Research Group: *Entanglement of Complex Quantum Systems*
Max–Planck–Institute of Quantum Optics

Recovery Maps for Approximate Markov Chains—a Matrix Product State Perspective

Bachelor's Thesis

Julian Sieber

Examiner:	Prof. Dr. Norbert Schuch
Advisor:	Yimin Ge
Submission Date:	September 18, 2017

Contents

Acknowledgments

The author is deeply indebted to the supervisor of this thesis, Prof. Dr. Norbert Schuch, for proposing this interesting and manifold topic. During the months of work on this thesis, he was a constant source of inspiration as well as motivation and his soft guidance shaped the result significantly. I really appreciate the effort and time he spent in order to successfully supervise this thesis.
Furthermore, I owe gratitude to Yimin Ge for the helpful discussions around this thesis. He did not hesitate taking time to answer several questions which arose during the work on the topic.
I would like to thank my fellow student Johannes R. Kager for proofreading this thesis and for pointing out several typos.
Last but not least, the author kindly acknowledges the partial financial support from the German National Merit Foundation.

Zusammenfassung

Die vorliegende Arbeit bietet einen Einstieg in die Quanten– sowie Quanteninformationstheorie. Im ersten Kapitel führen wir die für unsere Arbeit notwendigen Konzepte der Quantentheorie ein. Besonderes Interesse liegt hierbei, neben dem Studium vollständig positiver Abbildungen, auf der Untersuchung der Schatten p–Normen, deren Hölder–Ungleichung integraler Bestandteil des Beweises eines zentralen Lemmas im hinteren Teil der Arbeit ist. Im darauffolgenden Kapitel beschäftigen wir uns mit den grundlegenden Konzepten der klassischen Informationstheorie und stellen den Zusammenhang derer zu ihren Quantenanaloga her. Danach werden wir in Kapitel drei eine kurze Einführung in die Thematik der Matrixproduktzustände geben, welche wir im abschließenden Kapitel verwenden, um das Hauptresultat dieser Arbeit zu zeigen. Dieses ist eine Fehlerabschätzung des Prozesses, welchen wir nachfolgend kurz erläutern möchten. Wir betrachten eine Spinkette mit N Plätzen, jeder mit dem assoziierten Hilbert–Raum $\mathbb{C}^d$. Unter der Voraussetzung, dass der Zustand des Gesamtsystems rein ist, zeigen wir die Existenz eines Matrixproduktzustands, welcher diesen sehr gut approximiert. Wir starten mit dem ausgespurten Zustand der ersten beiden Plätze und rekonstruieren den Gesamtzustand durch sukzessive Anwendung einer «recovery map». Dabei bauen wir maßgeblich auf den Resultaten von Fawzi und Renner [FR15] auf. Ferner wird in selbigem Kapitel ein alternativer Ansatz von Wilde [Wil15] vorgestellt, welcher bessere Aussagen über die Struktur der «recovery map» macht. Abschließend untersuchen wir die Abhängigkeit der «bond dimension» von der Systemgröße und geben das notwendige Skalierungsverhalten der Quanten–Transinformation in Abhängigkeit der Systemgröße, um polynomielles bzw. quasi–polynomielles Wachstum zu erreichen, an.

Introduction

Back in 1858, the 16–year–old Max Planck arrived at the University of Munich in order to start his studies in physics and the local professor Philipp von Jolly discouraged him in the strongest terms possible. In his book [Pla33], Planck quoted von Jolly with the words

> «In this field, almost everything is already discovered, and all that remains is to fill a few holes.»

Fortunately, Planck was not swayed by this statement and laid the cornerstone of a completely new branch of physics, the quantum physics, by his seminal work towards an explanation of the black–body radiation 16 years later. Since then, quantum physics has become the most successful physical theory and has given impetus to numerous advances in almost all branches of mathematics. Quantum mechanics is ubiquitous and virtually all modern physical theories are based upon it. In 1935, Einstein, Podolsky and Rosen [EPR35] discovered the phenomenon of entanglement which is one of the most counterintuitive notion in quantum mechanics. Even these famous minds lacked an explanation of the fact that the successful concept of describing a system of point masses without interaction by computing the trajectory of each constituent individually fails completely in the quantum realm. The major difference is the necessity of considering a quantum system as a whole since even particles far away from the one of interest may effect the latter. Nevertheless, a few months later, Erwin Schrödinger coined the term «entanglement» and gave a first explanation of this phenomenon [Sch35; Sch36]. It is worthwhile to note that the notion of entanglement arises directly from the mathematical description of composite quantum systems, founded by the outstanding work of Dirac [Dir25] and von Neumann [Von32], as we will see in chapter 1.

In the more recent years, roughly in the 1990s, quantum information theory, a field in the intersection of physics and computer science, was established and has been pushed forward in a stunning pase since then. The theoretical framework of quantum computers is rather sophisticated; nevertheless we lack concrete large–scale technical implementations of this technology. Even in the light of the difficulties, there is a wide consensus in the scientific community that a commercial use of quantum computers is only a question of time. Some of the world's leading computer and software companies, eg. Microsoft, IBM, HP and NEC, host active quantum computation groups in

their research facilities.
Apart from the pecuniary aspects of quantum information, this theory has had a major impact in the field of quantum many–body physics. In particular, studying one– or higher–dimensional spin chains or systems, as we will do in this thesis, requires several notions from quantum information. The reader may note that we use the term «spin chain» in a wide sense as we consider a d–dimensional Hilbert space associated to each site of the chain. Unless stated otherwise, we do not vary this dimension throughout the system since this would usually lead to more complicated expressions rather than to new physics. At a first glance, studying spin chains might seem far away from the physical relevant particle systems. But in particular cases by invoking the Jordan–Wigner transformation [JW28], it can be shown that the spin chain setting is equivalent to a fermionic system on a lattice with short–range interactions.
This thesis is mainly motivated by the seminal work of Fawzi and Renner [FR15] which brought the notion of the Petz recovery map [Pet86; Pet88] back to the focus of research. Suppose we have spin chain S of finite length. S can be thought of a set containing all the sites of the chain. We partition the system into three pairwise disjoint subsystems A, B and C such that

$$S = A \dot{\cup} B \dot{\cup} C.$$

The work of Fawzi and Renner gives an upper bound on the error we make by recovering the total system's state ρ_S from the marginal ρ_{AB} by an application of a suitable map.

