=1}^{n} a_k^p\right)^{\theta}\left(\sum_{k=1}^{n} b_k^q\right)^{1-\theta}$$

Setting $\theta = \frac{1}{p}$ yields Hölder's inequality.

Note that since we can prove Hölder's inequality from Minkowski's inequality, we can also easily prove that the Cauchy-Schwarz inequality follows from Minkowski's inequality - from which we can show very easily that the AM-GM inequality follows from Minkowski's inequality as well: the starting point of our study of Hölder's inequality!

7

EXPLORING P-SPACES

All the results that we have derived up until now are restricted to finite sequences. In this section, we study a very important class of normed vector spaces: $\mathbb{R}^n$ equipped with the so-called p-norm for $p \geq 1$ and the standard inner product. Note that these spaces are complete under each norm and the metric induces by them. We will study some interesting properties of these spaces, and will lastly show how this structure is lost when we extend to infinite dimensional vector spaces. Note that we will be referring to many of the definitions introduced in Chapter 1.

7.1. INTRODUCING P-SPACES

We first recall the definition of the p-norm.

Definition 7.1. Let $\mathbf{x} \in \mathbb{R}^n$ and let $p \geq 1$. We then define the *p-norm* as:

$$\|\mathbf{x}\|_p = \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p}$$

Furthermore, for $p = \infty$, we set:

$$\|\mathbf{x}\|_\infty = \max_i \{|x_i|\}$$

We denote the normed space $\left(\mathbb{R}^n, \|\cdot\|_p\right)$ as $\mathbb{R}^n_p$.

One may wonder why the p-norm is only defined for $p \geq 1$. Let us consider what happens to Minkowski's inequality (Theorem 6.9) if $p < 1$. As was mentioned briefly in Chapter 6, Minkowski's inequality can be seen as a generalisation of the triangle inequality Note that the triangle inequality follows from Minkowski's inequality by setting $\mathbf{a} \equiv \mathbf{x} - \mathbf{z}$ and $\mathbf{b} \equiv \mathbf{z} - \mathbf{y}$:

$$d_p(\mathrm{x}, \mathrm{y}) = \|\mathbf{x} - \mathbf{y}\|_p = \|\mathbf{a} + \mathbf{b}\|_p \leq \|\mathbf{a}\| + \|\mathbf{b}\|_p = \|\mathbf{x} - \mathbf{z}\|_p + \|\mathbf{y} - \mathbf{z}\|_p = d_p(\mathrm{x}, \mathrm{z}) + d_p(\mathrm{z}, \mathrm{y})$$

Let us refer back to the proof of Minkowksi's inequality on the basis of Jensen's inequality, in which we made use of the *concave* mapping $\phi(x) = \left(1 + x^{\frac{1}{p}}\right)^p$, from which we obtained

$$\frac{\sum_{i=1}^n w_i \left(1 + x_i^{\frac{1}{p}}\right)^p}{\sum_{k=1}^n w_k} \leq \left(1 + \left(\frac{\sum_{i=1}^n w_i x_i}{\sum_{k=1}^n w_i}\right)^{\frac{1}{p}}\right)^p$$

using Jensen's inequality (Theorem 4.1). Note, however that this final step is only valid if the mapping $x \to (1+x^{\frac{1}{p}})^p$ is *convex*. This is the case for $p \geq 1$ - for $p < 1$, however, this mapping becomes *concave*, and we obtain the inverse of the triangle inequality, i.e. we find that:

$$\|\mathbf{a}+\mathbf{b}\|_{p<1} \geq \|\mathbf{a}\|_{p<1} + \|\mathbf{b}\|_{p<1}$$

which yields:

$$d_p(\mathrm{x},\mathrm{y}) \geq d_p(\mathrm{x},\mathrm{z}) + d_p(\mathrm{z},\mathrm{y})$$

This, of course, conflicts with properties of metrics, and as such, the requirement that $p \geq 1$ is a necessary requirement for us to have a valid metric space.

7.2. PROPERTIES OF P-SPACES

In this section, we will prove inequalities on the class of p-norms. These inequalities will reveal a cyclic relationship between two different p-norms. We will then show how this cyclic relationship ensures that a lot of structure on the different $\mathbb{R}_p^n$ is essentially equal.

Theorem 7.1. *Let* $1 < p_2 < p_1 < \infty$. *Then* $\forall \mathbf{x} \in \mathbb{R}^n$*:*

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_{p_1} \leq \|\mathbf{x}\|_{p_2} \leq \|\mathbf{x}\|_1 \leq n\|\mathbf{x}\|_\infty$$

Proof. We start off by proving the leftmost inequality.

$$\|\mathbf{x}\| = \max_i(|x_i|) = \left(|x_i|^{p_1}\right)^{1/p_1} \leq \left(\sum_{i=1}^n |x_i|^{p_1}\right)^{1/p_1} = \|\mathbf{x}\|_{p_1}$$

Now consider the last two inequalities. We start with the first of these.

$$\|\mathbf{x}\|_{p_2} = \left(\sum_{i=1}^n |x_i|^{p_2}\right)^{1/p_2} \leq \sum_{i=1}^n |x_i| = \|\mathbf{x}\|_1 \iff \sum_{i=1}^n |x_i|^{p_2} \leq \left(\sum_{i=1}^n |x_i|\right)^{p_2} \iff \left(\sum_{i=1}^n \frac{|x_i|}{\sum_{i=1}^n |x_i|}\right)^{p_2} \leq 1$$

which, evidently, is true.

$$\|\mathbf{x}\|_1 = \sum_{k=1}^n |x_i| \leq \sum_{k=1}^n \max_i(|x_i|) = n \cdot \max_i(|x_i|) = n\|\mathbf{x}\|_\infty$$

□

We now come to the last of these inequalities.

$$\|\mathbf{x}\|_{p_1} = \left(\sum_{i=1}^n |x_i|^{p_1}\right)^{1/p_1} \leq \left(\sum_{i=1}^n |x_i|^{p_2}\right)^{1/p_2} = \|\mathbf{x}\|_{p_2} \iff \left(\sum_{i=1}^n |x_i|^{p_1}\right)^{p_2/p_1} \leq \left(\sum_{i=1}^n |x_i|^{p_2}\right)$$

But this comes down to proving that:

$$\|\mathbf{x}^{p_2}\|_{p_1/p_2} \leq \|\mathbf{x}^{p_2}\|_1$$

which, for $t \equiv p_1/p_2 > 1$, reduces to proving that $\|\mathbf{y}\|_t \leq \|\mathbf{y}\|_1$ (where $y_i = |x_i|^{p_2}$). But this has been done already, and as such, we are done.

Corollary 7.1.1. *Let n be fixed. Then all $\mathbb{R}_p^n$ share the same convergent sequences, the same open sets and the same compact sets.*

Proof. Consider a sequence of vectors $(\mathbf{x}_k)_{k=1}^{\infty}$ such that $\mathbf{x}_k \xrightarrow{p_1} \mathbf{x}$ for some $p_1 \in [1,\infty]$. We then have:

$$\forall \epsilon > 0 : \exists N \in \mathbb{N} : \forall k \geq N : \|\mathbf{x} - \mathbf{x}_k\|_{p_1} < \frac{\epsilon}{n}$$

Let us now consider $p_2 \in [1,\infty)$. Two cases can be distinguished.
Case I: $p_1 < p_2$.
We now simply have that for given ϵ and N:

$$\|\mathbf{x} - \mathbf{x}_k\|_{p_2} \leq \|\mathbf{x} - \mathbf{x}_k\|_{p_1} < \frac{\epsilon}{n} \quad \forall k \geq N$$

because of Theorem 7.1.
Case II: $p_1 > p_2$.
In this case, we find that $\|\mathbf{x}\|_{p_2} \leq n\|\mathbf{x}\|_{\infty} \leq n\|\mathbf{x}\|_{p_1}$. We thus find that $\forall \epsilon > 0 : \exists N \in \mathbb{N} : \forall k \geq N$:

$$\|\mathbf{x} - \mathbf{x}_k\|_{p_2} \leq n \cdot \|\mathbf{x} - \mathbf{x}_k\|_{p_1} \leq n \cdot \frac{\epsilon}{n} = \epsilon$$

So in either case, we have that $\mathbf{x}_k \xrightarrow{p_2} \mathbf{x}$. We thus see that all p-spaces share the same convergent sequences.

We know that open sets and convergent sets are related in the following way:

$$X \subseteq \mathbb{R}_p^n \text{ is open } \iff \forall\, (\mathbf{x}_k)_{k=1}^{\infty} \in \mathbb{R}_p^n \setminus X \text{ convergent: } \mathbf{x}_k \to \mathbf{x}, \text{ for } \mathbf{x} \in \mathbb{R}_p^n \setminus X$$

But since all p-metrics yield the same convergent sequences, they also share the same open sets.

But then, all p-spaces share the same complete sets, since convergent Cauchy sequences in one p-space must also converge under all other p metrics.
But then, since these p-spaces share the same open sets, they also share the same totally bounded sets. But then, they also share the same compact sets. □

This proof shows that results on any R_p^n space regarding for example convergence of sequences or openness of sets can be extended to all other p-metrics. But this is not all! We can even prove that all such results hold for *any* arbitrary norm on $\mathbb{R}^n$, and as such are completely completely independent under the choice of norm!

Theorem 7.2. *Let n be fixed and let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. Then for all $\|\cdot\|_p$, we have that $\|\cdot\|$ and $\|\cdot\|_p$ generate the same convergent sequences, the same open sets and the same compact sets.*

Proof. We first prove that $\exists C$ such that

$$\|\mathbf{x}\| \leq C\|\mathbf{x}\|_{\infty}$$

Let us first write $\mathbf{x}$ in terms of the basis vectors:

$$\|\mathbf{x}\| = \|\sum_{i=1}^{n} x_i \mathbf{e}_i\| = \sum_{i=1}^{n} |x_i| \|\mathbf{e_i}\| \leq C\left(\sum_{i=1}^{n} |x_i|\right) = C\|\mathbf{x}\|_1$$

with $C \equiv \max_i (\|\mathbf{e}_i\|)$. Using Theorem 7.1, we find: $\|\mathbf{x}\| \leq Cn\|x\|_{\infty}$.
We now prove that $\|\mathbf{x}\| \geq m\|\mathbf{x}\|_{\infty}$. Let us consider the following compact subset of $\mathbb{R}_{\infty}^n : K \equiv$

$\{\mathbf{x} \in \mathbb{R}^n | \|\mathbf{x}\|_\infty = 1\}$ and the function $f : \mathbf{x} \to \|\mathbf{x}\|$. We then use the inverse triangle inequality to derive:

$$|f(x) - f(y)| = |\|\mathbf{x}\| - \|\mathbf{y}\|| \leq \|\mathbf{x} - \mathbf{y}\| \leq C \cdot n \|\mathbf{x} - \mathbf{y}\|_\infty$$

And thus, f is Lipschitz continuous on K. Note, however, that since K is a compact set, we can find a minimum and maximum of f on K, i.e. $\exists m, M \in \mathbb{R}_{>0} : m \leq f(k) \leq M$. Thus, given some $\mathbf{x} \in \mathbb{R}$:

$$\|\mathbf{x}\| = \| \|\mathbf{x}\|_\infty \frac{\mathbf{x}}{\|\mathbf{x}\|_\infty} \| = \|\mathbf{x}\|_\infty \| \frac{\mathbf{x}}{\|\mathbf{x}\|_\infty} \| \geq m \|\mathbf{x}\|_\infty$$

which follows from the fact that $\frac{\mathbf{x}}{\|\mathbf{x}\|_\infty} \in K$. As such, we find:

$$m \|\mathbf{x}\|_\infty \leq \|\mathbf{x}\| \leq n \cdot C \|\mathbf{x}\|_\infty$$

By continuing in the exact same manner as was done in the proof of the previous theorem, we can thus show that we indeed obtain the same convergent sequences as well as the same open, complete and compact sets under both norms. □

This theorem thus shows that studying the topology of the n dimensional real vector space, which can be restricted to just studying to $\mathbb{R}^n$, can simply be done by choosing one's favourite metric! Note that convergence (or, equivalently, openness) is a very strong property, as other, more complex structures, like for example continuity of functions over the vector field at hand, can completely by characterised by convergent sequences (or, equivalently, open sets)!

7.3. EXTENSION TO $\mathbb{R}^\infty$

At last, we have arrived at the point at which we can reflect on the finitude of the spaces that we are considering - specifically, we are interested in seeing whether the strong structure preserving properties of the spaces at hand are still present on infinite dimensional spaces. In order to assess this, we limit ourselves to $\mathbb{R}^\infty$, and consider whether or not the norm-independent structure present in the finite dimensional case is still to our disposal. Let us now introduce a linear subspace of $\mathbb{R}^\infty$, which we will study in this section, which we will dub V:

$$V = \{x = (x_n)_{n=1}^\infty \text{ with } x_n = 0 \, \forall n > n_0 \text{ for some} n_0.\}$$

Note that we can extend the aforementioned p-norms to this set: since for every sequence there is some n_0 such that $x_n = 0 \forall n \geq n_0$, so for that specific sequence, V acts as $\mathbb{R}^{n_0}$ under the p-norms. Therefore, we still have:

$$\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_{p_1} \leq \|\mathbf{x}\|_{p_2} \leq \|\mathbf{x}\|_1$$

for $1 < p_2 < p_1 < \infty$ and for every $\mathbf{x} \in V$. However, the topologies generated under each of these norms are distinct! Let us prove this in the following theorem.

Theorem 7.3. *Given the set V as defined above. Let* $1 < p_2 < p_1 < \infty$, *then:*

$$\tau_\infty \subset \tau_{p_1} \subset \tau_{p_2} \subset \tau_1$$

Proof. Let us first prove that:

$$\tau_\infty \subseteq \tau_{p_1} \subseteq \tau_{p_2} \subseteq \tau_1$$

We can restrict ourselves to the case for $q_1 \in (1, \infty)$ and $q_2 \in [1, q_1)$ and prove that

$$\tau_{q_1} \subseteq \tau_{q_2}$$

without loss of generalisation. Suppose $X \in \tau_{q_1}$. Then X is open under the q_1-norm. We prove that X is open under the q_2-norm. Take $\mathbf{x} \in X$. Then $\exists \epsilon$ such that $B_\epsilon^{q_1}(\mathbf{x}) \subseteq X$. Now take $\mathbf{y} \in B_\epsilon^{q_2}(\mathbf{x})$. Then since $\|\mathbf{x}-\mathbf{y}\|_{q_1} \leq \|\mathbf{x}-\mathbf{y}\|_{q_2} < \epsilon$, we have that $\mathbf{y} \in B_\epsilon^{q_1}(\mathbf{x})$. As such, we have that $B_\epsilon^{q_2}(\mathbf{x}) \subseteq B_\epsilon^{q_1}(\mathbf{x})$. But then surely $\forall \mathbf{x} \in X : \exists \epsilon : \mathbf{x} \in B_\epsilon^{q_2}(\mathbf{x})$. Thus, all open sets under the q_1-norm are open under the q_2-norm.

We now prove that τ_{q_1} is a *strict subset* of τ_{q_2}. We do this by showing that $B_\delta^{q_2}(\mathbf{x}) \notin \tau_{q_1}$. More specifically, we show that any q_1-open set cannot be fully contained in $B_1^{q_2}$ under the q_1-norm.
We first note that we can make use of the fact that $B_\delta^{q_2}(\mathbf{x}) = \mathbf{x} + B_\delta^{q_2}(\mathbf{0})$, such that we can limit ourselves to the case $B_\delta^{q_2}(\mathbf{0})$ without loss of generality.

We prove that $B_\epsilon^{q_1}(\mathbf{0}) \not\subseteq B_\delta^{q_2}(\mathbf{0})$ under the q_1-norm for all $\epsilon > 0$.
Let us consider the sequence $x_n = \frac{1}{n^{1/q_2}}$. Then:

$$\sum_{n=1}^{\infty} x_n^{q_2} \text{ diverges, while}$$
$$\sum_{n=1}^{\infty} x_n^{q_1} \text{ converges}$$

Now let $\epsilon > 0$. Then: $\exists N_0 : \left(\sum_{n=N_0}^{\infty} x_n^{q_1}\right)^{1/q_1} < \epsilon$. Furthermore, we also have that $\exists N_1 : \sum_{n=N_0}^{N_1} x_n^{q_2} > \delta$. But then, surely $\left(\sum_{n=N_0}^{N_1} x_n^{q_1}\right)^{1/q_1} < \epsilon$. Thus, if we define a new sequence $\mathbf{y} \equiv \left(x_{N_0}, x_{N_0+1} \ldots, x_{N_1}, 0, 0, \ldots\right)$, then:

$$\mathbf{y} \in B_\epsilon^{q_1}(\mathbf{0})$$
$$\mathbf{y} \notin B_\delta^{q_2}(\mathbf{0})$$

Lastly, we want to prove that $\tau_\infty \subset \tau_p$ for all $p \in [1, \infty)$. Suppose that for some $p : \tau_\infty = \tau_p$. Then let $q \in (p, \infty)$. Then $\tau_q = \tau_p$. $\Rightarrow\Leftarrow$. Thus $\tau_\infty \subset \tau_p$. $\square$

We see that, contrary to the finite dimensional case, the topologies on $\mathbb{R}^\infty$ are not equivalent.