Our Main Results

Based on the theorem developed by Fawzi and Renner, we investigate an iterative recovery procedure on a spin chain of length N starting with the marginal state describing the first two sites and show that we can construct a matrix product state $|\tilde{\Psi}\rangle$ provided the total system's state $\sigma_{1,\dots,N}$ is pure, that is, $\sigma_{1,\dots,N} = |\Psi\rangle\langle\Psi|$. Additionally, we prove the bound

$$\| \, |\tilde{\Psi}\rangle\langle\tilde{\Psi}| - |\Psi\rangle\langle\Psi| \, \|_1 \leqslant C\sqrt{N-2}$$

for some explicitly determined $C > 0$ and study the dependence between this upper bound on the error and the bond dimension of the matrix product state. To this end, we impose different assumptions on the saturation behavior of the von–Neumann entropy in order to achieve polynomial, quasi–polynomial and an exponential scaling of the bond dimension with the system size.

A Reader's Guide

We aim at giving a self–contained treatment of the subject which requires no previous knowledge about quantum mechanics. Only a solid background in linear algebra, undergraduate analysis and basic complex analysis is assumed. In particular, we do not review the notion of the tensor product

of Hilbert spaces since the reader may find a thorough exposition of this topic in the classical book of Reed and Simon [RS81] as well as in the recently published comprehensive course on analysis by Simon [Sim15]. Any considered Hilbert space will be of finite dimension* and hence no techniques from functional analysis are needed. During the course of this thesis, we will encounter the need of studying the linear maps between Hilbert spaces which are in finite dimensions better known as matrices. Nevertheless, we adopt the notation $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_B)$ (instead of $\mathrm{Hom}(\mathcal{H}_A, \mathcal{H}_B)$) with $\mathcal{B}(\mathcal{H}_A, \mathcal{H}_A) = \mathcal{B}(\mathcal{H}_A)$ (instead of $\mathrm{End}(\mathcal{H}_A)$) widely used in the relevant literature. In chapter 1, we consider matrices whose entries are matrices itself. To this end, it is convenient to take a more algebraic point of view and to denote the C^*-algebra of complex $n \times n$ matrices by $\mathfrak{M}_n = (\mathcal{B}(\mathbb{C}^n), +, \cdot)$ with the ordinary matrix sum and product.

- Chapter 1 provides the fundamental notions of quantum theory and some of the mathematical basics. The reader may find it convenient that we present the proofs of the theorems and lemmas. Moreover, we study the classification of maps $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ and prove two major theorems for most important class, the completely positive maps. Eventually, we introduce and discuss the Schatten p–norms and present another distance measure of quantum states, the fidelity.

- Even if this is a physics thesis, we have to understand a few notions of computer science. In particular, chapter 2 provides a brief review of classical information theory as well as deeper one of the quantities of quantum information which are needed in the subsequent chapters. Furthermore, we investigate the interplay of probability theory and computer science; especially we give an introduction to Markov chains in both fields.

- Matrix product states provide a powerful ansatz in order to simulate one–dimensional quantum systems. Chapter 3 shows how this notion arises naturally from the study of spin chains. Moreover, we introduce a convenient graphical notation for tensor networks and provide the reader with a few states with an easy matrix product representation.

- In chapter 4, we use all the acquired knowledge to study the recoverability of states on a spin chain. We present two different approaches which both lead to the same explicit form of the recovery map as well as an identical upper bound on the fidelity between the recovered and the original state in terms of the quantum conditional mutual information of the tripartite system.

- Our own results concerning the reconstruction procedure mentioned above are encapsulated in chapter 5. We show that we can construct a matrix product approximation of the state $\sigma_{1,\ldots,N} = |\Psi\rangle\langle\Psi|$ and discuss

*The result of Fawzi and Renner even applies to infinite–dimensional but separable Hilbert spaces.

upper bounds on the inevitable error we have made. Moreover, we establish conditions on the quantum mutual information in order to have a polynomial, or at most quasi–polynomial, ascent of the bond dimension as the system size increases.

- Eventually, we provide a conclusion of this thesis and present an open question which might be tackled by further work.

We postpone the proof of the important theorem of Uhlmann (*cf.* theorem 1.26) to appendix A since it is not a central but a very helpful result in numerous proofs of this thesis. In this section the reader may also find an exposition on the theory of measurement in quantum mechanics, especially on the notion of positive operator valued measures, which complements the content of chapter 1 but is of no need for the rest of our work. Last but not least, we review the proof of Wilde's recoverability theorem [Wil15] in order to achieve a deeper understanding of the structure of a possible recovery map.

Chapter 1

A Primer on Quantum Theory

The main purpose of this chapter is the presentation of most of the definitions used later on in this thesis rather than developing any original thoughts. For the reader's convenience we present the theorems which underlay our work and prove them afterwards. Further information can be found in the textbooks of Nielsen and Chuang [NC10] and Wilde [Wil17].

1.1 Quantum States and Quantum Channels

From an abstract point of view, any experiment in the quantum realm can be divided into two sections. The first one is the so-called **preparation**. It is an intrinsic property of the considered quantum system. In contrast to macroscopic physics, we cannot make any deterministic predictions. The only chance we have is to make an educated guess about the probability distribution of the possible outcomes we may obtain when conducting a **measurement**. Note that two states which at a first glance look quite different may produce the same probability distribution. Thus, from a mathematical point of view, we define a **state** of the system as equivalence class of preparations which produce the same probability distribution in all measurements. Our experiment does not only consist of the preparation: every experimental setup is developed to do measurements. In contrast to classical physics, the process of measurement might change the system drastically. We have to keep this fact in mind while developing our theory. A physicist refers to this phenomenon as «collapse of the wave function». In analogy to the definition of the state of a system, we refer to an **observable** as the equivalence class of measurements which give the same probability distribution in all preparations. Figure 1.1 depicts the explained situation. Due to the probabilistic structure of quantum mechanics, the experiment's outcome is only a (conditional) probability distribution. Thus, the experimenter has to be rather durable and conduct several measurements of a given preparation to obtain reasonable results.
In the following paragraphs, the preparation of the experiment is investigated *en détail*. An introduction to the measurement part and the concept of

positive operator valued measures can be found in appendix A.1.

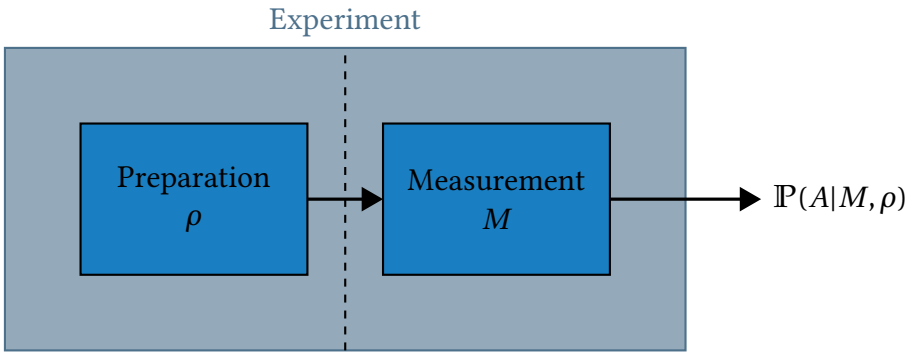


Figure 1.1: Schematic structure of a quantum experiment. Its outcome is the conditional probability of an event A given a measurement M and a preparation ρ. The preparation of the system determines the possible measurement outcomes completely.

The most fundamental notion of quantum theory is a quantum state or density operator (matrix).

Definition 1.1. Let $\mathscr{H}$ be a Hilbert space. An operator $\rho \in \mathscr{B}(\mathscr{H})$ fulfilling $\rho \geqslant 0$ as well as $\mathrm{tr}[\rho] = 1$ is called a **quantum state**. We denote by

$$\mathscr{D}(\mathscr{H}) = \left\{\rho \in \mathscr{B}(\mathscr{H}) \,\middle|\, \rho \text{ is a quantum state}\right\}$$

the set of all quantum states acting on the Hilbert space $\mathscr{H}$.

An immediate consequence of the definition above is that the set $\mathscr{D}(\mathscr{H})$ is convex, that is, for any $\rho_1, \rho_2 \in \mathscr{D}(\mathscr{H})$ also $\rho = \lambda\rho_1 + (1-\lambda)\rho_2$ is a valid density operator if $\lambda \in [0,1]$. Due to the continuity of the trace, we need not restrict to finite convex combinations. For example, $\sum_{n\in\mathbb{N}} 2^{-n}\rho_n$ is a valid quantum state for $(\rho_n)_{n\in\mathbb{N}} \subset \mathscr{D}(\mathscr{H})$. Since every density matrix ρ is positive semidefinite, it is especially Hermitian. Hence, it admits a spectral decomposition of the form

$$\rho = \sum_{i=1}^{\dim(\mathscr{H})} \lambda_i \,|i\rangle\langle i|$$

with the non–negative eigenvalues λ_i summing up to one and an orthonormal basis (onb) $\{|i\rangle\}_{i=1}^{\dim(\mathscr{H})}$ of the Hilbert space $\mathscr{H}$ consisting of ρ's eigenvectors. For a unit vector $|\psi\rangle \in \mathscr{H}$, we refer to $|\psi\rangle\langle\psi| \in \mathscr{D}(\mathscr{H})$ as a **pure state**. Thus, the spectral decomposition suggests the interpretation of ρ as a convex combination of pure states. From a physical point of view, the eigenvalue decomposition represents a statistical mixture of pure states $|i\rangle\langle i|$ which occur with a–priori probabilities λ_i. We call such a density operator a **mixed state** hereafter. Note that the decomposition of a mixed state into pure ones

is not unique. In fact, defining $\mu_i = \langle\sqrt{\rho}\varphi_i|\sqrt{\rho}\varphi_i\rangle$ for an onb $\{|\varphi_i\rangle\}_i \subset \mathscr{H}$ and the unit vectors $|\psi_i\rangle = \mu_i^{-1/2}\sqrt{\rho}\,|\varphi_i\rangle$ for $\mu_i \neq 0$ yields

$$\rho = \sum_{i:\mu_i\neq 0} \mu_i\,|\psi_i\rangle\langle\psi_i|\,.$$

To see this, let $|\eta\rangle \in \mathscr{H}$ be arbitrary and observe that

$$\left\langle \eta \middle| \left(\sum_{i:\mu_i\neq 0} \mu_i\,|\psi_i\rangle\langle\psi_i| \right)\eta \right\rangle = \sum_{i=1}^{\dim(\mathscr{H})} \left|\langle\sqrt{\rho}\eta|\varphi_i\rangle\right|^2 = \left\|\sqrt{\rho}\eta\right\|^2 = \langle\eta\,|\,\rho\eta\rangle\,.$$

Eventually, the fact that the μ_i sum up to one is shown by a straight–forward computation:

$$\sum_{i=1}^{\dim(\mathscr{H})} \mu_i = \sum_{i=1}^{\dim(\mathscr{H})} \langle\varphi_i|\rho\varphi_i\rangle = \mathrm{tr}[\rho] = 1.$$

Now we are able to prove an upper bound of the trace of ρ^2. Computing ρ^2 explicitly yields

$$\rho^2 = \sum_{j,k} \lambda_j\lambda_k\,|\varphi_j\rangle\langle\varphi_j|\varphi_k\rangle\langle\varphi_k| = \sum_j \lambda_j^2\,|\varphi_j\rangle\langle\varphi_j| \neq \rho$$

in general. The trace is then given as

$$\mathrm{tr}[\rho^2] = \sum_{i,j} \lambda_j^2 |\langle\varphi_i|\varphi_j\rangle|^2 = \sum_j \lambda_j^2 \leqslant \left(\sum_j \lambda_j\right)^2 = 1.$$

In order to decide whether a given state is pure or mixed, the following theorem provides a strong criterion.

Theorem 1.2. *Let $\rho \in \mathscr{D}(\mathscr{H})$. Then the following statements are equivalent:*

(i) *ρ is a pure state.*

(ii) $\mathrm{tr}\left[\rho^2\right] = 1$

Proof. First assume that ρ is a pure state, that is, $\rho = |\varphi\rangle\langle\varphi|$. Note that $\rho^2 = \|\varphi\|^2\rho = \rho$ and thus $\mathrm{tr}[\rho] = \mathrm{tr}[\rho^2]$.
For the other implication, we observe that the trace of ρ^2 reads

$$\mathrm{tr}[\rho^2] = \sum_j \lambda_j^2$$

where λ_j are the eigenvalues of ρ. The assumption $\mathrm{tr}[\rho^2] = 1 = \mathrm{tr}[\rho]$ implies that there is only one non–zero eigenvalue λ_k which is equal to one. Hence, $\rho = |\varphi_k\rangle\langle\varphi_k|$. □

The following definition introduces a very important operator which maps a state on the bipartite system $\mathscr{H}_A \otimes \mathscr{H}_B$ to a density operator on $\mathscr{H}_A$.