CONCLUDING PART I

In the past chapters, we have given several proofs for the **Cauchy-Schwarz Inequality**, the **AM-GM inequality**, **Hölder's Inequality** and **Minkowski's inequality**, and discussed the intimate relationship between these inequalities by showing how each of them can be derived from the others. We have also proved how the Cauchy-Schwarz inequality and the AM-GM inequality follow from the simple and almost trivial fact that $(a-b)^2 \geq 0$.
We saw that this trivial fact hinted towards the study of convexity, which proved to be a surprisingly strong property of which **Jensen's Inequality** could be derived.
We also considered a generalised class of means, and proved a generalisation of the AM-GM inequality from that: the **Power Means Inequality**.
In retrospect, all the other inequalities could be considered to be (often) simple consequences of Jensen's inequality - which in and of itself was a simple consequence of the definition of convexity.
The connection between all the inequalities we studied is represented in the following diagram.

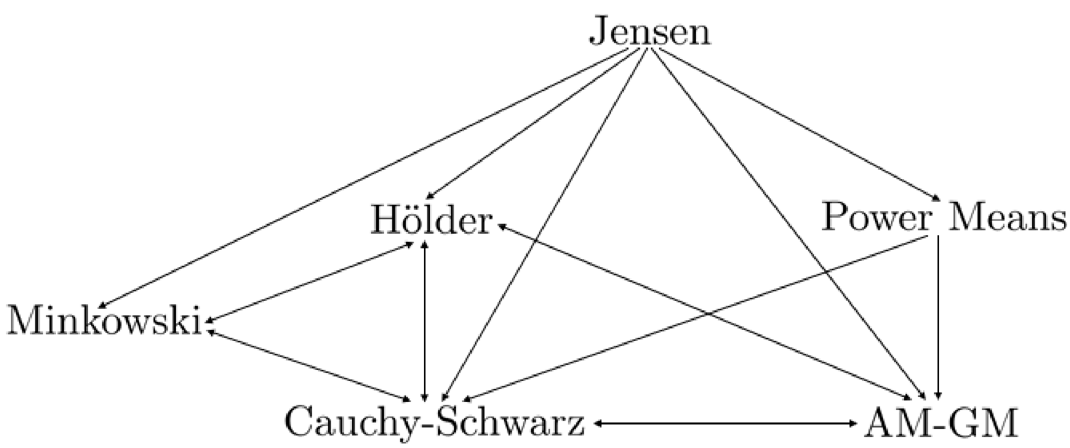


Figure 7.1: The connection between the main inequalities of this thesis is displayed in this diagram. The inequality at the end of an arrow can be derived using the inequality at the base of the arrow.

Lastly, we studied inequalities relating to the **p-norms** which followed from Jensen's inequality, and showed how the topologies under each of these norms (and under every other norm!) were the same on finite dimensional real vector spaces, and as such, how a lot of structure on real vector is preserved when switching between different norms. We ended by showing that this rich structure is lost when extending to the infinite dimensional case.

Part II

Introduction to Part II

After having studied mathematical inequalities extensively in Part I, we can finally turn to the physical part of this thesis, which concerns the detection of quantum entanglement, and how it relates to the mathematical inequalities that we have studied so far.

We will first introduce some fundamental quantum mechanical concepts in Chapter 8.

Then, in Chapter 9, we will study the concept of quantum entanglement and delve into various entanglement methods. This section will mostly be based on the results found in Gühne and Tóth (2008).

Then, finally, we will join our prior enquiries into mathematical inequalities from Part I with quantum entanglement detection, and we will introduce entanglement criteria derived in Wölk et al. (2014) in Chapter 10.

After having proved these criteria, we will assess them in Chapter 11 by considering which states they can and cannot detect, as well as by comparing them to the criteria introduced in Chapter 9.

8

QUANTUM MECHANICAL FUNDAMENTALS

This chapter marks the beginning of the second part of this thesis, in which the applications of the inequalities that have heretofore been studied to the field of quantum entanglement detection will be studied. First, however, fundamental quantum mechanical concepts and theorems will be introduced. These concepts will form the basis for the next few chapters, in which these concepts will be supplemented by more in-depth discussion of quantum entanglement and its detection. Note that this section will mostly be based on Nielsen and Chuang (2012).

8.1. THE AXIOMATIC FOUNDATION OF QUANTUM MECHANICS

Four postulates form the basis of the mathematical description of the physical description of quantum mechanical phenomena. These four postulates will be introduced below.

Postulate I: Any isolated physical system has an associated Hilbert space, called the *state space*, in which the system can be completely described by a unit vector, the so called *state vector*, henceforth denoted by $|\Psi\rangle$.

Postulate II: The evolution of a closed quantum system is described by a *unitary transformation.* The time evolution of the state of any closed quantum system is described by the **Schrödinger equation***:

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\Psi\rangle = \hat{H}|\Psi\rangle$$

where $\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + \hat{V}$ denotes the Hamiltonian operator.

Postulate III: Quantum measurements are described by a collection of measurement operators $\{M_m\}$, where the indices refer to the possible outcomes. For a state $|\Psi\rangle$, the probability of measuring outcome m is:

$$\mathbb{P}(M = m) = \langle\Psi| M_m^\dagger M_m |\Psi\rangle$$

After measurement, the state collapses to a new state $|\hat{\psi}\rangle$:

$$|\hat{\psi}\rangle = \frac{M_m|\psi\rangle}{\sqrt{\langle\psi| M_m^\dagger M_m |\psi\rangle}}$$

There is also a special class of measurements, the so-called *projective measurements*. These are described by an *observable* M, which is hermitian on the state space of the system and has a spectral decomposition, i.e.:

$$M = \sum_m m P_m$$

where m denotes a possible outcome of the measurement and P_m is the projector onto the corresponding eigenspace. The probability of measuring m is thus:

$$\mathbb{P}(M = m) = \langle \psi | P_m | \psi \rangle$$

This property leads to nice results for the expected value and the standard deviation:

$$\langle M \rangle \equiv \mathbb{E}(M) = \sum_m \mathbb{P}(M = m) = \langle \psi | M | \psi \rangle$$

$$\sigma(M) \equiv \sqrt{\langle (M - \langle M \rangle)^2 \rangle} = \sqrt{\langle \psi | M^2 | \psi \rangle - \left(\langle \psi | M | \psi \rangle\right)^2}$$

Postulate IV: The state phase of a composite physical system is described by the tensor product of its constituting component systems.

We now proceed to studying the concepts introduced in these postulates more thoroughly.

8.2. Extending the Axiomatic Framework

In this section, we extend upon the postulates of the field of quantum mechanics. First, we delve deeper into the description of states, and then we define quantum operators on more thoroughly.

8.2.1. Describing States

We extend on the concept of a state vector $|\Psi\rangle$ as introduced above. First, note that linear operators can be represented using the so-called *outer product representation*.

Definition 8.1. Let V, W be Hilbert spaces and let $v \in V, w \in W$. Then the outer product $|w\rangle \langle v|$ acting on $V \times W$ is defined as the operator which for every other vector $\hat{v} \in V$: satisfies:

$$(|w\rangle \langle v|) |\hat{v}\rangle = |w\rangle \langle v|\hat{v}\rangle = (\langle v|\hat{v}\rangle) |w\rangle$$

State vectors, however, only described pure states. As such, mixtures or statistical ensembles of different quantum states cannot be represented by state vectors. In order to describe these, we make use of the outer product representation and introduce the concept of a *density matrix*.

Definition 8.2. An ensemble of pure states $|\psi_i\rangle$, occurring with probabilities p_i, a so-called *mixed state*, can be represented by a density operator $\hat{\rho}$, defined by:

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

Density matrices possess very nice properties which makes the density operator formalism very useful. Some of the relevant properties of this are given in the following theorem.

Theorem 8.1. *Let $\hat{\rho}$ be a density operator. Then $\hat{\rho}$ satisfies the following properties:*

1. *$\hat{\rho}$ is positive semi-definite.*
2. *$\hat{\rho}$ has a trace of unity.*
3. *The evolution of a system through the unitary operator U is given by $U\rho U^\dagger$*
4. *For a measurement described by M_m, we find*[1]*:*

$$\mathbb{P}(M = m) = \mathrm{Tr}\left(M_m^\dagger M_m \rho\right)$$

After measurement, the state collapses to:

$$\frac{M_m \rho M_m^\dagger}{\mathrm{Tr}\left(M_m^\dagger M_m \rho\right)}$$

5. $\mathrm{Tr}\left(\rho^2\right) \leq 1$ *with equality if and only if ρ represents a pure state.*
6. *Given an observable A:*

$$\langle A \rangle = \mathrm{Tr}\left(\hat{\rho} A\right)$$

7. *if ρ represents the state of systems A and B, we can describe the state of system A with the reduced density operator ρ_A.:*

$$\rho_A = \mathrm{Tr}_B(\rho)$$

Where Tr_B *denotes the partial trace. The partial trace is defined as follows:*

$$\mathrm{Tr}_B(|a_1\rangle\langle a_2| \otimes |b_1\rangle\langle b_2|) \equiv |a_1\rangle\langle a_2| \mathrm{Tr}(|b_1\rangle\langle b_2|)$$

Proof. We only prove property 6.

$$Tr(\hat{\rho}A) = \sum_k \langle k|\hat{\rho}A|k\rangle = \sum_{i=1}^n p_i \sum_k \langle k|\left(|\psi_i\rangle\langle\psi_i|\right) A|k\rangle = \sum_{i=1}^n p_i \sum_k \langle\psi_i| A|k\rangle\langle k|\psi_i\rangle = \sum_{i=1}^n p_i \langle\psi_i| A|\psi_i\rangle$$

where the last step follows from the completeness of $\{k\}$. □

8.2.2. DESCRIBING OPERATORS

In this section, we proceed to define quantum operators more rigorously by approaching them axiomatically. But first, since these operators act on Hilbert spaces, we can define an inner product between them as follows:

Definition 8.3. Given two operators A, B on a finite Hilbert space H. Then the mapping $\langle\cdot,\cdot\rangle$ defined on $H \times H$ as:

$$\langle A, B\rangle \equiv \mathrm{Tr}\left(A^\dagger B\right)$$

is called the Hilbert-Schmidt inner product

We now proceed with the axiomatic approach to quantum operators. Before introducing these axioms, some preliminary concepts need to be introduced first.

Definition 8.4. Let A be an operator acting on a Hilbert Space $\mathscr{H}$. Then A is called positive if $\forall\, v \in \mathscr{H}$, we have that $\langle v|Av\rangle \geq 0$.

[1] Note that $Tr(A) \equiv \sum_n A_{nn}$ denotes the *trace* of A.

Theorem 8.2. *(Nielsen and Chuang (2012)) Let A be a positive operator. Then A is self-adjoint, and therefore also Hermitian.*

Definition 8.5. A mapping $\Lambda : \mathscr{H}_{\mathscr{A}} \to \mathscr{H}_{\mathscr{B}}$ is called *positive* if $\forall\, a \geq 0$, $\Lambda(a) \geq 0$. Furthermore, Λ is called a *completely positive* mapping if it still yields positive mappings under composition with an arbitrary other system, i.e. if the mapping $I \otimes \Lambda : \mathscr{C} \otimes \mathscr{H}_{\mathscr{A}} \to \mathscr{D}$ is still positive, regardless of $\mathscr{C}$.

Definition 8.6. Let A be an operator acting on $\mathscr{H}_A \otimes \mathscr{H}_B$. Then A is called *decomposable* if $\exists\, P_1, P_2$ are completely positive maps such that:

$$A = P_1 + P_2 \otimes T$$

where T denotes the transposition map.

Having introduces these preliminary concepts, every quantum operator Λ is described by the following axioms:

Axiom I: Given an initial state ρ, $\mathrm{Tr}\big(\Lambda(\rho)\big)$ yields the probability that the process described by Λ occurs.

Axiom II: The map Λ is a *convex linear map* on the set of density operators.

Axiom III: Λ is a completely positive map.

Note that this axiomatic approach is equivalent to the approach taken before. This can be explicated in the following theorem.

Theorem 8.3. *(Nielsen and Chuang (2012)) Λ satisfies the axioms given above if and only if $\Lambda(\rho)$ can be represented as:*

$$\Lambda(\rho) = \sum_i A_i \rho A_i^\dagger$$

where $\{A_i\}$ is a set of operators satisfying $\sum E_i^\dagger E_i \leq I$.

Two important maps can be introduced, which will prove to be relevant for entanglement detection later on.

Definition 8.7. The map $\rho^{T_B} = I \otimes \theta : \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}} \to \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$ where $\theta : B \mapsto B^T$ is called the *partial transpose* map.

Remark. The transpose map is a *positive*, but not completely positive, map. Therefore, the partial transpose is not necessarily a positive map.

Using this knowledge, we can classify states according to the positivity of their partial transpose as follows:

Definition 8.8. Let $\rho \in \mathscr{H}_{\mathscr{A}} \otimes \mathscr{H}_{\mathscr{B}}$. If ρ^{T_A} is a positive map, so if ρ has a *Positive Partial Transpose*, we call ρ *PPT*. If ρ is not PPT, we call it *NPT*.

8.2.3. SCHMIDT DECOMPOSITIONS

Since quantum states and operators are represented using linear algebraic objects, these can be decomposed in many ways. One decomposition that will prove to be of utmost importance throughout our enquiry is given below, namely the Schmidt Decomposition.

Theorem 8.4. ***Schmidt Decomposition.*** *(Gühne and Tóth (2008)) Given a pure state* $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ *of a composite system* AB*. Then there exists an orthonormal basis* $|i_A\rangle$ *for system* A *and an orthonormal basis* $|i_B\rangle$ *for system* B *such that, given* $d = \min\{\dim \mathcal{H}_A, \mathcal{H}_B\}$*:*

$$|\psi\rangle = \sum_{i=1}^{d} \lambda_i |i_A\rangle \otimes |i_B\rangle$$

where $\lambda_i \geq 0 \forall i$ *are called the Schmidt Coefficients such that* $\sum_i \lambda_i = 1$*. Furthermore,* $S = \#\{\lambda_i \mid \lambda_i > 0\}$ *is called the Schmidt Rank of the wave function* $|\psi\rangle$*.*

Note that the concept of a Schmidt decomposition can be extended to density matrices of bipartite systems as follows:

Theorem 8.5. ***Schmidt Decomposition.*** *(Gühne and Tóth (2008)) Given a density matrix* $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$*. Then a basis* G_k^A *of* $\mathcal{H}_A$*,* G_k^B *of* $\mathcal{H}_B$ *consisting of Local Orthogonal Observables (LOO) exists, i.e.* $\langle G_i^A, G_j^A\rangle = \langle G_i^B, G_i^B\rangle = \delta_{ij}$*, such that:*

$$\rho = \sum_k \lambda_k G_k^A \otimes G_k^B$$

with $\lambda_k \geq 0$*.*

8.3. THE QUANTUM CAUCHY-SCHWARZ INEQUALITY

In the language of quantum mechanics, we retrieve the following formulation for the Cauchy-Schwarz Inequality:

Theorem 8.6. ***Cauchy-Schwarz Inequality.*** *(Nielsen and Chuang (2012)) Given* $|\psi\rangle$*,* $|\phi\rangle$*, then:*

$$|\langle\psi|\phi\rangle|^2 \leq \langle\psi|\psi\rangle\langle\phi|\phi\rangle$$

Furthermore, the Cauchy-Schwarz inequality can be adapted to operators, as these satisfy the Hilbert-Schmidt inner product.

Theorem 8.7. ***Cauchy-Schwarz for expectations.*** *(Wölk et al. (2014)) Given a density matrix* ρ *and two operators* A, B*, both acting on the same Hilbert space. Then:*

$$|\mathrm{Tr}(AB\rho)|^2 \leq \mathrm{Tr}\left(AA^\dagger\rho\right)\mathrm{Tr}\left(B^\dagger B\rho\right)$$

Proof. Note that $\langle\hat{A}, \hat{B}\rangle \equiv Tr(\hat{A}^\dagger\hat{B})$ is the Hilbert-Schmidt inner product. Setting $\hat{A} = \sqrt{\rho}A^\dagger, \hat{B} = B\sqrt{\rho}$, applying Cauchy-Schwarz' inequality and applying the cyclic property of the Trace operator yields the result. □

A famous application of the Cauchy-Schwarz inequality is in proving the well-known uncertainty relation of Heisenberg, of which we present a proof below.

Theorem 8.8. ***Heisenberg's uncertainty principle.*** *Let* A, B *be Hermitian operators and let* $|\psi\rangle$ *denote a quantum state. Then:*

$$\sigma(A)\sigma(B) \geq \frac{|\langle\psi|[A,B]|\psi\rangle|}{2}$$

Proof. First consider the operators $\hat{A} = A - <A>$, $\hat{B} = B - <B>$. Then:

$$|\langle\psi|[\hat{A},\hat{B}]|\psi\rangle|^2 + |\langle\psi|\{\hat{A},\hat{B}\}|\psi\rangle|^2 = 4|\langle\psi|\hat{A}\hat{B}|\psi\rangle|^2$$

Since:

$$
\begin{aligned}
|\langle\psi|\,\hat{A}\hat{B}\pm\hat{B}\hat{A}\,|\psi\rangle|^2 &= |\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle \pm \langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle|^2 \\
&= \left(\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle \pm \langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle\right)\overline{\left(\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle \pm \langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle\right)} \\
&= |\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle|^2 \pm \langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle\overline{\langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle} \pm \overline{\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle}\langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle + |\langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle|^2 \\
&= 2|\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle|^2 \pm \langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle\overline{\langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle} \pm \overline{\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle}\langle\psi|\,\hat{B}\hat{A}\,|\psi\rangle
\end{aligned}
$$

where the last step follows because the hermitian conjugation conserves lengths for scalars (inner products) and because A, B are hermitian. But now:

$$
|\langle\psi|\,\hat{A}\hat{B}\,|\psi\rangle|^2 \geq \frac{|\langle\psi|\,[\hat{A},\hat{B}]\,|\psi\rangle|^2}{4}
$$

Taking the square root of both sides and substituting $\hat{A} = A - \langle A\rangle\,, \hat{B} = B - \langle B\rangle$ finishes the proof. □

Note that the other inequalities presented in the previous chapters can also be written in terms of the quantum mechanical formalism. This will not be done here, since the quantum mechanical representation of these inequalities is considered to be very evident. Nevertheless, they will prove to be very important - the definition of the density matrix as the convex sum of outer products hints at the use of Jensen's Inequality or the AM-GM inequality, for example. This will become more clear as we turn to the topic of quantum entanglement, which will be introduced in the next chapter.

9

An Enquiry into Entanglement Detection

In this chapter, an enquiry will be made into the various detection methods for quantum entanglement. But, prior to that, the concept of entanglement will be introduced first. This chapter will be divided into two sections - one on bipartite entanglement and one on multipartite entanglement - as this approach offers more insight in the topic of entanglement detection.

9.1. Bipartite Entanglement and its detection

We first introduce the definition of entanglement for two-particle states and then proceed to discuss the detection of bipartite entanglement.

9.1.1. Introduction to Bipartite Entanglement

Let us first introduce the definition of separability and entanglement for bipartite systems.

Definition 9.1. Given a pure bipartite state $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$. $|\Psi\rangle$ is called *separable* if it can be written as the tensor product of pure states, i.e. if $\exists \left|\psi_1\right\rangle \in \mathcal{H}_1 and \left|\psi_2\right\rangle \in \mathcal{H}_2$, such that:

$$|\Psi\rangle = \left|\psi_1\right\rangle \otimes \left|\psi_2\right\rangle$$

If $|\Psi\rangle$ is not separable, it is called *entangled*.

An example of a bipartite entangled state is for example the ψ_- *Bell state*, which is given by:

$$\left|\psi_-\right\rangle = \frac{|10\rangle - |01\rangle}{\sqrt{2}}$$

This result can be generalised to mixed bipartite systems.

Definition 9.2. Let $\rho \in \mathcal{H}_1 \otimes \mathcal{H}_2$ be a density operator. We then distinguish three classes of density matrices.

1. If $\rho = \rho^A \otimes \rho^B$, then ρ is called a *product state*.
2. If ρ can be written as a convex sum of product states, i.e. $\rho = \sum_i p_i \rho_i^A \otimes \rho_i^B$, then ρ is called *separable*. Note that any product state is also separable.
3. If ρ is not separable, ρ is called *entangled*.

Given some mixture of quantum states - so, some density matrix - the question whether this mixture is separable or entangled follows naturally, especially considering that this knowledge is of great importance for many applications in for example quantum computing. This question results in the so-called *separability problem*, which has not been resolved up to this date (Gühne and Tóth (2008)). Some criteria, however, have been derived - these will be discussed below.

9.1.2. BIPARTITE ENTANGLEMENT DETECTION

There are various methods for the detection of entanglement. In this section, the most important bipartite entanglement detection methods will be treated. We first discuss detection methods which are dependent on performing direct action on the given mixture.

THE PPT, REDUCTION AND MAJORISATION CRITERIA

We first give three criteria and prove them, and then show how they relate to one another. Let us first state and prove the so-called PPT criterion (Wolf (2003), Gühne and Tóth (2008))

Theorem 9.1. ***(PPT Criterion)*** *Let ρ be a bipartite, separable state. Then ρ is PPT.*

Proof. Let ρ be separable, i.e. $\rho = \sum_{i=1}^{n} pi\rho_i^A \otimes \rho_i^B$. We can take the partial transpose:

$$\rho^{T_A} = \sum_{i=1}^{n} p_i \left(\rho_i^A\right)^T \otimes \rho_i^B$$

Note that this again is a density matrix and as such a positive definite operator. Thus ρ is PPT. □

If the converse of the implication were also true, the separability problem for bipartite states could be reduced to the question whether these states are PPT. This is, unfortunately, not always the case. There are some cases of great practical relevance - like two-qubit systems - for which this is the case, however:

Theorem 9.2. ***(Horodecki's Theorem)*** *(Horodecki et al. (1996)) Let ρ be a $2x2$ or a $2x3$ system. If ρ is PPT, then ρ is separable.*

Another criterion that is related to positive but not completely positive maps is based on the so-called *Reduction map*.

Theorem 9.3. ***(Reduction criterion)*** *Let $\Lambda^R(X) = \mathrm{Tr}(X)\mathbb{1} - X$. If ρ is separable, then $I \otimes \Lambda^R(\rho) \geq 0$.*

One may wonder how this criterion relates to the PPT criterion in terms of its detection strength. The following theorem shows that the Reduction criterion is actually weaker than the PPT criterion:

Theorem 9.4. *Let ρ be entangled and detectable by the Reduction criterion. Then ρ is detectable by the PPT criterion.*

Proof. First of all, note that Λ^R is decomposable (see Horodecki and Horodecki (1999)), i.e. $\exists P_1, P_2$ completely positive maps such that $\Lambda^R = P_1 + P_2 \otimes T$. But since P_1 and P_2 are completely positive, ρ can only be detected if $I \otimes T$ is not positive. □

Note that decomposability is not unique to this specific positive map - it can be generalised even further.

Theorem 9.5. *(Woronowicz (1976)) Let $\Lambda : \mathcal{H}_A \to \mathcal{H}_B$ with $\dim \mathcal{H}_a = 2$ and $\dim \mathcal{H}_B \leq 3$. Then Λ is decomposable.*

It can actually be shown that any positive but not completely positive map is completely positive over the separable states. This yields an interesting theorem, of which the Reduction Criterion is just a corollary.

Theorem 9.6. *(Horodecki et al. (1996)) ρ is separable $\iff$ $\forall \Lambda$ positive, we have that $(I \otimes \Lambda)(\rho) \geq 0 \quad \forall \rho$.*

Thus, the separability problem reduces to the identification of all positive maps. From this, it easily follows that for low dimensions (at most 2 x 3), the PPT criterion is a necessary and sufficient condition for entanglement, and as such, proves the Horodecki criterion.

Furthermore, there is another criterion that - at first sight - might appear to be completely unrelated, which we state here without proof.

Theorem 9.7. ***(Majorisation Criterion)*** *(Hiroshima (2003)) Let $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$ be a density matrix and consider $\rho^A \equiv \mathrm{Tr}_B(\rho)$. Consider the eigenvalues of ρ arranged in decreasing order, say $P = \left(p_1, p_2 \ldots\right)$ and analogously those of ρ^A, say $Q = \left(q_1, q_2 \ldots\right)$. If ρ is separable, then $\forall k$:*

$$\sum_{i=1}^{k} p_i \leq \sum_{i=1}^{k} q_i$$

The question how the majorisation criterion relates to the reduction criterion arises naturally. This question is addressed by the following theorem.

Theorem 9.8. *(Hiroshima (2003)) Any entangled state that is detected by the majorisation criterion can also be detected by the reduction criterion.*

We thus see that the PPT criterion is the stronger one of these criteria. There are, however, criteria that can detect states which are left undetected by the PPT criterion. One of such entangled PPT states is (de Vicente (2007)):

$$\rho = \frac{1}{36}\begin{bmatrix} \frac{7}{2} & \frac{7}{2} & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ \frac{7}{2} & \frac{7}{2} & -1 & -1 & -1 & \frac{7}{2} & -1 & -1 & -1 \\ -1 & -1 & \frac{7}{2} & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & \frac{7}{2} & -1 & -1 & \frac{7}{2} & -1 & -1 \\ -1 & -1 & -1 & -1 & 8 & -1 & -1 & -1 & -1 \\ -1 & \frac{7}{2} & -1 & -1 & -1 & \frac{7}{2} & -1 & -1 & -1 \\ -1 & -1 & -1 & \frac{7}{2} & -1 & -1 & \frac{7}{2} & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & \frac{7}{2} & \frac{7}{2} \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & \frac{7}{2} & \frac{7}{2} \end{bmatrix}$$

Some of these criteria are discussed next.

THE CCNR CRITERION, THE CMC CRITERION AND OTHERS

We start off by introducing a criterion which is based on the Schmidt decomposition in operator space: the so-called *Computable Cross Norm or Realignment Criterion.*

Theorem 9.9. ***(CCNR Criterion)*** *(Chen and Wu (2003)). Let ρ be a density matrix with the following Schmidt decomposition: $\rho = \sum_k \lambda_k G_k^A \otimes G_k^B$, where $\{G_k^A\}$ forms an orthonormal basis of $\mathcal{H}_A$ and $\{G_k^B\}$ forms an orthonormal basis of $\mathcal{H}_B$. If ρ is separable, then:*

$$\sum_k \lambda_k \leq 1$$

A similar criterion is the so-called *Covariance Matrix Criterion*.

Theorem 9.10. ***(CMC Criterion)*** *(Gittsovich et al. (2008)) Let G_i^A, G_j^B denote local orthogonal observables acting on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Then by forming the block matrix γ by setting $\gamma_{ij} = \langle G_i^A G_j^B \rangle - \langle G_i^A \rangle \langle G_j^B \rangle$. Then if $\rho \in \mathcal{H}_A \otimes \mathcal{H}_B$ is separable, one can find states $\psi_k^A \in \mathcal{H}_A$ and $\psi_k^k \in \mathcal{H}_B$ such that:*

$$\gamma \geq \begin{bmatrix} \sum_k p_k \gamma \left| \psi_k^A \right\rangle & 0 \\ 0 & \sum_k p_k \gamma \left| \psi_k^B \right\rangle \end{bmatrix}$$

We note that these criteria are intimately connected, as is shown in the following theorem:

Theorem 9.11. *(de Vicente (2007)) Any state ρ detected by the CNC criterion is also detected by the CCNR criterion.*

Note that the CCNR and the PPT criterion do not detect the same states. There is, however, a criterion, which relates these two criteria based on the idea of linear contractions and permutations.

Theorem 9.12. *(Horodecki et al. (2006)) Given a density matrix ρ expanded in a product basis, so $\rho = \sum_{ij,kl} \rho_{ijkl} \left|i\right\rangle \left\langle j\right| \otimes \left|k\right\rangle \left\langle l\right|$. Then the following two separability criteria can be derived:*

$$\|\rho_{(ijlk)}\| \leq 1 \quad \textit{and} \quad \|\rho_{(ikjl)}\| \leq 1$$

where $(\dots)$ indicates a permutation of ijkl and the Hilbert-Schmidt norm is used.

Note that the first of the two separability requirements is just the PPT criterion, while the second criterion comes down to the CCNR criterion.

BIPARTITE WITNESS DETECTION

The methods presented up until here all presuppose knowledge of the mixture of states at hand. In this section, a class of methods that do not require such knowledge will be presented: *entanglement witness detection methods.* We first define a witness.

Definition 9.3. Let W be an observable. W is called an *entanglement witness* if:

$$\begin{aligned} \forall \rho_s \text{ separable} &: \mathrm{Tr}\left(W \rho_s\right) \geq 0 \\ \exists \rho_e \text{ entangled} &: \mathrm{Tr}\left(W \rho_e\right) < 0 \end{aligned}$$

In fact, all entangled states can be detected by some witness:

Theorem 9.13. *(Horodecki et al. (1996)) Let ρ be an entangled state. Then there is an entanglement witness W which detects ρ.*

We thus find another reformulation of the separability problem: if we know how to construct a entanglement witness for any entangled state, we can always determine whether a state is separable or not.
We can now discuss two examples of constructions of witnesses, namely for states that violate the PPT criterion and for states that violate the CCNR criterion.

Theorem 9.14. *(Gühne and Tóth (2008)) Let ρ be NPT with a negative eigenvalue λ_- for some eigenvector $\left|\eta\right\rangle$ of ρ^{T_A}. Then:*

$$W = \left|\eta\right\rangle \left\langle \eta\right|^{T_A}$$

is a witness detecting ρ.

We can generalise this theorem to any positive map as follows:

Theorem 9.15. *(Gühne and Tóth (2008)) Let ρ be an entangled state detected by the positive operator Λ. Then:*

1. *$\mathbb{1} \otimes \Lambda(\rho)$ has some eigenvalue $\lambda_- < 0$ for some eigenvector $|\eta\rangle$, and*

2. *$W = I \otimes \Lambda^*(|\eta\rangle\langle\eta|)$ is a witness which detects ρ.*

We can also derive a witness based on the CCNR criterion.

Theorem 9.16. *(Yu and Liu (2005)) Let ρ violate the CCNR criterion. Then:*

$$W = I - \sum_k G_k^A \otimes G_k^B$$

is a witness that detects ρ.

The question might rise whether there is a connection between entanglement detection via positive maps and entanglement detection using witnesses. This question can be addressed by studying a specific isomorphism, which will be discussed in the next section.

CONNECTING POSITIVE MAPS AND WITNESSES

In the search of a connection between the maps between two spaces and the operators acting on their tensor product, a logical step would be to try to find an isomorphism between these two classes of objects. We introduce one such isomorphism below.

Definition 9.4. Consider the operators E acting on $L(\mathcal{H}_A) \otimes L(\mathcal{H}_B)$ and consider the maps $\epsilon : L(\mathcal{H}_A) \to L(\mathcal{H}_B)$. Then the *Choi-Jamiołkowski isomorphism* is given by:

$$\epsilon(\rho) = \mathrm{Tr}_A(E\rho^T \otimes \mathbb{1}_B)$$

The relevance of the Choi-Jamiołkowski isomorphism will become clear from the next theorem.

Theorem 9.17. *(Choi (1975)) Let E be an operator acting on $L(\mathcal{H}_A) \otimes L(\mathcal{H}_B)$ and consider the map $\epsilon : L(\mathcal{H}_A) \to L(\mathcal{H}_B)$. Suppose that ϵ and E are mapped onto each other by the Choi-Jamiołkowski isomorphism. Then the following properties hold:*

1. *ϵ is CP $\iff$ E is a positive semi-definite operator.*

2. *ϵ is positive but not CP $\iff$ E is an entanglement witness.*

3. *ϵ is decomposable $\iff$ E is a decomposable entanglement witness.*

This theorem shows that the Choi-Jamiołkowski isomorphism relates positive but not completely positive maps to entanglement witnesses, effectively linking the maps studied in the previous sections like the partial transpose to the witnesses discussed later on. As such, we see that the separability problem can also be interpreted to be the classification of all entanglement witnesses.

NON-LINEAR WITNESS DETECTION

Note that the inequalities that were derived up until now for witness detection methods are all inequalities in the mean of the given witness. These inequalities can, however, be extended to include non-linearities in the form of e.g. the variance of the observables at hand. For example, all witnesses of the form presented in Theorem 9.14 can be given a nonlinear extension (see: Gühne and Lütkenhaus (2005)). This extension offers significant improvement over the linear entanglement criteria. We will discuss several of these non-linear criteria in this chapter, and will derive some other non-linear criteria using the inequalities studied in the previous chapters in the next chapter.

Two of these non-linear criteria were derived by **Hillery and Zubairy**.

Theorem 9.18. *Let A, B be operators acting on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Then for separable states ρ we have:*

$$|\langle A^\dagger \otimes B\rangle|^2 \leq \langle A^\dagger A \otimes B^\dagger B\rangle$$
$$|\langle A^\dagger \otimes B\rangle|^2 \leq \langle AA^\dagger\rangle \langle B^\dagger B\rangle$$

Another bipartite entanglement criteria was derived by **Shchukin and Vogel**.