Definition 1.3. The operator $\mathrm{tr}_B : \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B) \to \mathscr{D}(\mathscr{H}_A)$, $\rho \mapsto \rho_A$, defined via

$$\mathrm{tr}[\rho(X \otimes \mathbb{1})] = \mathrm{tr}[\rho_A X] \quad \forall X \in \mathscr{B}(\mathscr{H}_A) \tag{1.1}$$

is called the **partial trace**. The matrix ρ_A is called the **reduced density operator** of ρ w.r.t. $\mathscr{H}_A$.
We refer to a state $\rho_{AB} \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$ as **product state** if there are $\rho_A \in \mathscr{D}(\mathscr{H}_A)$ and $\rho_B \in \mathscr{D}(\mathscr{H}_B)$ such that $\rho_{AB} = \rho_A \otimes \rho_B$.

Since we called the operator tr_B «the» partial trace in the preceding definition, we have to ensure that the construction is unique. To see this fact, suppose we have two density operators $\rho_A, \tilde{\rho}_A$ which both satisfy equation (1.1) for the same $\rho \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$. Then, in particular $\mathrm{tr}[\rho_A X] = \mathrm{tr}[\tilde{\rho}_A X]$ holds. Hence, we find $\mathrm{tr}[(\rho_A - \tilde{\rho}_A)X] = \langle \rho_A - \tilde{\rho}_A | X \rangle_{HS} = 0$ for every $X \in \mathscr{B}(\mathscr{H}_A)$ where $\langle A|B\rangle_{HS} = \mathrm{tr}[A^\dagger B]$ denotes the **Hilbert–Schmidt inner product** (a more detailed discussion can be found in § 1.2.1). Eventually, we conclude $(\rho_A - \tilde{\rho}_A) \in \mathscr{B}(\mathscr{H}_A)^\perp = \{0\}$.
At the first glance, the definition of the partial trace might seem a bit cryptic to the unfamiliar reader. Therefore, we present the following two examples in order to make equation (1.1) more vivid.

Example 1.4.

(i) For a product state $\rho = |\psi\rangle\langle\varphi| \otimes |\xi\rangle\langle\eta| \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$ the partial trace is given by $\mathrm{tr}_B[\rho] = \langle\eta|\xi\rangle\, |\psi\rangle\langle\varphi|$.

(ii) Consider the vector $|\psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle \otimes |1\rangle + |2\rangle \otimes |2\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2$ (this state is known as Bell or EPR state). Here and in the following, we occasionally denote the i–th element of an onb by $|i\rangle$. The density matrix is given as

$$\begin{aligned}\rho = |\psi\rangle\langle\psi| &= \frac{1}{2}(|1\rangle \otimes |1\rangle + |2\rangle \otimes |2\rangle)(\langle 1| \otimes \langle 1| + \langle 2| \otimes \langle 2|) \\ &= \frac{1}{2}(|1\rangle\langle 1| \otimes |1\rangle\langle 1| + |1\rangle\langle 2| \otimes |1\rangle\langle 2| \\ &\quad + |2\rangle\langle 1| \otimes |2\rangle\langle 1| + |2\rangle\langle 2| \otimes |2\rangle\langle 2|).\end{aligned}$$

Computing the partial trace gives

$$\rho_A = \mathrm{tr}_B(\rho) = \frac{1}{2}(|1\rangle\langle 1| + |2\rangle\langle 2|).$$

Observe that $\mathrm{tr}[\rho_A^2] = 1/2$ and hence theorem 1.2 shows the important fact that tracing out a pure state yields a mixed one in general.

The following proposition ensures that by applying the partial trace to a state acting on a bipartite system we always obtain a density operator on the re-

maining Hilbert space.

Proposition 1.5. *Let $\rho \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$, then*

(i) $\operatorname{tr}[\rho_A] = \operatorname{tr}[\rho]$,

(ii) $\rho \geqslant 0 \Rightarrow \rho_A \geqslant 0$.

Proof. Ad (i): $\operatorname{tr}[\rho_A] = \operatorname{tr}[\rho_A \mathbb{1}] = \operatorname{tr}[\rho(\mathbb{1} \otimes \mathbb{1})] = \operatorname{tr}[\rho]$.
Ad (ii): Expanding the scalar product in the onb $\{|\psi_j\rangle\}_j$ yields

$$\langle \varphi \,|\, \rho_A \varphi \rangle = \sum_j \langle \psi_j \,|\, \rho_A \varphi \rangle \langle \varphi \,|\, \psi_j \rangle = \operatorname{tr}[\rho_A |\varphi\rangle\langle\varphi|] = \operatorname{tr}[\rho[(|\varphi\rangle\langle\varphi|) \otimes \mathbb{1}]] \geqslant 0$$

since $(|\varphi\rangle\langle\varphi|) \otimes \mathbb{1} \geqslant 0$. □

As we have seen in example 1.4, the partial trace of a pure state may be mixed. The following theorem shows that every mixed state can be obtained by tracing out a pure state.

Theorem 1.6 (Purification). *Let $\rho \in \mathscr{B}(\mathscr{H})$ be a density operator. Then there is a normalized vector $|\psi\rangle \in \mathscr{H} \otimes \mathscr{H}_E$ such that ρ_A is the reduced density operator of $|\psi\rangle\langle\psi|$ w.r.t. $\mathscr{H}$.*

Proof. The density operator ρ admits a spectral decomposition of the form $\rho = \sum_j \lambda_j |\varphi_j\rangle\langle\varphi_j|$ with $\lambda_j \geqslant 0$. The vector

$$|\psi\rangle = \sum_{k:\lambda_k>0} \sqrt{\lambda_k}(|\varphi_j\rangle \otimes |\varphi_j\rangle)$$

has the desired property. Let us check it shortly:

$$|\psi\rangle\langle\psi| = \sum_{i,j} \sqrt{\lambda_i \lambda_j}(|\varphi_i\rangle\langle\varphi_j| \otimes |\varphi_i\rangle\langle\varphi_j|)$$

and thus

$$\operatorname{tr}_E[|\psi\rangle\langle\psi|] = \sum_{i,j} \sqrt{\lambda_i \lambda_j}\, \langle\varphi_i|\varphi_j\rangle\, |\varphi_i\rangle\langle\varphi_j| = \sum_j \lambda_k |\varphi_j\rangle\langle\varphi_j| = \rho.$$

□

Remark 1.7. The purification of quantum state $|\psi\rangle\langle\psi|$ is only unique up to a unitary $\mathbb{1} \otimes U$ since $\operatorname{tr}_B[(\mathbb{1} \otimes U)|\psi\rangle\langle\psi|(\mathbb{1} \otimes U)^\dagger] = \operatorname{tr}_B[|\psi\rangle\langle\psi|]$.