Theorem 9.19. *() Let a, b denote the annihilation operators in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Then:*

$$|\langle a^{\dagger m} a^n a^{\dagger p} a^q \otimes b^{\dagger s} b^r b^{\dagger k} b^l\rangle| \leq \langle a^{\dagger m} a^n a^{\dagger n} a^m \otimes b^{\dagger l} b^k b^{\dagger k} b^l\rangle \langle a^{\dagger q} a^p a^{\dagger p} a^q \otimes b^{\dagger s} b^r b^{\dagger r} b^s\rangle$$

Lastly, a criterion has been derived by **Huber et al.**

Theorem 9.20. *Let Π denote the permutation operator and let Φ denote a product state of an m-tupled system for some $m \in \mathbb{N}_{\geq 2}$. Then:*

$$\sqrt{\mathrm{Re}\left(\langle\Phi| (\mathbb{1}_A \otimes \Pi_B)^\dagger \rho^{\otimes m} (\Pi_A \otimes \mathbb{1}_B) |\Phi\rangle\right)} \leq \sqrt{\langle\Phi| \rho^{\otimes m} |\Phi\rangle}$$

It can be shown that the cases for $m \geq 3$ can be brought back to the case for $m = 2$ with a different product state Φ. We therefore restrict ourselves to the case for $m = 2$.

9.2. MULTIPARTITE ENTANGLEMENT AND ITS DETECTION

In this section we introduce entanglement and entanglement detection methods for multipartite systems. This will not be done in a completely analogous manner to the two-particle case, as extra particles introduce more complex structures than the bipartite case, since multiple degrees of entanglement become possible.

9.2.1. INTRODUCTION INTO MULTIPARTITE ENTANGLEMENT

We start off by introducing the concept of separability, and from there on delve into the possible forms of entanglement.

Definition 9.5. Given a pure multipartite state $|\Psi\rangle \in \bigotimes_{i=1}^n \mathcal{H}_i$. $|\Psi\rangle$ is called *fully separable* if it can be written as the tensor product of pure states, i.e. if $\exists |\psi_1\rangle \in \mathcal{H}_1, \ldots, |\psi_n\rangle \in \mathcal{H}_n$, such that:

$$|\Psi\rangle = \bigotimes_{i=1}^n |\psi_i\rangle$$

Any state that is not fully separable, is somewhat entangled. We now introduce structure into the degrees of entanglement possible. First, we introduce a weaker form of separability.

Definition 9.6. Let $|\Psi\rangle \in \bigotimes_{i=1}^{n} \mathcal{H}_i$. $|\Psi\rangle$ is called *m-separable* with $1 < m < n$ if there are m parts $P_1, \ldots, P_m$ such that:

$$|\Psi\rangle = \bigotimes_{i=1}^{m} |\phi\rangle_{P_i}$$

We call $|\Psi\rangle$ *truly n-partite entangled* if it is neither fully separable nor m-separable.

Let us now introduce some examples of important multipartite entangled states. We first start off with a class of states that is very strongly entangled: the Greenberger–Horne–Zeilinger (GHZ) state.

Definition 9.7. The GHZ-state for n qubits is:

$$|GHZ_n\rangle = \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}}$$

Another class of entangled qubit states are called Dicke-states.

Definition 9.8. The (n,k) symmetric Dicke state on n qubits is defined as:

$$\left|D_{k,n}\right\rangle = \frac{\sum_j P_j\left(|1\rangle^{\otimes k} \otimes |0\rangle^{\otimes n-k}\right)}{\sqrt{\binom{n}{k}}}$$

where $P_j()$ denotes a permutation over the order of the qubits.

To make these definitions more tangible, let us consider their tripartite versions. For three qubits, the GHZ-state is becomes:

$$|GHZ_3\rangle = \frac{|000\rangle + |111\rangle}{\sqrt{2}}$$

and one of the Dicke states, the (1,3)-Dicke state (which is also known as the W state), becomes:

$$|W_3\rangle = \frac{|100\rangle + |010\rangle + |001\rangle}{\sqrt{3}}$$

9.2.2. MULTIPARTITE ENTANGLEMENT DETECTION METHODS

In this section, we will primarily focuss on multipartite witness detection, as this will prove to be relevant for our enquiry. Nevertheless, we first introduce a multipartite entanglement criterion which generalises the PPT and CCNR criteria by using permutations.

MULTIPARTITE ENTANGLEMENT PERMUTATION CRITERIA

Theorem 9.21. ***(Permutation Criteria)*** *(Horodecki et al. (2006)) Given ρ describing an n-dimensional system with an expansion in a product basis of the following form:*

$$\rho = \sum_{i_1, j_1, \ldots i_n, j_n} \rho_{i_1, j_1, \ldots i_n, j_n} \bigotimes_{k=1}^{n} |i_k\rangle \langle j_k|$$

Then ρ is separable if for all permutations (denoted as σ), we have:

$$\|\rho_{\sigma(i_1, j_1, \ldots i_n, j_n)}\| \leq 1$$

MULTIPARTITE WITNESS DETECTION

We can also introduce the concept of entanglement witnesses for multipartite systems. It should first be noted that several types of witnesses exist for different types of multipartite entanglement. One of these witnesses are GHZ-witnesses.

Definition 9.9. W_{GHZ} is called a *GHZ-class witness* if:

$$\forall \rho \text{ not GHZ} : \mathrm{Tr}\big(W_{GHZ}\rho\big) \geq 0 \tag{9.1}$$

$$\exists \rho_g \text{ GHZ} : \mathrm{Tr}\big(W_{GHZ}\rho_g\big) < 0 \tag{9.2}$$

This definition can be adapted to derive witnesses for other classes. These witnesses can be easily constructed as follows. Let $\left|psi\right\rangle$ denote the state that we want the witness to detect. Then

$$W_\psi = \alpha \mathbb{1} - \left|\psi\right\rangle \left\langle\psi\right|$$

Then W_ψ is a witness detecting $\left|\psi\right\rangle$ with α denoting the maximum overlap between $\left|\psi\right\rangle$ and other inseparable states. Note that this allows for a degree of freedom, as said inseparable states can be chosen to be e.g. biseparable or fully separable states. One can derive, for example, a witness that detects genuine multipartite detection in specific Dike states, namely $\left|D_{\frac{N}{2},N}\right\rangle$ to be:

$$W_D = \frac{N}{2(N-1)}\mathbb{1} - \left|D_{\frac{N}{2},N}\right\rangle \left\langle D_{\frac{N}{2},N}\right|$$

Note that these linear witness criteria are very basic, and can be extended greatly, for example by considering stabiliser witnesses. This is, however, beyond the scope of this thesis.

We now consider five cases of non-linear witnesses for multipartite systems.
We first start off by considering a criterion for tripartite systems that was derived by **Gühne and Seevinck**.

Theorem 9.22. *Consider the density matrix* ρ*, expressed in the standard basis* $(|000\rangle, |001\rangle, \ldots, |111\rangle)$. *Then for separable states and biseparable qubit states, we find that:*

$$|\rho_{000,111}| \leq \sqrt{\rho_{001,001}\rho_{110,110}} + \sqrt{\rho_{010,010}\rho_{101,101}} + \sqrt{\rho_{011,011}\rho_{100,100}}$$

Note that this Theorem also has practical uses, as it provides a both necessary and sufficient condition for full separability of GHZ-states that are mixed with white noise (Gühne and Seevinck (2018)). We will prove this theorem in a later chapter.
Furthermore, it is suggested in Wölk et al. (2014) that the condition by Gühne and Seevinck can be extended using the AM-GM inequality (Theorem 3.1). Note, however, that *this extension is of little use* in our attempt to detect non-separable states, as this only loosens the upper bound that has been derived, and as such, will fail to detect even more not completely separable states.

We now turn to more general multipartite systems. First of all, the criterion by **Hillary and Zubairy** can be generalised to multipartite systems.

Theorem 9.23. *Let* A_k *be operators acting on* $\mathscr{H}_k$ *for* $k = 1, \ldots, N$. *Then:*

$$\|\langle \bigotimes_{k=1}^{N} A_k \rangle|^2 \leq \langle \bigotimes_{k=1}^{j} A_k^\dagger A_k \rangle \langle \bigotimes_{k=j+1}^{N} A_k A_k^\dagger \rangle$$

is satisfied by separable states as well as biseparable states with respect to the partition $(1, 2, \ldots, j), (j+1, \ldots, N)$.

In Hillery et al. (2010), **Hillary et al.** sought to derive stronger versions of this extension. Note that these criteria can proved using the inequalities that we have studied. We present two of them here.

Theorem 9.24. *Consider operators A_k acting on $\mathscr{H}_k$ for $k = 1, \ldots, n$. Then for separable states, the following holds:*

$$|\langle \bigotimes_{k=1}^{N} A_k \rangle| \leq \prod_{k=1}^{N} \langle \left(A_k^\dagger A_k \right)^{\frac{n}{2}} \rangle^{\frac{1}{n}}$$

Theorem 9.25. *Consider operators A_k acting on $\mathscr{H}_k$ for $k = 1, \ldots, n$. Then for separable states, the following holds:*

$$|\langle \bigotimes_{k=1}^{N} A_k \rangle|^n \leq \langle \left(\frac{1}{n} \sum_{k=1}^{N} A_k^\dagger A_k \right)^{\frac{n}{2}} \rangle$$

The criterion by **Huber et al.** can also be extended to the multipartite case. We specifically consider the generalisation for $m = 2$.

Theorem 9.26. *For all biseparable states, we find that:*

$$\sqrt{\langle \Phi | \rho^{\otimes 2} \Pi | \Phi \rangle} \leq \sum_j \sqrt{\langle \Phi | \mathscr{P}_j^\dagger \rho^{\otimes 2} \mathscr{P}_j | \Phi \rangle}$$

where Π acts on all subsystems simultaneously, while $\mathscr{P}_j$ only acts on the j^{th} subsystem.

In the next chapter, we derive some non-linear witnesses based on the inequalities studied in previous chapters, and then assess these criteria by comparing them to the ones presented here.

10

Entanglement Criteria based on Cauchy-Schwarz and Hölder

In this chapter, we will present entanglement criteria developed by Wölk, Huber and Gühne. Just like in the previous chapter, we first start by presenting the two-particle case first. Afterwards, the results from these two-particle states will be extended to multipartite systems, after which we will compare these results with some of the criteria discussed in the previous chapter.

10.1. A Bipartite Entanglement Criterion

In this section, we present two upper bounds for quantum states: one that holds for all states, and a second one which only holds for separable states. Throughout this section, states $\psi \in \mathcal{H}_A \otimes \mathcal{H}_B$ as well as operators A_1, A_2 acting on $\mathcal{H}_A$ as well as operators B_1, B_2 acting on $\mathcal{H}_B$ will be considered.
We now present and prove an upper bound which holds for all states.

Theorem 10.1. *For all operators A_1, A_2, B_1, B_2, the following inequality holds:*

$$|\langle A_1 A_2 \otimes B_1 B_2 \rangle|^2 \leq \langle A_1 A_1^\dagger \otimes B_1 B_1^\dagger \rangle \langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \rangle$$

Proof. Note that this result follows trivially from the Cauchy-Schwarz Inequality on the Hilbert-Schmidt norm (Theorem 8.7) by taking $A = A_1 \otimes B_1, B = A_2 \otimes B_2$. □

We now present and prove a result which only holds for separable states.

Theorem 10.2. *For separable states and for all operators A_1, A_2, B_1, B_2, the following inequality holds:*

$$|\langle A_1 A_2 \otimes B_1 B_2 \rangle|^2 \leq \langle A_1 A_1^\dagger \otimes B_2^\dagger B_2 \rangle \langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger \rangle$$

Proof. First assume $\left|\psi\right\rangle = |a\rangle \otimes |b\rangle$. Then:

$$|\langle a \otimes b| A_1 A_2 \otimes B_1 B_2 |a \otimes b\rangle|^2 = |\langle a| A_1 A_2 |a\rangle|^2 |\langle b| B_1 B_2 |b\rangle|^2 \leq \langle a| A_1 A_1^\dagger |a\rangle \langle a| A_2^\dagger A_2 |a\rangle \langle b| B_1 B_1^\dagger |b\rangle \langle b| B_2^\dagger B_2 |b\rangle$$

where the first step follows from the separability of $\left|\psi\right\rangle$ and the inequality follows from Theorem 8.7. The proof for pure states is finished by realising that:

$$\begin{aligned}\langle A_1 A_1^\dagger \otimes B_2^\dagger B_2 \rangle \langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger \rangle &= \langle a| A_1 A_1^\dagger |a\rangle \langle b| B_2^\dagger B_2 |b\rangle \langle a| A_2^\dagger A_2 |a\rangle \langle b| B_1 B_1^\dagger |b\rangle \\ &= \langle a| A_1 A_1^\dagger |a\rangle \langle a| A_2^\dagger A_2 |a\rangle \langle b| B_1 B_1^\dagger |b\rangle \langle b| B_2^\dagger B_2 |b\rangle\end{aligned}$$

Let us now consider mixed states. Set $\rho = \sum_{k=1}^{n} p_k \rho_k$, where $\rho_k \equiv |\psi_k\rangle\langle\psi_k|$ for pure states $|\psi_k\rangle$. Then each ρ_k represents a pure state, thus:

$$\langle A_1 A_2 \otimes B_1 B_2\rangle = \mathrm{Tr}\left((A_1 A_2 \otimes B_1 B_2 \rho)\right) = \mathrm{Tr}\left((A_1 A_2 \otimes B_1 B_2)\sum_{k=1}^{n} p_k \rho_k\right) = \sum_{k=1}^{n} p_k \mathrm{Tr}\left((A_1 A_2 \otimes B_1 B_2)\,\rho_k\right)$$

One can now deduce that:

$$\begin{aligned}
\sum_{k=1}^{n} p_k \mathrm{Tr}\left((A_1 A_2 \otimes B_1 B_2)\,\rho_k\right) &\le \sum_{k=1}^{n} p_k \sqrt{\mathrm{Tr}\left(\left(A_1 A_1^\dagger \otimes B_2^\dagger B_2\right)\rho_k\right)\mathrm{Tr}\left(\left(A_2^\dagger A_2 \otimes B_1 B_1^\dagger\right)\rho_k\right)} \\
&= \sum_{k=1}^{n} \sqrt{p_k \mathrm{Tr}\left(\left(A_1 A_1^\dagger \otimes B_2^\dagger B_2\right)\rho_k\right)} \cdot \sqrt{p_k \mathrm{Tr}\left(\left(A_2^\dagger A_2 \otimes B_1 B_1^\dagger\right)\rho_k\right)} \\
&\le \sqrt{\sum_{k=1}^{n} p_k \mathrm{Tr}\left(\left(A_1 A_1^\dagger \otimes B_2^\dagger B_2\right)\rho_k\right)\sum_{k=1}^{n} p_k \mathrm{Tr}\left(\left(A_2^\dagger A_2 \otimes B_1 B_1^\dagger\right)\rho_k\right)} \\
&= \sqrt{\mathrm{Tr}\left(\left(A_1 A_1^\dagger \otimes B_2^\dagger B_2\right)\sum_{k=1}^{n} p_k \rho_k\right)\mathrm{Tr}\left(\left(A_2^\dagger A_2 \otimes B_1 B_1^\dagger\right)\sum_{k=1}^{n} p_k \rho_k\right)} \\
&= \sqrt{\mathrm{Tr}\left(\left(A_1 A_1^\dagger \otimes B_2^\dagger B_2\right)\rho\right)\mathrm{Tr}\left(\left(A_2^\dagger A_2 \otimes B_1 B_1^\dagger\right)\rho\right)} \\
&= \sqrt{\langle A_1 A_1^\dagger \otimes B_2^\dagger B_2\rangle\langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger\rangle}
\end{aligned}$$

where the two inequalities follow from the Cauchy-schwarz Inequality (Theorem 2.4). □

Any state that violates this criterion is thus entangled. Note that this criterion only works if the upper bound found for separable states is smaller than the upper bound we have derived for general states.

10.2. DERIVING MULTIPARTITE ENTANGLEMENT CRITERIA

In this section, we will derive a generalisation of the criterion derived in the previous section. Note that this derivation cannot be done completely analogously without loss of the strength of the criterion - that is, the criterion will not only be sensitive to (genuine) multipartite entanglement. We will first show why this is the case, and will then proceed to derive the criterion.

Let us first proceed by deriving a criterion in an analogous manner. Let us consider operators $A_k^1 A_k^2$ acting on Hilbert spaces $\mathcal{H}_k$ for $k = 1,\ldots,N$. For product states, we can then easily derive the following inequality - simply through substitution:

$$|\langle \bigotimes_{k=1}^{N} A_k^1 A_k^2 \rangle| \le \prod_{k=1}^{N} \sqrt{\langle A_k^1 (A_k^1)^\dagger\rangle\langle (A_k^2)^\dagger A_k^2\rangle}$$

We can then introduce any recombination of the operators on the right hand side, following a similar argument as was done in the bipartite case. We introduce a new index based on the permutation σ over the current set of indices, to find:

$$|\langle \bigotimes_{k=1}^{N} A_k^1 A_k^2 \rangle| \le \prod_{k=1}^{N} \sqrt{\langle A_k^1 (A_k^1)^\dagger (A_{\sigma(k)}^2)^\dagger A_{\sigma(k)}^2\rangle}$$

Note, however, that these inequalities are not convex, and as such do not generally hold for separable mixed states. As such, no criterion can be derived that guarantees to be valid for separable states in this fashion.

We thus derive another set of multipartite entanglement criteria, which is an explicit generalisation of our bipartite criterion, which is based on a development scheme presented in Wölk et al. (2014), in which we will use the **Cauchy-Schwarz inequality** (Theorem 2.4), **Holder's inequality** (Theorem 6.5) and **Jensen's inequality** (Theorem 4.1).

Theorem 10.3. *Given operators A_k^1, A_k^2 acting on a Hilbert Space $\mathscr{H}_k$ for $k = 1, \ldots, N$. Then for any permutation σ acting on $\{1, \ldots, N\}$ with ord$(\sigma) = n$, the following inequality holds for separable states:*

$$|\langle \bigotimes_{k=1}^{N} A_k^1 A_k^2 \rangle|^2 \leq \prod_{k=1}^{N} \left(\langle \left(A_k^1 \left(A_k^1\right)^\dagger (A_{\sigma(k)}^2)^\dagger A_{\sigma(k)}^2 \right)^{\frac{N}{2}} \rangle \right)^{\frac{2}{N}}$$

Proof. We first set $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$, where $|\psi_i\rangle$ are product states in $\bigotimes_{k=1}^N \mathscr{H}_k$. Because we are considering a separable state, we find that:

$$|\langle \bigotimes_{k=1}^{N} A_k^1 A_k^2 \rangle|^2 = |\sum_i p_i \langle \bigotimes_{k=1}^{N} A_k^1 A_k^2 \rangle_i |^2 \leq \sum_i p_i |\langle \bigotimes_{k=1}^{N} A_k^1 A_k^2 \rangle_i |^2$$

Where the final step follows from the triangle inequality for absolute values. We can now split the right hand side up in the same manner as was done in the bipartite case, as this concerns product states. We thus proceed by applying the Cauchy-Schwarz inequality:

$$\sum_i p_i \langle \bigotimes_{k=1}^{N} A_k^1 A_k^2 \rangle_i |^2 \leq \sum_i p_i |\prod_{k=1}^{N} \langle A_k^1 \left(A_k^1\right)^\dagger \rangle_i \langle \left(A_k^2\right)^\dagger A_k^2 \rangle_i |$$

We now introduce the permutation function σ to regroup the expected values:

$$\sum_i p_i |\prod_{k=1}^{N} \langle A_k^1 \left(A_k^1\right)^\dagger \rangle_i \langle \left(A_k^2\right)^\dagger A_k^2 \rangle_i | = \sum_i p_i |\prod_{k=1}^{N} \langle A_k^1 \left(A_k^1\right)^\dagger \left(A_{\sigma(k)}^2\right)^\dagger A_{\sigma(k)}^2 \rangle_i |$$

We now combine two of the results that we have derived earlier: a generalisation of Hölder's inequality (6.5) as well as the weighted Hölder's inequality (6.6). Using these results, we find:

$$\sum_i p_i |\prod_{k=1}^{N} \langle A_k^1 \left(A_k^1\right)^\dagger \left(A_{\sigma(k)}^2\right)^\dagger A_{\sigma(k)}^2 \rangle_i | \leq \prod_{k=1}^{N} \sum_i p_i |\langle A_k^1 \left(A_k^1\right)^\dagger \left(A_{\sigma(k)}^2\right)^\dagger A_{\sigma(k)}^2 \rangle_i^{\frac{N}{2}} |^{\frac{2}{N}}$$

We now make use of Jensen's inequality (4.1), which yields that for any positive operator P and any $x \geq 1$, we have that $\langle P \rangle^x \leq \langle P^x \rangle$, such that:

$$\prod_{k=1}^{N} \sum_i p_i |\langle A_k^1 \left(A_k^1\right)^\dagger \left(A_{\sigma(k)}^2\right)^\dagger A_{\sigma(k)}^2 \rangle_i^{\frac{N}{2}} |^{\frac{2}{N}} \leq \prod_{k=1}^{N} \sum_i p_i |\langle \left(A_k^1 \left(A_k^1\right)^\dagger \left(A_{\sigma(k)}^2\right)^\dagger A_{\sigma(k)}^2 \right)^{\frac{N}{2}} \rangle_i |^{\frac{2}{N}}$$

Note that the right hand side of the last inequality can be rewritten in terms of expected values over the mixed state, so:

$$\prod_{k=1}^{N} \sum_i p_i |\langle \left(A_k^1 \left(A_k^1\right)^\dagger \left(A_{\sigma(k)}^2\right)^\dagger A_{\sigma(k)}^2 \right)^{\frac{N}{2}} \rangle_i |^{\frac{2}{N}} = \prod_{k=1}^{N} \left(\langle \left(A_k^1 \left(A_k^1\right)^\dagger (A_{\sigma(k)}^2)^\dagger A_{\sigma(k)}^2 \right)^{\frac{N}{2}} \rangle \right)^{\frac{2}{N}}$$

Combining the first and the final result finishes this proof. □

Note that this is just one of many possibilities, as the decision to group the k values with just one of their permutations was made on an arbitrary basis - these can be grouped in any other manner, and as such, this derivation allows for criteria of different forms than the one presented above to be derived.
We present four of the many alternatives below which apply to tripartite systems. In the next chapter, these specific cases will prove their worth when we consider a special tripartite state and further assess our criteria.

Theorem 10.4. *Let A_1, A_2, B_1, B_2 and C_1, C_2 be operators acting on the Hilbert Spaces $\mathcal{H}_A, \mathcal{H}_B$ and $\mathcal{H}_C$, respectively. Then the following condition holds for all separable states:*

$$|\langle A_1A_2 \otimes B_1B_2 \otimes C_1C_2\rangle| \leq \Big(\langle A_1A_1^\dagger \otimes B_1B_1^\dagger \otimes C_2^\dagger C_2\rangle \langle A_1A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_1C_1^\dagger\rangle$$
$$\dots \langle A_2^\dagger A_2 \otimes B_1B_1^\dagger \otimes C_1C_1^\dagger\rangle \langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle\Big)^{\frac{1}{4}}$$

Proof. We proceed by following the scheme presented in the proof of the previous theorem to find:

$$|\langle A_1A_2 \otimes B_1B_2 \otimes C_1C_2\rangle| \leq \sum_i p_i \sqrt{\langle A_1A_1^\dagger\rangle_i \langle A_2^\dagger A_2\rangle_i \langle B_1B_1^\dagger\rangle_i \langle B_2^\dagger B_2\rangle_i \langle C_1C_1^\dagger\rangle_i \langle C_2^\dagger C_2\rangle_i}$$

Since each of the terms contains operators of the form $XX^\dagger$ and is thus a positive operator, we can safely set $\langle XX^\dagger\rangle = \left(\langle XX^\dagger\rangle^n\right)^{\frac{1}{n}}$. Setting n = 2, we find:

$$|\langle A_1A_2 \otimes B_1B_2 \otimes C_1C_2\rangle| \leq \sum_i p_i \Big(\langle A_1A_1^\dagger\rangle_i \langle A_1A_1^\dagger\rangle_i \langle A_2^\dagger A_2\rangle_i \langle A_2^\dagger A_2\rangle_i \langle B_1B_1^\dagger\rangle_i \langle B_1B_1^\dagger\rangle_i$$
$$\dots \langle B_2^\dagger B_2\rangle_i \langle B_2^\dagger B_2\rangle_i \langle C_1C_1^\dagger\rangle_i \langle C_1C_1^\dagger\rangle_i \langle C_2^\dagger C_2\rangle_i \langle C_2^\dagger C_2\rangle_i\Big)^{\frac{1}{4}}$$

We now combine these expected values in groups of **three** in order to find:

$$|\langle A_1A_2 \otimes B_1B_2 \otimes C_1C_2\rangle| \leq \sum_i p_i \Big(\langle A_1A_1^\dagger \otimes B_1B_1^\dagger \otimes C_2^\dagger C_2\rangle_i \langle A_1A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_1C_1^\dagger\rangle_i$$
$$\dots \langle A_2^\dagger A_2 \otimes B_1B_1^\dagger \otimes C_1C_1^\dagger\rangle_i \langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle_i\Big)^{\frac{1}{4}}$$

We now proceed in the same manner as before and apply a combination of Theorem 6.5 and Theorem 6.6 in order to find:

$$|\langle A_1A_2 \otimes B_1B_2 \otimes C_1C_2\rangle| \leq \left(\left(\sum_i p_i \langle A_1A_1^\dagger \otimes B_1B_1^\dagger \otimes C_2^\dagger C_2\rangle_i\right)\left(\sum_i p_i \langle A_1A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_1C_1^\dagger\rangle_i\right)\right.$$
$$\left.\dots \left(\sum_i p_i \langle A_2^\dagger A_2 \otimes B_1B_1^\dagger \otimes C_1C_1^\dagger\rangle_i\right)\left(\sum_i p_i \langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle_i\right)\right)^{\frac{1}{4}}$$

Each of the sums on the right can be rewritten to the expected value with regards to the density matrix at hand. This finishes the proof. □

Similarly, one can prove the following theorem:

Theorem 10.5. *Let A_1, A_2, B_1, B_2 and C_1, C_2 be operators acting on the Hilbert Spaces $\mathcal{H}_A, \mathcal{H}_B$ and $\mathcal{H}_C$, respectively. Then the following condition holds for all separable states:*

$$|\langle A_1A_2 \otimes B_1B_2 \otimes C_1C_2\rangle| \leq \Big(\langle A_1A_1^\dagger \otimes B_1B_1^\dagger \otimes C_1C_1^\dagger\rangle \langle A_1A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle$$
$$\dots \langle A_2^\dagger A_2 \otimes B_1B_1^\dagger \otimes C_2^\dagger C_2\rangle \langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \otimes C_1C_1^\dagger\rangle\Big)^{\frac{1}{4}}$$

The next two extensions can be found below. As these can be proved in a completely analogous manner as the previous criteria, we omit their proof - one can follow the previous proof and take $n = 3$ instead and regroup the operators differently.

Theorem 10.6. *Let A_1, A_2, B_1, B_2 and C_1, C_2 be operators acting on the Hilbert Spaces $\mathcal{H}_A, \mathcal{H}_B$ and $\mathcal{H}_C$, respectively. Then for separable states, we have that:*

$$|\langle A_1 A_2 \otimes B_1 B_2 \otimes C_1 C_2\rangle| \leq \Big(\langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \otimes C_1 C_1^\dagger\rangle \langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger \otimes C_2^\dagger C_2\rangle \langle A_1 A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_1 C_1^\dagger\rangle \\ \dots \langle A_1 A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle \langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger \otimes C_1 C_1^\dagger\rangle \langle A_1 A_1^\dagger \otimes B_1 B_1^\dagger \otimes C_2^\dagger C_2\rangle\Big)^{\frac{1}{6}}$$

Theorem 10.7. *Let A_1, A_2, B_1, B_2 and C_1, C_2 be operators acting on the Hilbert Spaces $\mathcal{H}_A, \mathcal{H}_B$ and $\mathcal{H}_C$, respectively. Then for separable states, we have that:*

$$|\langle A_1 A_2 \otimes B_1 B_2 \otimes C_1 C_2\rangle| \leq \Big(\langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \otimes C_1 C_1^\dagger\rangle \langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger \otimes C_2^\dagger C_2\rangle \langle A_1 A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle \\ \dots \langle A_1 A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle \langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger \otimes C_1 C_1^\dagger\rangle \langle A_1 A_1^\dagger \otimes B_1 B_1^\dagger \otimes C_1 C_1^\dagger\rangle\Big)^{\frac{1}{6}}$$

11

Assessment of the Inequalities-based Entanglement Criteria

In this section, we assess the entanglement criteria that have been derived using the different inequalities by comparing them to some of the criteria mentioned in the Chapter 9. Furthermore, we the optimality condition as well as the limitations of these criteria. This section will again be split up into two parts: the bipartite case and the multipartite case.

11.1. Assessing the Bipartite Criterion

In this section, we will assess the criteria at hand in two ways. First, we will derive the optimal choice of criteria for a given bipartite state, after which we will compare the criteria at hand to various of the criteria that have been discussed in Chapter 9.

11.1.1. Determining an optimal criterion

Let us now study the optimal choice of operators. Note that the optimal choice corresponds to the choice of operators which gives the tightest upper bound, as this ensures that we can detect as many states as possible. This optimal choice will prove to be rather elegant.

Theorem 11.1. *The optimal choice for operators in Theorem 10.2 are of the form:*

$$A_1 = |\alpha\rangle\langle\phi| \qquad A_2 = |\phi\rangle\langle\gamma|$$

where α, γ and ϕ are pure states. An analogous result holds for the operators acting on $\mathcal{H}_B$.

Proof. First, note that given an orthonormal basis $\{\phi_j\}$ of $\mathcal{H}_A$, any operators A_1, A_2 acting on $\mathcal{H}_A$ can be written as follows:

$$A_1 = \sum_j a_j |\alpha_j\rangle\langle\phi_j| \qquad A_2 = \sum_j g_j |\phi_j\rangle\langle\gamma_j|$$

Note that Theorem 10.2 can now be written as:

$$|\langle\left(\sum_j a_j |\alpha_j\rangle\langle\phi_j| \sum_j g_j |\phi_j\rangle\langle\gamma_j|\right) \otimes B_1 B_2\rangle|^2 \leq \langle\left(\sum_j a_j |\alpha_j\rangle\langle\phi_j| \sum_j a_j^* |\phi_j\rangle\langle\alpha_j|\right) \otimes B_2^\dagger B_2\rangle \cdot$$
$$\langle\left(\sum_j g_j^* |\gamma_j\rangle\langle\phi_j| \sum_j g_j |\phi_j\rangle\langle\gamma_j|\right) \otimes B_2^\dagger B_2\rangle$$

which, using the orthonormality of $\{\phi_j\}$ and the linearity of the expectation value operator, can be rewritten as:

$$|\sum_j a_j g_j \langle |\alpha_j\rangle\langle\gamma_j| \otimes B_1 B_2\rangle|^2 \leq \sum_j |a_j|^2 \langle |\alpha_j\rangle\langle\alpha_j| \otimes B_2^\dagger B_2\rangle \sum_i |g_i|^2 \langle |\gamma_i\rangle\langle\gamma_i| \otimes B_2^\dagger B_2\rangle$$

This will be our starting point. We will prove that this upper bound has some slack - meaning, that there exists a more tight upper bound, and that this slack is only zero under the choice of optimal operators proposed in this theorem.
We start off by introducing more compact notation. Set:

$$x_j \equiv \langle |\alpha_j\rangle\langle\alpha_j| \otimes B_2^\dagger B_2\rangle$$
$$y_i \equiv \langle |\gamma_i\rangle\langle\gamma_i| \otimes B_2^\dagger B_2\rangle$$

We can then rewrite the right-hand side of our criterion as follows:

$$\sum_j |a_j|^2 x_j \sum_i |\gamma_i|^2 y_i = \sum_j |\alpha_j\gamma_j|^2 x_j y_i + \sum_i \sum_{k>i} \left(|\alpha_i\gamma_k|^2 x_i y_k + |\alpha_k\gamma_i|^2 x_k y_i\right)$$

Let us now consider the left hand side. We will now derive a more tight upper bound on the left hand side, and will then show that this upper bound is equivalent to the right hand side of our criterion if and only if we choose the proposed optimal operators. We can apply the triangle inequality on the right hand side, and then apply Theorem 10.2 on each of the operators in the summand to find:

$$|\sum_j a_j g_j \langle |\alpha_j\rangle\langle\gamma_j| \otimes B_1 B_2\rangle|^2 \leq \left(\sum_j |\alpha_j\gamma_j| \sqrt{x_j y_j}\right)^2$$

in which we have equality if we chose the operators as proposed. □

Note that under the optimal choice of operators, our criterion reduces to:

$$|\langle\alpha_1 \otimes \beta_1|\rho|\alpha_2 \otimes \beta_2\rangle|^2 \leq \langle\alpha_1 \otimes \beta_1|\rho|\alpha_1 \otimes \beta_1\rangle \cdot \langle\alpha_2 \otimes \beta_2|\rho|\alpha_2\beta_2\rangle$$

11.1.2. Comparing the Bipartite Criterion

Comparison with the Criterion by Hillery and Zubairy

We start off by considering the entanglement criteria that have been derived by **Hillery and Zubairy** (Theorem 9.18). Note that the first of these criteria is a direct consequence of our bipartite inequality. This can easily be seen by setting:

$$A_1 = A^\dagger \qquad A_2 = I_A$$
$$B_1 = I_B \qquad B_2 = B$$

As such, this criterion is a *direct* implication of the criterion we derived based on the Cauchy-Schwarz inequality. Note that we can therefore consider our criterion to be a generalisation of the criterion of Hillery and Zubairy!
Furthermore, note that the second of these criteria has the shape of Theorem 10.1, and as such, a full comparison with our criterion is not possible.