The concept of purification simplifies our lives a lot. In particular, we might always consider a general quantum state as a pure one acting on a higher dimensional system which given by the original Hilbert space tensorized with some appropriate ancillary system. The phrase «going to the church of the larger Hilbert space» coined by John Smolin depicts this dilation in felicitous manner.

1.1.1 Entanglement of Quantum States

The phenomenon of entanglement arises immediately from the mathematical notion of the tensor product. This feature lays at the core of quantum information theory. A quantifaction of entanglement is obtained by considering the Schmidt decomposition which we introduce in this section.

Definition 1.8. A vector $|\eta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is called **factorized** if there are $|\varphi\rangle \in \mathcal{H}_A$ and $|\psi\rangle \in \mathcal{H}_B$ such that $|\eta\rangle = |\varphi\rangle \otimes |\psi\rangle$. Otherwise we say $|\eta\rangle$ is **entangled**.

The best way of illustrating the notion of an entangled vector is to study an example.

Example 1.9. Consider the orthonormal vectors $|\varphi_1\rangle, |\varphi_2\rangle \in \mathcal{H}$ and the complex scalars $\alpha, \beta \in \mathbb{C} \setminus \{0\}$ with $|\alpha|^2 + |\beta|^2 = 1$. Then we claim that the unit vector

$$|\eta\rangle = \alpha(|\varphi_1\rangle \otimes |\varphi_2\rangle) + \beta(|\varphi_1\rangle \otimes |\varphi_2\rangle) \in \mathcal{H} \otimes \mathcal{H}$$

is entangled. To prove the assertion we complete $\{|\varphi_1\rangle, |\varphi_2\rangle\}$ to an onb of $\mathcal{H}$. Thus, any vector $|\varphi\rangle \otimes |\psi\rangle \in \mathcal{H} \otimes \mathcal{H}$ can be written as

$$|\varphi\rangle \otimes |\psi\rangle = \left(\sum_i c_i |\varphi_i\rangle\right) \otimes \left(\sum_j d_j |\varphi_j\rangle\right) \tag{1.2}$$

with $c_i, d_j \in \mathbb{C}$. In order to represent our vector $|\eta\rangle$, we get the following non–trivial constraints:

$$c_1 d_1 = \alpha, \quad c_2 d_2 = \beta, \quad c_1 d_2 = c_2 d_1 = 0 \quad \text{and } c_j = d_j = 0 \; \forall \, j > 2.$$

Hence, $\alpha\beta = c_1 d_2 c_2 d_1 = 0$ which implies $\alpha = 0$ or $\beta = 0$. Consequently, we obtain a contradiction and we conclude that $|\eta\rangle$ is entangled.

Equation (1.2) can be seen as motivation of the Schmidt decomposition which we want to present in the following theorem.

Theorem 1.10 (Schmidt decomposition). *Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be two Hilbert spaces with $d_A = \dim(\mathcal{H}_A)$ and $d_B = \dim(\mathcal{H}_B)$. For any vector $|\eta\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ there exist non–negative numbers $\lambda_i \geqslant 0$, which are called the **Schmidt coefficients** of $|\eta\rangle$, as well as onb's $\{|e_1\rangle, \ldots, |e_{d_A}\rangle\} \subset \mathcal{H}_A$ and $\{|f_1\rangle, \ldots, |f_{d_B}\rangle\} \subset \mathcal{H}_b$ such that*

$$|\eta\rangle = \sum_{i=1}^{\min\{d_A, d_B\}} \sqrt{\lambda_i} \, |e_i\rangle \otimes |f_i\rangle \, .$$

The integer $R_S := |\{\lambda_i \mid \lambda_i > 0\}|$ *is named the* ***Schmidt rank*** *of* $|\eta\rangle$.

Proof. First assume $d_A = d_B =: d$. As equation (1.2) implies, we may write

$$|\eta\rangle = \left(\sum_j c_j |\varphi_j\rangle\right) \otimes \left(\sum_k d_k |\varphi_k\rangle\right) = \sum_{j=1}^{d} \sum_{k=1}^{d} \underbrace{c_j d_k}_{=:c_{jk}} |\varphi_j\rangle \otimes |\psi_k\rangle$$

with onb's $\{|\varphi_1\rangle, \ldots, |\varphi_d\rangle\} \subset \mathscr{H}_A$ and $\{|\psi_1\rangle, \ldots, |\psi_d\rangle\} \subset \mathscr{H}_B$. The matrix $C := (c_{jk}) \in \mathbb{C}^{d\times d}$ has a singular value decomposition $C = U\Sigma V^\dagger$ with unitaries U, V and a diagonal matrix $\Sigma = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_d})$. Thus, the components of C are given as

$$c_{jk} = \sum_{\ell=1}^{d} u_{j\ell} \sqrt{\lambda_\ell} v^\dagger_{\ell k}$$

and we find

$$|\eta\rangle = \sum_{\ell=1}^{d} \sqrt{\lambda_\ell} \underbrace{\left(\sum_{j=1}^{d} u_{j\ell} |\varphi_j\rangle\right)}_{=:|e_\ell\rangle} \otimes \underbrace{\left(\sum_{k=1}^{d} v^\dagger_{\ell k} |\psi_k\rangle\right)}_{=:|f_\ell\rangle}.$$

Since U, V are unitary matrices, $\{|e_\ell\rangle\}_\ell$ and $\{|f_\ell\rangle\}_\ell$ form onbs.
In the case $d_A \neq d_B$, we know $\min\{d_A, d_B\} \geqslant \mathrm{rank}(C) = \mathrm{rank}(\Sigma)$. Thus Σ has at most $\min\{d_A, d_B\}$ non-zero elements. □

From the Schmidt decomposition, we want to draw some simple conclusions. The vector $|\eta\rangle$ is entangled if and only if the Schmidt rank is larger than or equal to two. Moreover, the Schmidt decomposition provides a convenient way to compute reduced density operators. Assume we got a state $\rho = |\psi\rangle\langle\psi|$ with Schmidt decomposition $|\psi\rangle = \sum_j \lambda_j^{1/2} (|e_j\rangle \otimes |f_j\rangle)$. Then the reduced density operators are given by

$$\rho_A = \sum_j \lambda_j |e_j\rangle\langle e_j|,$$

$$\rho_B = \sum_j \lambda_j |f_j\rangle\langle f_j|.$$

Thus, ρ_A and ρ_B have the same eigenvalues, only the multiplicity of the eigenvalue zero may differ.
After we have investigated the concept of entanglement for vectors, we now generalize it to density operators.