Nevertheless, let us consider the example of a qubit system, and let us show that our criterion can be employed to yield a more tight entanglement detection criterion than both of the criteria from Hillery and Zubairy. Consider, for example, the following operators:

$$A_1 = |1\rangle\langle 0| \qquad A_2 = |0\rangle\langle 0|$$
$$B_1 = |1\rangle\langle 0| \qquad B_2 = |0\rangle\langle 0|$$

Let us first consider what we find for the first criterion of Theorem 9.18. *Note that these results were presented incorrectly in Wölk et al. (2014).* We present the correct results below. The first criterion reduces to:

$$|\langle A_1 \otimes B_1\rangle|^2 \leq \langle A_1 A_1^\dagger \otimes B_1^\dagger B_1\rangle$$

The left hand side yields:

$$|\langle A_1 \otimes B_1\rangle|^2 = |\langle |1\rangle\langle 0| \otimes |1\rangle\langle 0|\rangle|^2 = |\langle |11\rangle\langle 00|\rangle|^2 = \rho_{00,11}$$

Similarly, the right hand side yields:

$$\langle A_1 A_1^\dagger \otimes B_1^\dagger B_1\rangle = \langle |1\rangle\langle 0|0\rangle\langle 1| \otimes |0\rangle\langle 1|1\rangle\langle 0|\rangle = \langle |10\rangle\langle 10|\rangle = \rho_{10,10}$$

From which we finally find the following criterion:

$$|\rho_{00,11}|^2 \leq \rho_{10,10}$$

In an exactly analogous manner, we find that our criterion becomes:

$$|\rho_{00,11}|^2 \leq \rho_{01,01}\rho_{10,10}$$

Furthermore, we find that the second criterion yields:

$$|\rho_{00,11}|^2 \leq \left(\rho_{00,00} + \rho_{01,01}\right)\left(\rho_{00,00} + \rho_{10,10}\right) = \rho_{10,10}\rho_{01,01} + \rho_{00,00}^2 + \rho_{00,00}\left(\rho_{01,01} + \rho_{10,10}\right)$$

where we took $A = A_1$ instead, in order to find the same left hand side. *Note that these results were also presented incorrectly in Wölk et al. (2014).*
However, since density operators are Hermitian, positive operators with a trace of unity, we have that $0 \leq \rho_{ij,ij} \leq 1$, and as such, we find that our criterion outperforms the criteria of Hillary and Zubairy in almost every two-qubit case.

COMPARISON WITH THE CRITERION BY SHCHUKIN AND VOGEL

Note that by choosing the following operators:

$$A_1 = \left(a^\dagger\right)^m a^n \qquad A_2 = \left(a^\dagger\right)^p a^q$$
$$B_1 = \left(b^\dagger\right)^s a^r \qquad B_2 = \left(b^\dagger\right)^k a^l$$

and substituting this in our criterion in Theorem 10.2, we find the criterion by Shchuckin and Vogel (Theorem 9.19) - thus this criterion is merely a special case of our criterion!

COMPARISON WITH THE CRITERION BY HUBER ET AL.

We now compare our criterion with the criterion by **Huber et al.** (Theorem 9.20). Note that if we take a product state $|\Phi\rangle = |\alpha\beta\gamma\delta\rangle$, the right hand side yields:

$$\sqrt{\langle\Phi|\rho\otimes\rho|\Phi\rangle} = \sqrt{\langle\alpha\beta|\rho|\alpha\beta\rangle\langle\gamma\delta|\rho|\gamma\delta\rangle}$$

whereas the left hand side becomes:

$$\begin{aligned}\sqrt{\mathrm{Re}\left(\langle\alpha\beta\gamma\delta\right|(\mathbb{1}\otimes\Pi)^{\dagger}\rho\otimes\rho\,(\Pi\otimes\mathbb{1})\left|\alpha\beta\gamma\delta\rangle\right)} &= \sqrt{\mathrm{Re}\left(\langle\alpha\beta\gamma\delta\,(\mathbb{1}\otimes\Pi)\right|\rho\otimes\rho\left|(\Pi\otimes\mathbb{1})\,\alpha\beta\gamma\delta\rangle\right)}\\ &= \sqrt{\mathrm{Re}\left(\langle\alpha\delta\gamma\beta\right|\rho\otimes\rho\left|\gamma\beta\alpha\delta\rangle\right)}\\ &= \sqrt{\mathrm{Re}\left(\langle\alpha\delta|\,\rho\left|\gamma\beta\rangle\,\rho\,\langle\gamma\beta\right|\alpha\delta\rangle\right)}\\ &= |\langle\alpha\delta|\,\rho\left|\gamma\delta\rangle\right|\end{aligned}$$

We combine these to rewrite the criterion as:

$$|\langle\alpha\delta|\,\rho\left|\gamma\delta\rangle\right| \leq \sqrt{\langle\alpha\beta|\,\rho\left|\alpha\beta\rangle\langle\gamma\delta\right|\rho\left|\gamma\delta\rangle\right.}$$

which is just the optimal form of our bipartite criterion!

COMPARISON WITH THE PPT CRITERION

Note that, first of all, that this criterion is only limited to NPT states.

Theorem 11.2. *The criterion in Theorem 10.2 does not detect PPT states.*

Proof. Suppose ρ is PPT and violates the upper bound. First note that:

$$|\langle A_1A_2\otimes B_1B\rangle| = |\mathrm{Tr}\left((A_1A_2\otimes B_1B_2)\,\rho\right)| = |\mathrm{Tr}\left(\left((A_1A_2)^T\otimes B_1B_2\right)\rho^{T_A}\right)| = |\mathrm{Tr}\left(\left(A_2^TA_1^T\otimes B_1B_2\right)\rho^{T_A}\right)|$$

Note that the the trace is conserved by the partial transpose operator.
Now, since ρ^{T_A} is PPT and thus a valid density operator, the Cauchy-Schwarz inequality (Theorem 8.7) can be used to find that:

$$|\mathrm{Tr}\left(\left(A_2^TA_1^T\otimes B_1B_2\right)\rho^{T_A}\right)| \leq \sqrt{\mathrm{Tr}\left(\left(A_2^T\left(A_2^T\right)^{\dagger}\otimes B_1B_1^{\dagger}\right)\rho^{T_A}\right)\mathrm{Tr}\left(\left(\left(A_1^T\right)^{\dagger}A_1^T\otimes B_2^{\dagger}B_2\right)\rho^{T_A}\right)}$$

Thus:

$$\begin{aligned}|\langle A_1A_2\otimes B_1B\rangle| &\leq \sqrt{\mathrm{Tr}\left(\left(A_2^T\left(A_2^T\right)^{\dagger}\otimes B_1B_1^{\dagger}\right)\rho^{T_A}\right)\mathrm{Tr}\left(\left(\left(A_1^T\right)^{\dagger}A_1^T\otimes B_2^{\dagger}B_2\right)\rho^{T_A}\right)}\\ &= \sqrt{\langle A_1A_1^{\dagger}\otimes B_1B_1^{\dagger}\rangle\langle A_2^{\dagger}A_2\otimes B_2^{\dagger}B_2\rangle}\end{aligned}$$

which yields the upper bound from Theorem 10.2. $\Rightarrow\Leftarrow$ □

Therefore, our criterion can only be used to determine entanglement in NPT states. One bipartite PPT state that is entangled, and can therefore not be detected by our criterion, is the following double qutrit state:

$$\rho_a = \frac{1}{8a+1}\begin{bmatrix} a & 0 & 0 & 0 & a & 0 & 0 & 0 & a\\ 0 & a & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & a & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & a & 0 & 0 & 0 & 0 & 0\\ a & 0 & 0 & 0 & a & 0 & 0 & 0 & a\\ 0 & 0 & 0 & 0 & 0 & a & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1+a}{2} & 0 & \frac{\sqrt{1-a^2}}{2}\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a & 0\\ a & 0 & 0 & 0 & a & 0 & \frac{\sqrt{1-a^2}}{2} & 0 & \frac{1+a}{2} \end{bmatrix} \qquad (0<a<1)$$

Note, however, that the PPT criterion yields that NPT states are *always* entangled. As such, our criterion is weaker than the PPT criterion in the most general case, as there are NPT states that cannot be detected by our criterion. In the next section we will see, however, that our criterion does perform sufficiently well in the case of the most important bipartite state: two-qubit states.

11.1.3. USEFUL APPLICATIONS OF THE BIPARTITE CRITERION

Let us now consider the extent to which this criterion is useful. An important application of this criterion lies in its ability to detect entanglement in two-qubit systems. Before we can prove this, we prove a relevant lemma.

Lemma 11.3. *Let ρ be a density matrix representing a bipartite state. Then ρ violates the criterion given in Theorem 10.2 under a few conditions. First, ρ^{T_A} needs to have eigenstates $\left|\lambda_j^+\right\rangle$ corresponding to positive eigenvalues and eigenstates $\left|\lambda_j^-\right\rangle$ corresponding to negative eigenvalues and $\mathscr{H}_A$ (respectively $\mathscr{H}_B$) has states $|a_k\rangle$ ($|b_k\rangle$) with $k = 1,2$ such that:*

1. $\left\langle a_2 \otimes b_1 \middle| \lambda_j^+ \right\rangle = c^+ \left\langle a_1 \otimes b_2 \middle| \lambda_j^+ \right\rangle$
2. $\left\langle a_2 \otimes b_1 \middle| \lambda_j^- \right\rangle = c^- \left\langle a_1 \otimes b_2 \middle| \lambda_j^- \right\rangle$

where $c^\pm$ are not dependent on j and $c^+ c^- < 0$. Then, for arbitrary $\alpha \in \mathscr{H}_B$ and $\beta \in \mathscr{H}_B$, set:

$$A_1 = \left|a_1^*\right\rangle\langle\alpha| \qquad B_1 = |b_1\rangle\langle\beta| \tag{11.1}$$

$$A_2 = |\alpha\rangle\left\langle a_2^*\right| \qquad B_2 = |\beta\rangle\langle b_2| \tag{11.2}$$

Proof. We make use of the proof of Theorem 11.2 to show that:

$$|\mathrm{Tr}\left(\left(A_2^T A_1^T \otimes B_1 B_2\right)\rho^{T_A}\right)| \geq \sqrt{\mathrm{Tr}\left(\left(A_2^T \left(A_2^T\right)^\dagger \otimes B_1 B_1^\dagger\right)\rho^{T_A}\right)\mathrm{Tr}\left(\left(\left(A_1^T\right)^\dagger A_1^T \otimes B_2^\dagger B_2\right)\rho^{T_A}\right)}$$

We consider both sides of the equation separately. We first fill in the left-hand sight of the equation.

$$\begin{aligned}
|\mathrm{Tr}\left(\left(A_2^T A_1^T \otimes B_1 B_2\right)\rho^{T_A}\right)| &= |\mathrm{Tr}\left(\left(\left(|a_2\rangle\left\langle\alpha^*\right|\right)\left(\left|\alpha^*\right\rangle\langle a_1|\right)\right)\otimes\left(\left(|b_1\rangle\langle\beta|\right)\left(|\beta\rangle\langle b_2|\right)\right)\rho^{T_A}\right)| \\
&= |\mathrm{Tr}\left(\left(|a_2\rangle\left(\left\langle\alpha^*\middle|\alpha^*\right\rangle\right)\langle a_1|\right)\otimes\left(|b_1\rangle\left(\langle\beta|\beta\rangle\right)\langle b_2|\right)\rho^{T_A}\right)| \\
&= |\mathrm{Tr}\left(\left(\left(|a_2\rangle\langle a_1|\right)\otimes\left(|b_1\rangle\langle b_2|\right)\right)\rho^{T_A}\right)| \\
&= |\mathrm{Tr}\left(\left(|a_2\otimes b_1\rangle\langle a_1\otimes b_2|\right)\rho^{T_A}\right)| \\
&= |\langle a_1\otimes b_2|\,\rho^{T_A}\,|a_2\otimes b_1\rangle|
\end{aligned}$$

We now introduce a spectral decomposition of ρ^{T_A}, i.e. $\rho^{T_A} = \sum_j \lambda_j^\pm \left|\lambda_j^\pm\right\rangle\left\langle\lambda_j^\pm\right|$, from which we find:

$$\begin{aligned}
|\mathrm{Tr}\left(\left(A_2^T A_1^T \otimes B_1 B_2\right)\rho^{T_A}\right)| &= |\langle a_1\otimes b_2|\left(\sum_j \lambda_j^\pm\left|\lambda_j^\pm\right\rangle\right)\left\langle\lambda_j^\pm\middle|a_2\otimes b_1\right\rangle| \\
&= |\sum_j \lambda_j^\pm\left(\langle a_1\otimes b_2|\left|\lambda_j^\pm\right\rangle\right)\left(\left\langle\lambda_j^\pm\middle|a_2\otimes b_1\right\rangle\right)| \\
&= |\sum_j\left(|c^\pm|\lambda_j^\pm\left\langle a_1\otimes b_2\middle|\lambda_j^\pm\right\rangle\right)| \\
&= |\sum_j |c^+|\lambda_j^+\left\langle a_1\otimes b_2\middle|\lambda_j^+\right\rangle + \sum_j |c^-|\lambda_j^-\left\langle a_1\otimes b_2\middle|\lambda_j^-\right\rangle|
\end{aligned}$$

We now consider the right hand side of the equation. We work out the results of one of the two sides of the product under the square root, since the derivation of the other goes completely analogously.

$$\begin{aligned}\mathrm{Tr}\left(\left(A_2^T (A_2^T)^\dagger \otimes B_1 B_1^\dagger\right)\rho^{T_A}\right) &= \mathrm{Tr}\left(\left((|a_2\rangle\left\langle\alpha^*\middle|\alpha^*\right\rangle\langle a_2|)\otimes|b_1\rangle\left\langle\beta\middle|\beta\right\rangle\langle b_1|\right)\rho^{T_A}\right)\\ &= \mathrm{Tr}\left(((|a_2\rangle\langle a_2|)\otimes|b_1\rangle\langle b_1|)\,\rho^{T_A}\right)\\ &= \mathrm{Tr}\left((|a_2\otimes b_1\rangle\langle a_2\otimes b_1|)\rho^{T_A}\right)\\ &= \langle a_2\otimes b_1|\,\rho^{T_A}\,|a_2\otimes b_1\rangle\end{aligned}$$

We now use the spectral decomposition of ρ^{T_A} to find:

$$\mathrm{Tr}\left(\left(A_2^T (A_2^T)^\dagger \otimes B_1 B_1^\dagger\right)\rho^{T_A}\right) = \sum_j \lambda_j^\pm \left|\left\langle a_2\otimes b_1\middle|\lambda_j^\pm\right\rangle\right|^2 = \sum_j |c^\pm|^2 \lambda_j^\pm \left|\left\langle a_1\otimes b_2\middle|\lambda_j^\pm\right\rangle\right|^2$$

Proceeding in a similar manner, we find that the right hand side equals:

$$\sqrt{\left(|c^+|^2\sum_j\lambda_j^+\left|\left\langle a_1\otimes b_2\middle|\lambda_j^+\right\rangle\right|^2+|c^-|^2\sum_j\lambda_j^-\left|\left\langle a_1\otimes b_2\middle|\lambda_j^-\right\rangle\right|^2\right)\left(\sum_j\lambda_j^+\left|\left\langle a_1\otimes b_2\middle|\lambda_j^+\right\rangle\right|^2+\sum_j\lambda_j^-\left|\left\langle a_1\otimes b_2\middle|\lambda_j^-\right\rangle\right|^2\right)}$$

This can be worked out to:

$$\sqrt{|c^+|^2\left(\sum_j\left|\left\langle a_1\otimes b_2\middle|\lambda_j^+\right\rangle\right|^2\right)^2+|c^-|^2\left(\sum_j\left|\left\langle a_1\otimes b_2\middle|\lambda_j^-\right\rangle\right|^2\right)^2-\left(|c^+|^2+|c^-|^2\right)\sum_j\lambda_j^+\left|\left\langle a_1\otimes b_2\middle|\lambda_j^+\right\rangle\right|^2\sum_j|\lambda_j^-|\left|\left\langle a_1\otimes b_2\middle|\lambda_j^-\right\rangle\right|^2}$$

Now, using the starting point of this thesis, namely: $\left(|c^+|-|c^-|\right)^2\geq 0$, we can show that the right hand side is bound as follows:

$$\begin{aligned}&\leq \left||c^+|\sum_j\lambda_j^+\left|\left\langle a_1\otimes b_2\middle|\lambda_j^+\right\rangle\right|^2-|c^-|\sum_j|\lambda_j|\left|\left\langle a_1\otimes b_2\middle|\lambda_j^-\right\rangle\right|^2\right|\\ &\leq |c^+|\sum_j\lambda_j^+\left|\left\langle a_1\otimes b_2\middle|\lambda_j^+\right\rangle\right|^2+|c^-|\sum_j|\lambda_j|\left|\left\langle a_1\otimes b_2\middle|\lambda_j^-\right\rangle\right|^2\end{aligned}$$

which shows that the right hand side is indeed smaller than the left hand side. □

Theorem 11.4. *Any two-qubit state ρ can be detected using Theorem 10.2.*

We proceed in a different manner than was done in Wölk et al. (2014) and first introduce a lemma to show that an entangled 2-qubit state's partial transpose has one negative and three positive eigenvalues.