Definition 1.11. A state $\rho \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$ is called **separable** if there are $\{\lambda_j \geqslant 0\}_{j=1}^N$ with $\sum_{j=1}^N \lambda_j = 1$ and product states $\rho_A^{(j)} \otimes \rho_B^{(j)} \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$ such that

$$\rho = \sum_{j=1}^{N} \lambda_j (\rho_A^{(j)} \otimes \rho_B^{(j)}).$$

Note that due to the spectral decomposition we may restrict to the con-

sideration of pure product states in the convex decomposition above. If there is no such representation of ρ we call the state **entangled**.

For pure states the following theorem establishes a connection between the entanglement of vectors and density operators.

Theorem 1.12. *For a pure state $\rho = |\eta\rangle\langle\eta| \in \mathscr{D}(\mathscr{H}_1 \otimes \mathscr{H}_B)$ the following statements are equivalent:*

(i) *$|\eta\rangle$ has Schmidt rank one (i.e. it has the Schmidt decomposition $|\eta\rangle = |e\rangle \otimes |f\rangle$).*

(ii) *ρ is a product state.*

(iii) *ρ is separable.*

Proof. The only implication which needs to be proven is (iii)⇒(i). Since ρ is separable, we can write

$$|\eta\rangle\langle\eta| = \sum_j \lambda_j (|\varphi_j\rangle\langle\varphi_j| \otimes |\psi_j\rangle\langle\psi_j|)$$

where $\lambda_j > 0$. After multiplying the equation from the left as well as from the right by $|\eta\rangle\langle\eta|$, we find

$$|\eta\rangle\langle\eta| = \sum_j \lambda_j |\langle\eta|\varphi_j \otimes \psi_j\rangle|^2 |\eta\rangle\langle\eta| .$$

Due to Cauchy Schwarz, we get $|\langle\eta|\varphi_j \otimes \psi_j\rangle|^2 \leqslant 1$ and the trace normalization implies $\sum_j \lambda_j = 1$. Thus, $|\langle\eta|\varphi_j \otimes \psi_j\rangle|^2 = 1$ for all j and consequently $|\eta\rangle = z(|\varphi_j\rangle \otimes \psi_k)$ for some complex phase factor z. Eventually, the trace normalization implies $z = 1$. □

The following example shows that a mixed state of entangled vectors might be separable.

Example 1.13. Consider for $|\alpha|^2 + |\beta|^2 = 1$ the unit vectors

$$|\eta_\pm\rangle = \alpha(|\varphi_1\rangle \otimes |\varphi_2\rangle) \pm \beta(|\varphi_1\rangle \otimes |\varphi_2\rangle) \in \mathscr{H} \otimes \mathscr{H}$$

and recall from example 1.9 that they are entangled. But the mixed state

$$\begin{aligned}\rho &= \frac{1}{2}|\eta_+\rangle\langle\eta_+| + \frac{1}{2}|\eta_-\rangle\langle\eta_-| \\ &= |\alpha|^2(|\varphi_1\rangle\langle\varphi_1|) \otimes (|\varphi_1\rangle\langle\varphi_1|) + |\beta|^2(|\varphi_2\rangle\langle\varphi_2|) \otimes (|\varphi_2\rangle\langle\varphi_2|) \\ &= |\alpha|^2 |\varphi_1 \otimes \varphi_1\rangle\langle\varphi_1 \otimes \varphi_1| + |\beta|^2 |\varphi_2 \otimes \varphi_2\rangle\langle\varphi_2 \otimes \varphi_2|\end{aligned}$$

is clearly separable.

1.1.2 Classification of Matrix Maps

The standard reference for the matrix maps which are of interest in quantum information theory is the readable book of Bhatia [Bha09] which gives an exhaustive treatment of the subject. The following definition introduces the class of matrix maps which we study in detail later on.

Definition 1.14. We call a linear map $\mathcal{N} : \mathfrak{M}_n \to \mathfrak{M}_m$

- **positive** if $\mathcal{N}(A) \geqslant 0$ for any $A \geqslant 0$.
- **completely positive** if $\left(\mathbb{1}_p \otimes \mathcal{N}\right)$ is positive for any $p \in \mathbb{N}$
- **trace–preserving** if $\mathrm{tr}\left[\mathcal{N}(A)\right] = \mathrm{tr}\left[A\right]$ for any $A \in \mathfrak{M}_n$.
- **unital** if it preserves the identity, that is, $\mathcal{N}(\mathbb{1}) = \mathbb{1}$.

A completely positive and trace–preserving map is named a **quantum channel** (or quantum operation).

Example 1.15. We study three simple examples which are ubiquitous in quantum information theory. A common way to prove that a given map satisfies the requirements of a quantum channel is to show that it can be written as concatenation of these basic examples. The proof of the two demanded properties is obtained in all cases by a basic computation and is therefore left to the reader.

(i) **Extensions.** Let $\sigma \in \mathscr{D}(\mathscr{H}_B)$ a fixed density operator. The map $\mathcal{N} : \mathscr{D}(\mathscr{H}_A) \to \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$, $\rho \mapsto \rho \otimes \sigma$, is a quantum channel.

(ii) **Partial trace.** Proposition 1.5 together with the fact that the partial trace tensorized with $\mathbb{1}_n$ remains a partial trace shows that $\mathcal{N} : \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B)$, $\rho \mapsto \mathrm{tr}_B[\rho]$ is a quantum channel.

(iii) **Unitary evolution.** Let $U \in \mathfrak{M}_n$, where $n = \dim(\mathscr{H})$, be a unitary matrix. Then the map $\mathcal{N} : \mathscr{D}(\mathscr{H}) \to \mathscr{D}(\mathscr{H})$, $\rho \mapsto U\rho U^\dagger$, is a quantum channel.

A legitimate question which might arise at this point of the discussion is whether positivity and complete positivity are different properties. To answer this question we want to present the canonical example of the flip operator in the following.

Example 1.16. Fix a basis of $\mathfrak{M}_n$ and let

$$\begin{aligned} \Theta : \mathfrak{M}_n &\to \mathfrak{M}_n \\ A &\mapsto A^\dagger \end{aligned}$$

denote the Hermitian conjugation. We define the partial transposition

$\Theta \otimes \mathbb{1} : \mathscr{B}(\mathbb{C}^n \otimes \mathscr{H}) \to \mathscr{B}(\mathbb{C}^n \otimes \mathscr{H})$, $(\Theta \otimes \mathbb{1})(A \otimes B) = A^\dagger \otimes B$. Moreover, let us introduce the flip operator for a fixed onb $\{|i\rangle\}$ of $\mathbb{C}^n$

$$\mathbb{F} : \mathbb{C}^n \otimes \mathbb{C}^n \to \mathbb{C}^n \otimes \mathbb{C}^n$$
$$\mathbb{F} = \sum_{i,j=1}^{n} |i\rangle\langle j| \otimes |j\rangle\langle i|.$$

Observe that for $|\psi\rangle, |\varphi\rangle \in \mathbb{C}^n$, we have $\mathbb{F}(|\psi\rangle \otimes |\varphi\rangle) = |\varphi\rangle \otimes |\psi\rangle$ which is immediately verified by expanding the vectors in the onb $\{|i\rangle\}$. This property explains the name of the operator. The reason why we consider the flip operator is the fact that we find for the **maximally entangled state** $|\omega_n\rangle = 1/\sqrt{n} \sum_{i=1}^{n} |i\rangle \otimes |i\rangle$ (we have $R_S = n$)

$$(\mathbb{1} \otimes \Theta)(|\omega_n\rangle\langle\omega_n|) = \frac{1}{n} \sum_{i,j=1}^{n} |i\rangle|j\rangle \otimes |j\rangle|i\rangle = \frac{1}{n}\mathbb{F}.$$

If we can show that $\mathbb{F}$ is not positive, then we have found a trace–preserving and even unital operator which is not completely positive. In fact, we observe that $\mathbb{F}^2 = \mathbb{1} \otimes \mathbb{1}$ and hence $\sigma(\mathbb{F}) \subset \{\pm 1\}$. It is easy to check that $|\psi\rangle \otimes |\varphi\rangle - |\varphi\rangle \otimes |\psi\rangle$ is an eigenvector corresponding to the eigenvalue -1.

Observe that a linear map $\mathscr{N} : \mathfrak{M}_n \to \mathfrak{M}_m$ always induces another linear map acting on $\mathfrak{M}_p(\mathfrak{M}_n) = \mathfrak{M}_p \otimes \mathfrak{M}_n$ by

$$\mathbb{1}_p \otimes \mathscr{N} : \mathfrak{M}_p(\mathfrak{M}_n) \to \mathfrak{M}_p(\mathfrak{M}_m)$$
$$A = (A_{ij})_{1 \leqslant i,j \leqslant p} \mapsto \big(\mathscr{N}(A_{ij})\big)_{1 \leqslant i,j \leqslant p}$$

where we think of A as block matrix with entries $A_{ij} \in \mathfrak{M}_n$.
This interpretation turns out to be rather convenient in the course of the proof of the following theorem due to Choi. By this theorem we acquire a representation of completely positive maps which is frequently used throughout the whole field of quantum information theory.

Theorem 1.17 (Choi's Theorem [Cho75]). *A linear map $\mathscr{N} : \mathfrak{M}_n \to \mathfrak{M}_m$ is completely positive if and only if there exist $V_i \in \mathbb{C}^{n \times m}$, $1 \leqslant i \leqslant nm$, such that*

$$\mathscr{N}(A) = \sum_{i=1}^{nm} V_i^\dagger A V_i. \tag{1.3}$$

*The matrices V_i are called **Kraus operators**.*

Proof. The argument presented in the following is essentially Choi's original one. Nevertheless, there are several other proofs known, e.g. the approach presented by Wilde in [Wil17] and the proof based on the Choi-Jamiolkowski isomorphism in [Wol12].

We start our proof with establishing that every completely positive map admits a decomposition of the form (1.3). Let e_i denote the i–th vector of the standard basis of the $\mathbb{C}^n$. Clearly,

$$\operatorname{span}\left\{e_i e_j^\dagger \,\middle|\, 1 \leqslant i,j \leqslant n\right\} = \mathfrak{M}_n$$

and thus it suffices to prove the assertion for these rank one matrices. In order to deal with block vectors and matrices, we use the fact that $\mathbb{C}^{nm} \cong \bigoplus_{i=1}^n \mathbb{C}^m$. Hence, a vector $v \in \mathbb{C}^{nm}$ reads

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

with $v_1, \ldots, v_n \in \mathbb{C}^m$. Defining $V^\dagger = (v_1, \ldots, v_n) \in \mathbb{C}^{m \times n}$ we find

$$V^\dagger e_i e_j^\dagger V = \left[(v_1, \ldots, v_n)\, e_i \right] \left[(v_1, \ldots, v_n)\, e_j \right]^\dagger = v_i v_j^\dagger.$$

The block matrix

$$E = \begin{pmatrix} e_1 e_1^\dagger & \cdots & e_1 e_n^\dagger \\ \vdots & \ddots & \vdots \\ e_n e_1^\dagger & \cdots & e_n e_n^\dagger \end{pmatrix}$$

is a positive element of $\mathfrak{M}_n(\mathfrak{M}_n)$ since $\operatorname{rank}(E) = 1$ and $\operatorname{tr}[E] \geqslant 0$. Hence, $\left(\mathcal{N}(e_i e_j^\dagger)\right)_{1 \leqslant i,j \leqslant n} \in \mathfrak{M}_n(\mathfrak{M}_k)$ is positive, too. Eventually by the spectral theorem, we have

$$\left(\mathcal{N}(e_i e_j^\dagger)\right)_{1 \leqslant i,j \leqslant n} = \sum_{k=1}^{nm} v_k v_k^\dagger = \sum_{k=1}^{nm} \left(V_k^\dagger e_i e_j^\dagger V_k\right)_{1 \leqslant i,j \leqslant n}$$

where some of the vectors $\{v_k\}_{1 \leqslant k \leqslant nm} \subset \mathbb{C}^{nm}$ may vanish. Thus, we find for the entries of the block matrix

$$\mathcal{N}(e_i e_j^\dagger) = \sum_{k=1}^{nm} V_k^\dagger e_i e_j^\dagger V_k$$

which proves the assertion.
The remaining implication is established by a simple computation. Let $p \in \mathbb{N}$ and $x \in \mathbb{C}^{pm} \setminus \{0\}$, then