Lemma 11.5. *The partial transpose of the density matrix of any entangled 2-qubit state has one negative eigenvalue, say λ^-, and three positive eigenvalues.*

Proof. Let $|\Psi\rangle$ represent any arbitrary entangled two-qubit state. Using the Schmidt decomposition, we can then express $|\Psi\rangle$ in terms of a two-dimensional, orthogonal basis, say $\{|\chi_0\chi_0\rangle, |\chi_1\chi_1\rangle\}$, such that:

$$|\Psi\rangle = \chi_0\left|\chi_0\chi_0\right\rangle+\chi_1\left|\chi_1\chi_1\right\rangle,\quad \chi_0,\chi_1\in\mathbb{R}_{>0}$$

with $\chi_0^2+\chi_1^2=1$. Then:

$$\rho = |\Psi\rangle\langle\Psi| = \chi_0^2 |\chi_0\chi_0\rangle\langle\chi_0\chi_0| + \chi_0\chi_1 |\chi_0\chi_0\rangle\langle\chi_1\chi_1| + \chi_1\chi_0 |\chi_1\chi_1\rangle\langle\chi_0\chi_0| + \chi_1^2 |\chi_1\chi_1\rangle\langle\chi_1\chi_1|$$

$$= \begin{bmatrix} \chi_0^2 & 0 & 0 & \chi_0\chi_1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \chi_1\chi_0 & 0 & 0 & \chi_1^2 \end{bmatrix}$$

We can then express ρ^{T_A} as follows:

$$\rho^{T_A} = \begin{bmatrix} \chi_0^2 & 0 & 0 & 0 \\ 0 & 0 & \chi_0\chi_1 & 0 \\ 0 & \chi_1\chi_0 & 0 & 0 \\ 0 & 0 & 0 & \chi_1^2 \end{bmatrix}$$

Which has the following eigenvectors (belonging to eigenvalue χ_0^2, χ_1^2, $\chi_0\chi_1$ and $-\chi_0\chi_1$, respectively):

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

This finishes the proof. □

We now prove that entangled two-qubit systems can always be detected by our criteria.

Proof. Using the lemma we just proved, we can express the negative eigenvalue in its own Schmidt basis, say:

$$|\lambda^-\rangle = \eta_0 |\eta_0\eta_0\rangle + \eta_1 |\eta_1\eta_1\rangle$$

Since all positive eigenvectors need to be orthogonal to the negative eigenvector, they must necessarily be of the form:

$$|\lambda_k^+\rangle = \gamma_k\eta_0 |\eta_0\eta_0\rangle - \gamma_k\eta_1 |\eta_1\eta_1\rangle + \phi_k |\eta_0\eta_1\rangle + \delta_k |\eta_1\eta_0\rangle$$

for arbitrary γ_k, δ_k and ϕ_k which cannot all be zero for the same value of k and with $\gamma_k \neq 0$ for at least one k (since these eigenvalues need to be orthogonal as well). We can now use 11.3 by setting:

$$|a_1 \otimes b_2\rangle = |\eta_1\eta_1\rangle$$
$$|a_2 \otimes b_1\rangle = |\eta_0\eta_0\rangle$$

with $c^+ = -\frac{\eta_1}{\eta_0}$ and $c^- = \frac{\eta_0}{\eta_1}$. □

11.1.4. LIMITATIONS TO THE BIPARTITE CRITERION

First of all, note that since our criterion cannot be used to detect entanglement in PPT states, we are restricted to detecting entanglement in NPT states. The question arises, however, whether all entangled NPT states can be detected. Before tackling this question, let us first introduce two definitions.

Definition 11.1. A state ρ is called *distillable* if for some $N \in \mathbb{N}$, $\rho^{\otimes N}$ can be brought to a maximally entangled state (e.g. $|\Psi_-\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}$) solely through the use of local operations and classical communication (LOCC).

Remark. If a state ρ is distillable, and say $\rho^{\otimes n} \xrightarrow{\text{LOCC}} \Psi_-$ for some smallest possible n, then ρ is called n-distillable.

Definition 11.2. If ρ is entangled and distillable, ρ is called *freely entangled*. Otherwise, ρ is called *bound entangled*.

Now that we have introduced these definitions, let us address the core issue in this section. First of all, note the optimality condition we found that the criterion gives us four independent vectors:

$$|\langle \alpha_1 \otimes \beta_1 | \rho | \alpha_2 \otimes \beta_2 \rangle|^2 \leq \langle \alpha_1 \otimes \beta_1 | \rho | \alpha_1 \otimes \beta_1 \rangle \cdot \langle \alpha_2 \otimes \beta_2 | \rho | \alpha_2 \beta_2 \rangle$$

Note, however, that $|\alpha_1\rangle$ can be decomposed into a part in $|\alpha_2\rangle$ and $|\alpha_2^\perp\rangle$. As such, the criterion yields an invariant result under projection in the qubit subspace $\mathbb{1}_A = |\alpha_1\rangle\langle\alpha_1| + |\alpha_1^\perp\rangle\langle\alpha_1^\perp|$. An analogous argument holds for the subsystem $\mathscr{H}_B$. Thus, we see that detecting entanglement using this criterion is equivalent to detecting entanglement a subsystem isomorphic to $\mathbb{C}_1^2 \otimes \mathbb{C}_1^2$. But from Horodecki et al. (1998) it therefore follows that detection of entanglement in such systems necessarily implies 1-distillability. As such, any bound entangled state or state that is freely entangled but not 1-distillable cannot be detected by our criterion!

We now discuss two examples of such states. First of all, let us consider the Werner states (Werner (1989)). Most generally, they are characterised by the parameter $p \in [0,1]$ and the dimension d, such that:

$$\rho_W(p,d) = p\frac{P^-}{N^-} + (1-p)\frac{P^+}{N^+}$$

where $N^\pm = \frac{d(d\pm1)}{2}$ and $P^\pm = \frac{I\pm F}{2}$, where $F \equiv \sum_{i,j} = |i\rangle\langle j| \otimes |j\rangle\langle i|$ is the so-called flip operator.
For $p \leq \frac{1}{2}$, these states are PPT. Furthermore, for $p > \frac{3(d-1)}{2(2d-1)}$, they are distillable. while for $\frac{1}{2} \leq p \leq \frac{3(d-1)}{2(2d-1)}$, they are conjectured to be bound entangled - and surely not 1-distillable (see Horodecki (2001)).
So, if we consider the case of a qutrit pair, we then find:

$$\rho_W = \frac{p}{6}\begin{bmatrix} 0&0&0&0&0&0&0&0&0\\ 0&1&0&-1&0&0&0&0&0\\ 0&0&1&0&0&0&-1&0&0\\ 0&-1&0&1&0&0&0&0&0\\ 0&0&0&0&0&0&0&0&0\\ 0&0&0&0&0&1&0&-1&0\\ 0&0&-1&0&0&0&1&0&0\\ 0&0&0&0&0&-1&0&1&0\\ 0&0&0&0&0&0&0&0&0 \end{bmatrix} + \frac{1-p}{12}\begin{bmatrix} 2&0&0&0&0&0&0&0&0\\ 0&1&0&1&0&0&0&0&0\\ 0&0&1&0&0&0&1&0&0\\ 0&1&0&1&0&0&0&0&0\\ 0&0&0&0&2&0&0&0&0\\ 0&0&0&0&0&1&0&1&0\\ 0&0&1&0&0&0&1&0&0\\ 0&0&0&0&0&1&0&1&0\\ 0&0&0&0&0&0&0&0&2 \end{bmatrix}$$

A bound entangled variant can be found by taking $\frac{1}{2} \leq p \leq \frac{3}{5}$. Setting $p = \frac{3}{5}$ yields:

$$\rho_W = \frac{1}{30}\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 0 & 0 & 0 & -2 & 0 & 0 \\ 0 & -2 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & -2 & 0 \\ 0 & 0 & -2 & 0 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}$$

which is an NPT state that is undetectable using our criterion.

A second class of not 1-distillable states is based on the results derived in Watrous (2004). There, the following Lemma is proved.

Lemma 11.6. *(Watrous (2004)) For any integers $d \geq 3$ and $n \geq 1$, there is some $\epsilon > 0$ such that the following state is not n-distillable:*

$$\rho_\epsilon = \frac{d+1+\epsilon}{d-1}(P^-)^{\otimes 2} + (P^+)^{\otimes 2}$$

Specifically for the case that $n = 1$, by following the proofs in Watrous (2004), one can derive that for $\epsilon < \frac{2d(1-\frac{2}{d})^2}{4(d-1)-(1-\frac{2}{d})^2}$, $\rho(\epsilon)$ is not 1-distillable. Working this out for $d = 3$, we find:

$$\epsilon < \frac{6}{71}$$

and ρ_ϵ becomes:

$$\rho_\epsilon = \frac{4+\epsilon}{2}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{\otimes 2} + \frac{1}{2}\begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}^{\otimes 2}$$

which yields a 81 x 81 matrix that represents a undetectable 9 x 9 state for any $\epsilon < \frac{6}{71}$.

11.2. ASSESSING THE MULTIPARTITE CRITERIA

11.2.1. APPLICATIONS OF ONE MULTIPARTITE CRITERION

Let us consider the criteria we obtained in Theorem 10.4 and Theorem 10.5 for one specific case of Let us consider the following state, which was introduced by A. Kay (Gühne (2010)) and which is separable under any bipartition, but not fully separable:

$$\rho_{abc} = \frac{1}{2+a+b+c+\frac{1}{a}+\frac{1}{b}+\frac{1}{c}} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & a & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & b & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{c} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & c & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{b} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{c} & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

for $a, b, c > 0$ and $abc \neq 1$. under the standard basis $\{|000\rangle, |001\rangle, \ldots, |111\rangle\}$. We now set:

$$A_1 = B_1 = C_1 = |1\rangle\langle 0|$$
$$A_2 = B_2 = C_2 = |0\rangle\langle 0|$$

We derive the result for Theorem 10.4 and posit the result for Theorem 10.5, as this is derived in an exactly analogous manner.
Substituting the operators on the left hand side of equation gives:

$$|\langle A_1 A_2 \otimes B_1 B_2 \otimes C_1 C_2\rangle| = |\langle (|1\rangle\langle 0|)^{\otimes 3}\rangle| = \langle 111|\rho|000\rangle = 1$$

Similarly, we can retrieve the right hand side:

$$\left(\langle A_1 A_1^\dagger \otimes B_1 B_1^\dagger \otimes C_2^\dagger C_2\rangle \langle A_1 A_1^\dagger \otimes B_2^\dagger B_2 \otimes C_1 C_1^\dagger\rangle \langle A_2^\dagger A_2 \otimes B_1 B_1^\dagger \otimes C_1 C_1^\dagger\rangle \langle A_2^\dagger A_2 \otimes B_2^\dagger B_2 \otimes C_2^\dagger C_2\rangle\right)^{\frac{1}{4}}$$
$$= \left(\langle (|1\rangle\langle 1|)^{\otimes 2} \otimes |0\rangle\langle 0|\rangle \langle |1\rangle\langle 1| \otimes |0\rangle\langle 0| \otimes |1\rangle\langle 1|\rangle \langle |0\rangle\langle 0| \otimes (|1\rangle\langle 1|)^{\otimes 2}\rangle \langle (|0\rangle\langle 0|)^{\otimes 3}\rangle\right)^{\frac{1}{4}}$$
$$= \left(\rho_{110,110}\rho_{101,101}\rho_{011,011}\rho_{000,000}\right)^{\frac{1}{4}}$$
$$= \left(\frac{1}{abc}\right)^{\frac{1}{4}} = (abc)^{\frac{-1}{4}}$$

Such that all states with:

$$\frac{1}{abc} > 1$$

are detected by Theorem 10.4. Similarly, we find that all states with:

$$abc > 1$$

are detected by Theorem 10.5. As such, *all* states of this form can be detected by the multipartite criteria we have derived. Furthermore, note that this can actually be done through measurements - even though A_1, B_1 and C_1 are not Hermitian operators and thus not observables, we have that $|1\rangle\langle 0| = \sigma_x - i\sigma_y$.

11.2.2. COMPARING THE MULTIPARTITE CRITERIA

COMPARISON WITH THE CRITERION BY GÜHNE AND SEEVINCK

We now discuss the criterion by **Gühne and Seevinck** (Theorem 9.22) for tripartite states. First, let us prove a lemma, from which this criterion immediately follows by taking a convex sum.

Lemma 11.7. *Let $\rho \in \mathscr{H}_A \times \mathscr{H}_B \otimes \mathscr{H}_C$. Then the following criteria hold for separable states:*

$$|\rho_{000,111}| \leq \sqrt{\rho_{001,001}\rho_{110,110}}$$
$$|\rho_{000,111}| \leq \sqrt{\rho_{010,010}\rho_{101,101}}$$
$$|\rho_{000,111}| \leq \sqrt{\rho_{011,011}\rho_{100,100}}$$

Proof. We only prove one of the equations, as the others follow from a completely analogous argument. Note that Theorem 10.2 still holds if we consider $\mathscr{H}_A = \mathscr{H}_1 \otimes \mathscr{H}_2$. As such, we can apply our bipartite criterion to the following operators:

$$A_1 = |00\rangle\langle 11| \qquad B_1 = |0\rangle\langle 0|$$
$$A_2 = |11\rangle\langle 11| \qquad B_2 = |0\rangle\langle 1|$$

Substituting these in Theorem 10.2, we find:

$$|\langle |00\rangle\langle 11|11\rangle\langle 11| \otimes |0\rangle\langle 0|0\rangle\langle 1|\rangle| \leq \sqrt{\langle |00\rangle\langle 11|11\rangle\langle 00 \otimes |1\rangle\langle 0|0\rangle\langle 1||\rangle \cdot \langle |11\rangle\langle 11|11\rangle\langle 11| \otimes |0\rangle\langle 0|0\rangle\langle 0|\rangle}$$
$$|\langle |000\rangle\langle 111|\rangle| \leq \sqrt{\langle |001\rangle\langle 001|\rangle\langle |110\rangle\langle 110|\rangle}$$
$$|\rho_{000,111}| \leq \sqrt{\rho_{001,001}\rho_{110,110}}$$

□

Note that each subcriterion presented in this lemma is the result of a different partition of the tripartite state. As such, the criterion given by Gühne and Seevinck, which sum these criteria, therefore does not detect biseparable states, and, as such, only detects genuine tripartite entanglement. Note, however, that our criterion is aimed at detecting any form of entanglement, and as such, performs better in for example the application considered in the previous section, since the criterion by Gühne and Seenvinck cannot detect entanglement in ρ_{abc}.

We could, of course, take a different approach than Gühne and Seevinck, and multiply the inequalities found in Lemma 11.7, in order to find:

$$|\rho_{000,111}| \leq \left(\rho_{001,001}\rho_{010,010}\rho_{011,011}\rho_{100,100}\rho_{101,101}\rho_{110,110}\right)^{\frac{1}{6}}$$

Evidently, this is a much stronger criterion than the one presented by them. Note, however, that by turning back to one form of the multipartite criteria that we derived (Theorem 10.6) and setting:

$$A_1 = B_1 = C_1 = |1\rangle\langle 0|$$
$$A_2 = B_2 = C_2 = |0\rangle\langle 0|$$

we find that this criterion is the direct consequence of our more general multipartite criterion!

Lastly, an extension that is proposed in Gühne and Seevinck (2018) is to apply substitutions, for example by setting $\rho_{2,2}\rho_{3,3} \mapsto \rho_{1,1}\rho_{4,4}$. This would yield:

$$|\rho_{000,111}| \leq \sqrt{\rho_{000,000}\rho_{011,011}^2\rho_{100,100}\rho_{101,101}\rho_{110,110}}$$

Note, now that referring back to Theorem 10.7 and by setting:

$$A_1 = B_1 = C_1 = |1\rangle\langle 0|$$
$$A_2 = B_2 = C_2 = |0\rangle\langle 0|$$

we find this inequality! Note that more generally, we have that **any substitution** which preserves the number of 0s and 1s for each index can be proved using a criterion of the form of Theorem 10.6 and 10.7, which are merely specific forms of our more general multipartite criteria!
Furthermore, it should be noted that this criterion can be considered to be a weighed **geometric mean** of the following two equations:

$$|\rho_{000,111}| \leq \sqrt{\rho_{011,011}\rho_{100,100}}$$
$$|\rho_{000,111}| \leq \sqrt[4]{\rho_{000,000}\rho_{011,011}\rho_{101,101}\rho_{110,110}}$$

which are given in Lemma 11.7 and Theorem 10.4 when taking the operators used in the example of ρ_{abc}. The weights, evidently, are $\frac{1}{3}$ and $\frac{2}{3}$, respectively. As such, checking these criteria separately yields a stronger entanglement detection scheme than using the substitution trick. Note that this conclusion extends to all allowed substitutions, and as such, these are not of any particular additional merit[1].

COMPARISON WITH THE CRITERION BY HILLERY ET AL.

First note that the suggested extension of the criterion by **Hillery** and **Zubairy** is also not violated by biseparable states, and as such, does not detect this form of entanglement, while our criterion does.

Now, let us turn to the criteria of **Hillery et al**. Let us start off by noting that the criterion in Theorem 9.24 is a consequence of the multipartite criteria that we have derived! Taking any σ in Theorem 10.3 and setting:

$$A_k^1 = I \qquad A_k^2 = A_k \ \forall k \in \{1,\dots,n\}$$

we find the criterion by Hillery et al. Furthermore, we should note that this criterion is not optimal. In detecting ρ_{abc}, for example, this criteria fails. Even worse, we can prove that this criterion is insensitive to a more general class of biseparable states!

Theorem 11.8. *(Wölk et al. (2014)) Let $\rho \in \bigotimes_{k=1}^{n} \mathcal{H}_k$ denote a density state that is biseparable for every k_1, k_2 over a bipartition $H_1|H_2$ such that $k_1 \in H_1$ and $k_2 \in H_2$. Then ρ cannot be detected by Theorem 9.24.*

Concerning Theorem 9.25, no general comment can be made regarding its performance for detecting entanglement in *general* states in comparison to our criterion (See: Hillery et al. (2010)), however, it should be noted that this criterion sometimes fails to detect entanglement, whereas the other of these criteria - which is a mere consequence of our criterion development scheme - can. Let us consider for example the tensor product of two generalised GHZ-states:

$$|\psi\rangle = (\cos(\theta_1)|0\rangle + \sin(\theta_1)|1\rangle) \otimes \left(\cos(\theta_2)|0\rangle^{\otimes 2} + \sin(\theta_2)|1\rangle^{\otimes 2}\right)$$

Let $A_k = |0\rangle\langle 1|$. We then find, for Theorem 9.24 and 9.25, respectively:

$$|\frac{1}{4}\sin(2\theta_1)\sin(2\theta_2)| \leq \left(\sin(\theta_1)^2\sin(\theta_2)^4\right)^{\frac{1}{3}}$$
$$|\frac{1}{4}\sin(2\theta_1)\sin(2\theta_2)| \leq \left(\frac{2}{3}\right)^{\frac{3}{2}}\cos^2(\theta_1)\sin^2(\theta_2) + \left(\frac{1}{3}\right)^{\frac{3}{2}}\sin^2(\theta_1)\cos^2(\theta_2) + \sin^2(\theta_1)\sin^2(\theta_2)$$

[1] at least, not from a theoretical point of view.

If we assume that the first state is a regular GHZ-state, so $\sin(\theta_1) = \cos(\theta_1) = \frac{1}{\sqrt{2}}$, we find:

$$|\cos(\theta_2)| \leq |4\sin(\theta_2)|^{\frac{1}{3}}$$

$$|\sin(\theta)\cos(\theta)| \leq 4\left(\left(1+\left(\frac{1}{3}\right)^{\frac{3}{2}}\right)\sin(\theta_2) - \left(\frac{2}{3}\right)^{\frac{3}{2}}\cos(\theta_2)\right)^2 + 8\left(\frac{2}{3}\right)^{\frac{3}{2}}\left(1+\left(\frac{1}{3}\right)^{\frac{3}{2}}\right)|\sin(\theta_2)\cos(\theta_2)|$$

Note that while the first criterion can be violated, the second cannot. As such, for this specific state, the first criterion of Hillery et al. outperforms the second. As such, we can conclude that there are at least states that are detectable with our criterion while being undetectable by Theorem 9.25.

COMPARISON WITH THE CRITERION BY HUBER ET AL.

Note that the multipartite generalisation of the criterion by **Huber et al.** is actually valid for all biseparable states (see Huber and Mintert (2010)). As such, this criterion is less strong than our criterion, for example when detecting states like ρ_{abc}.

12

CONCLUSION

In Part I of the thesis, the Cauchy-Schwarz inequality, Holder's inequality, the AM-GM inequality, Jensen's inequality, Minkowksi's inequality and the ladder of power means as well as their extensions and generalisations were proved and discussed on finite dimensional vector spaces. It was shown that an intimate relationship exists between these inequalities: first of all, all of the inequalities can be proved using just Jensen's inequality, which is a simple consequence of convexity. Secondly, the AM-GM inequality and the Cauchy-Schwarz inequality can be derived from one another. Lastly, even though Hölder's inequality is considered to be a generalisation of the Cauchy-Schwarz inequality and Minkowksi's inequality is considered to be a consequence of Hölder's inequality, it was shown that each these three inequalities can be used to prove the others.
Since these explorations were restricted to finite dimensional vector spaces, the consequences of extending the range of study to infinite dimensional vector spaces were considered. Specifically, it was shown that the rich structure that is present on $\mathbb{R}^n$ equipped with a p-norm (to which every finite dimensional normed vector space over the real numbers with dimension n is isomorphic) is lost in this extension, as the topologies under each of these norms are distinct.

In Part II, entanglement detection - and specifically non-linear entanglement witnesses derived in Wölk et al. (2014), which are based on the inequalities discussed extensively in Part I - is studied. This was first done for the case of two particles, and was later generalised to multipartite systems. Then, the bipartite criterion was derived and an implicit scheme in aforementioned paper for developing a general class of multipartite entanglement witnesses was explicated. These criteria were then assessed as follows: first, the optimality condition for the bipartite criterion was derived. Then, the criteria were assessed by comparing them with other well-known entanglement criteria. It was shown that while the inequality-based criterion is weaker than the Positive Partial Transpose criterion, it is more general and often even stronger than many entanglement witnesses criteria currently used. Lastly, it was shown that our criterion is nevertheless limited to some class of Negative Partial Transpose states, namely those that are 1-distillable.

ACKNOWLEDGEMENT

I want to express my sincere gratitude to my supervisors, Dr. Johannes Vermeer and
Dr. David Elkouss Coronas for sharing their insights, knowledge and passion with me.

Appendices

A

Supplement to Part I of the Thesis

We first introduce the proof to Theorem 1.2.

Proof. Let $\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{b_i}$ and $\mathbf{w} = \sum_{i=1}^{n} w_i \mathbf{b_i}$. Then:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \langle \sum_{i=1}^{n} v_i \mathbf{b_i}, \mathbf{w} \rangle = \sum_{i=1}^{n} v_i^{\dagger} \langle \mathbf{b_i}, \mathbf{w} \rangle = [v_1, \ldots, v_n]^{\dagger} \begin{bmatrix} \langle \mathbf{b_1}, \mathbf{w} \rangle \\ \vdots \\ \langle \mathbf{b_n}, \mathbf{w} \rangle \end{bmatrix} = [v_1, \ldots, v_n]^{\dagger} \begin{bmatrix} \langle \mathbf{b_1}, \sum_{i=1}^{n} w_i \mathbf{b_i} \rangle \\ \vdots \\ \langle \mathbf{b_n}, \sum_{i=1}^{n} w_i \mathbf{b_i} \rangle \end{bmatrix}$$

$$= [v_1, \ldots, v_n]^{\dagger} \begin{bmatrix} \langle \mathbf{b_1}, \sum_{i=1}^{n} w_i \mathbf{b_i} \rangle \\ \vdots \\ \langle \mathbf{b_n}, \sum_{i=1}^{n} w_i \mathbf{b_i} \rangle \end{bmatrix} = [v_1, \ldots, v_n]^{\dagger} \begin{bmatrix} \sum_{i=1}^{n} \langle \mathbf{b_1}, w_i \mathbf{b_i} \rangle \\ \vdots \\ \sum_{i=1}^{n} \langle \mathbf{b_n}, w_i \mathbf{b_i} \rangle \end{bmatrix}$$

$$= [v_1, \ldots, v_n]^{\dagger} \begin{bmatrix} \sum_{i=1}^{n} w_i \langle \mathbf{b_1}, \mathbf{b_i} \rangle \\ \vdots \\ \sum_{i=1}^{n} w_i \langle \mathbf{b_n}, \mathbf{b_i} \rangle \end{bmatrix} = [v_1, \ldots, v_n]^{\dagger} \begin{bmatrix} \sum_{i=1}^{n} \langle \mathbf{b_1}, \mathbf{b_i} \rangle \\ \vdots \\ \sum_{i=1}^{n} \langle \mathbf{b_n}, \mathbf{b_i} \rangle \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = [\mathbf{v}]_B^{\dagger} G_B [\mathbf{w}]_B$$

□

From this Theorem, the proof of Corollary 1.2.1 follows easily:

Proof. Note that for any orthonormal basis we have that $\langle b_i, b_j \rangle = \delta_{ij}$. Thus $G_B = I$. Now Theorem 1.2 yields the expected result. □

$$\langle \mathbf{u}, \mathbf{v}, \mathbf{w} \rangle = \langle \sum_{i=1}^{n} u_i \mathbf{b_i}, \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} u_i \langle \mathbf{b_i}, \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} u_i \langle \mathbf{v}, \mathbf{b_i}, \mathbf{w} \rangle = \sum_{i=1}^{n} u_i \langle \sum_{j=1}^{n} v_j \mathbf{b_j}, \mathbf{b_i}, \mathbf{w} \rangle$$

$$= \sum_{i=1}^{n} u_i \sum_{j=1}^{n} v_j \langle \mathbf{b_j}, \mathbf{b_i}, \mathbf{w} \rangle = \sum_{i=1}^{n} u_i \sum_{j=1}^{n} v_j \langle \mathbf{w}, \mathbf{b_j}, \mathbf{b_i} \rangle = \sum_{i=1}^{n} u_i \sum_{j=1}^{n} v_j \langle \sum_{k=1}^{n} w_k \mathbf{b_k}, \mathbf{b_j}, \mathbf{b_i} \rangle$$

$$= \sum_{i=1}^{n} u_i \sum_{j=1}^{n} v_j \sum_{k=1}^{n} w_k \langle \mathbf{b_k}, \mathbf{b_j}, \mathbf{b_i} \rangle = \sum_{i=1}^{n} u_i \sum_{j=1}^{n} v_j \sum_{k=1}^{n} w_k \delta_{ijk} = \sum_{i=1}^{n} u_i v_i w_i = \langle \mathbf{u}, \mathbf{v}, \mathbf{w} \rangle_B$$

B

Supplement to Part II of the Thesis

B.1. On the definition of partial traces

Given a measurement M^A performed on some closed system, say ρ^A. Let us now consider the composite system AB. Let M be the measurement operator on AB which yields the same result as M^A,so $M = \sum_m mP_m \otimes I_B = M^A \otimes I_B$, where P_m is the projector onto the the eigenspace corresponding to the eigenvalue m. If these measurement operators correspond to the same measurement, they should yield the same expectation value, so:

$$\mathrm{Tr}\left(f\left(\rho^{AB}\right)M^A\right) = \mathrm{Tr}\left(\rho^{AB}M\right) = \mathrm{Tr}\left(\rho^{AB}\left(M^A \otimes I_B\right)\right)$$

where we suppose that $f(\rho^{AB})$ is a function on ρ^{AB} which yields an appropriate density matrix describing subsystem A.

Theorem B.1. $f\left(\rho^{AB}\right) = \mathrm{Tr}_B(\rho^{AB})$ *and is uniquely defined.*

Proof. First, substitute $f\left(\rho^{AB}\right) = \mathrm{Tr}_B(\rho^{AB})$ and verify that still:

$$\mathrm{Tr}\left(f\left(\rho^{AB}\right)M^A\right) = \mathrm{Tr}\left(\rho^{AB}\left(M^A \otimes I_B\right)\right)$$

Note that $\rho^{AB} = \rho^A \otimes \rho^B$, and $\rho^A = \sum_i p_i |m_i\rangle\langle m_i|$ for some $\{p_i\}$. Now:

$$\begin{aligned}
\mathrm{Tr}\left(\mathrm{Tr}_B\left(\rho^{AB}\right)M^A\right) = Tr\left(\mathrm{Tr}_B\left(\rho^A \otimes \rho^B\right)M^A\right) &= \mathrm{Tr}\left(\rho^A \mathrm{Tr}\left(\rho^B\right)M^A\right)\\
&= \mathrm{Tr}\left(\rho^A M^A\right)\mathrm{Tr}\left(\rho^B I_B\right)\\
&= \mathrm{Tr}\left(\left(\rho^A \otimes \rho^B\right)\left(M^A \otimes I_B\right)\right)\\
&= \mathrm{Tr}\left(\rho^{AB}\left(M^A \otimes I_B\right)\right)
\end{aligned}$$

Consider an orthonormal basis M_i of the space of Hermitian operators with respect to the Hilbert-Schmidt inner product. Then:

$$\begin{aligned}
f\left(\rho^{AB}\right) = \sum_i M_i \mathrm{Tr}\left(M_i f\left(\rho^{AB}\right)\right) &= \sum_i M_i \mathrm{Tr}\left((M_i \otimes I_B)\rho^{AB}\right)\\
&= \sum_i M_i \mathrm{Tr}\left(\rho^{AB}(M_i \otimes I_B)\right)\\
&= \sum_i M_i \mathrm{Tr}\left(f\left(\rho^{AB}\right)M_i\right)\\
&= \sum_i M_i \mathrm{Tr}\left(\mathrm{Tr}_B\left(\rho^{AB}\right)M_i\right)
\end{aligned}$$

Since the right-hand side is invariant under our choice of orthonormal basis, $f(\rho^{AB})$ is unique. By making use of a Schmidt decomposition, this result can be generalised even further. □

B.2. SUPPLEMENTARY LINEAR ALGEBRA

We first introduce two definitions:

Definition B.1. Let A, B be two operators. The *commutator* $[A, B]$ of these operators is given by

$$[A, B] = AB - BA \tag{B.1}$$

Definition B.2. Let A, B be two operators. The *anti-commutator* $\{A, B\}$ of these operators is defined as

$$\{A, B\} = AB + BA \tag{B.2}$$

Bibliography

Bennett, C., Brassard, G., Crepeau, C., Jozsa, R., Peres, A., and Wootters, W. (1993). Teleporting an unknown quantum state via dual classical and einstein-podolsky-rosen channels.

Carothers, N. L. (2000). *Real Analysis*. CAMBRIDGE UNIVERSITY PRESS.

Chen, K. and Wu, L. (2003). A matrix realignment method for recognizing entanglement.

Choi, M. (1975). Completely positive linear maps on complex matrices.

de Vicente, J. (2007). Separability criteria based on the bloch representation of density matrices.

Einstein, A., Podolski, B., and Rosen, N. (1935). Reduction criterion of separability and limits for a class of distillation protocols.

Gittsovich, O., Gühne, O., Hyllus, P., and J. Eisert, J. (2008). Unifying several separability conditions using the covariance matrix criterion.

Grimmitt, G. and Welsh, D. (2014). *Probability: an Introduction*. Oxford University Press.

Gühne, O. (2010). Entanglement criteria and full separability of multi-qubit quantum states.

Gühne, O. and Lütkenhaus, N. (2005). Nonlinear entanglement witnesses.

Gühne, O. and Seevinck, M. (2018). Separability criteria for genuine multiparticle entanglement.

Gühne, O. and Tóth, G. (2008). Entanglement detection.

Hillery, M., Dung, H., and Zheng, H. (2010). Conditions for entanglement in multipartite systems.

Hiroshima, T. (2003). Majorization criterion for distillability of a bipartite quantum state.

Horodecki, M. (2001). Entanglement measures.

Horodecki, M. and Horodecki, P. (1999). Reduction criterion of separability and limits for a class of distillation protocols.

Horodecki, M., Horodecki, P., and Horodecki, R. (1996). Separability of mixed states: Necessary and sufficient conditions.

Horodecki, M., Horodecki, P., and Horodecki, R. (1998). Mixed-state entanglement and distillation: is there a “bound” entanglement in nature?

Horodecki, M., Horodecki, P., and Horodecki, R. (2006). Separability of mixed quantum states: Linear contractions and permutation criteria.

Huber, M. and Mintert, F. (2010). Detection of high-dimensional genuine multipartite entanglement of mixed states.

Ioannou, L. (2007). Computational complexity of the quantum separability problem.

Mattle, K., Weinfurter, H., Kwiat, P., and Zeilinger, A. (1995). Dense coding in experimental quantum communication.

Nielsen, M. and Chuang, I. (2012). *Quantum Computation and Quantum Information*. Cambridge University Press.

Rice, J. (2007). *Mathematical Statistics and Data Analysis*. Thomson Brooks/Cole.

Steele, J. M. (2004). *Cauchy-Schwarz Master Class*. Cambridge University Press.

Vermeer, J. (2017). *Dictaat TN2244WI: Lineaire Algebra, deel 2 voor Technische Natuurkunde*. Delft University of Technology.

Watrous, J. (2004). Many copies may be required for entanglement distillation.

Werner, R. (1989). Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model.

Wolf, M. (2003). Partial transposition in quantum information theory.

Wölk, S., Huber, M., and Gühne, O. (2014). A unified approach to entanglement criteria using the cauchy-schwarz and hölder inequalities.

Woronowicz, S. (1976). *Nonextendible Positive Maps*.

Yu, S. and Liu, N. (2005). Entanglement detection by local orthogonal observables.