$$\begin{aligned}
&\left\langle x \,\middle|\, (\mathbb{1}_p \otimes \mathcal{N})\left((A_{ij})_{1 \leqslant i,j \leqslant p}\right) \,\middle|\, x \right\rangle \\
&= \left\langle x \,\middle|\, \left(\mathcal{N}(A_{ij})\right)_{1 \leqslant i,j \leqslant p} \,\middle|\, x \right\rangle \\
&= \sum_{k=1}^{nm} \left\langle x \,\middle|\, (V_k^\dagger A_{ij} V_k)_{1 \leqslant i,j \leqslant p} \,\middle|\, x \right\rangle \\
&= \sum_{k=1}^{nm} \left\langle x \,\middle|\, (\mathbb{1}_p \otimes V_k)^\dagger A_{ij} (\mathbb{1}_p \otimes V_k) \,\middle|\, x \right\rangle \\
&= \sum_{k=1}^{nm} \left\langle y_k \,\middle|\, A_{ij} \,\middle|\, y_k \right\rangle \\
&\geqslant 0
\end{aligned}$$

where we abbreviate $y_k = (\mathbb{1}_p \otimes V_k)x$ and use the positivity of A_{ij}. Since this calculation holds true for any $p \in \mathbb{N}$ the complete positivity of $\mathcal{N}$ follows. □

Notice that we have actually proven another statement *en passant*, namely, if $\mathcal{N} : \mathfrak{M}_n \to \mathfrak{M}_m$ is n–positive, that is, $(\mathbb{1}_n \otimes \mathcal{N})(A) \geqslant 0$ for any $A \in \mathfrak{M}_n(M_n)$, then it is completely positive. Moreover, the Kraus operators are not unique, especially if some of the v_k vanish. If we impose the constraint that the set $\{v_k\}_{1 \leqslant k \leqslant nm}$ does not contain the zero vector, then the Kraus operators are at least unique up to a unitary conjugation.
Furthermore, theorem 1.17 implies that the map $\mathcal{N}$ is completely positive and unital if and only if $\sum_i V_i V_i^\dagger = \mathbb{1}$. Since the trace is invariant under cyclic permutations we additionally find that $\mathcal{N}$ is in addition trace–preserving if and only if the corresponding Kraus operators satisfy $\sum_i V_i^\dagger V_i = \mathbb{1}$.
The last theorem, we want to present in order to characterize completely positive maps, is Stinespring's dilation theorem. It can be seen as generalization of the purification of a quantum state (theorem 1.6), roughly spoken «going to the church of the larger Hilbert space» for quantum channels.

Theorem 1.18 (Stinespring's Dilation). *Let $\mathcal{N} : \mathscr{D}(\mathscr{H}_A) \to \mathscr{D}(\mathscr{H}_B)$ be a quantum channel. Then there exist an ancillary system $\mathscr{H}_E$, a vector $|\eta\rangle \in \mathscr{H}_E$ and a **Stinespring's dilation** $U : \mathscr{H}_A \to \mathscr{H}_B \otimes \mathscr{H}_E$ such that*

$$\mathcal{N}(\rho_A) = \mathrm{tr}_E \left[U \rho_A U^\dagger \right]$$

for all $\rho_A \in \mathscr{D}(\mathscr{H}_A)$. Moreover, the operators U are unitary, that is, $U^\dagger U = \mathbb{1}_A$.

Proof. From Choi's theorem 1.17, we have

$$\mathcal{N}(\rho_A) = \sum_i V_i \rho_A V_i^\dagger .$$

Defining $U = \sum_i V_i \otimes |i\rangle$ yields

$$U \rho_A U^\dagger = \left(\sum_i V_i \otimes |i\rangle \right) \rho_A \left(\sum_i V_i^\dagger \otimes \langle i| \right) = \sum_{i,j} V_i \rho_A V_j^\dagger \otimes |i\rangle \langle j| .$$

Using the linearity of the partial trace proves the claim. The unitarity of the Stinespring's dilation follows from the fact that $\sum_i V_i^\dagger V_i = \mathbb{1}_A$ (*cf.* the discussion subsequent to theorem 1.17). □

Actually, there is a more common form of Stinespring's dilation which can be easily deduced from the theorem above. Moreover, this version will be needed to prove the monotonicity of the fidelity in § 1.2.2.

Theorem 1.19 (Open System Representation). *Let $\mathcal{N} : \mathcal{H}_A \to \mathcal{H}_B$ be a quantum channel. Then there exist an ancillary system $\mathcal{H}_E$, a unitary operator $\tilde{U} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E)$ and a unit vector $|\eta\rangle \in \mathcal{H}_B \otimes \mathcal{H}_E$ such that*

$$\mathcal{N}(\rho_A) = \mathrm{tr}_{AE}\left[\tilde{U}\left(\rho_A \otimes |\eta\rangle\langle\eta|\right)\tilde{U}^\dagger\right].$$

Proof. By choosing the ancillary Hilbert space $\mathcal{H}_E$ large enough, namely $\dim(\mathcal{H}_E) = \dim\mathcal{H}_A \dim(\mathcal{H}_B)$, we can write the Stinespring's dilation U from theorem 1.18 as a unitary action on a suitable tensor product, that is, $U = \tilde{U}(\mathbb{1}_A \otimes |\eta\rangle)$ for some $|\eta\rangle \in \mathcal{H}_B \otimes \mathcal{H}_E$. This proves the theorem. □

1.2 Distance Measures for Quantum States

1.2.1 Schatten p-Norms

We want to introduce a metric on the set of bounded operators $\mathcal{B}(\mathcal{H})$ acting on the Hilbert space $\mathcal{B}(\mathcal{H})$. Since we only consider finite-dimensional Hilbert spaces establishing a norm on $\mathcal{B}(\mathcal{H}) \cong \mathbb{C}^{\dim(\mathcal{H})^2}$ is a quite arbitrary act. Nevertheless, the following norm is a wise choice since it has a quite natural operational interpretation as we will see below.

Definition 1.20. For $A \in \mathcal{B}(\mathcal{H})$ we define the **trace norm** via

$$\|A\|_1 = \mathrm{tr}[|A|] = \mathrm{tr}\left[\sqrt{A^\dagger A}\right].$$

The mapping

$$D : \mathcal{B}(\mathcal{H}) \times \mathcal{B}(\mathcal{H}) \to [0,\infty)$$
$$(A,B) \mapsto \frac{1}{2}\|A-B\|_1$$

defines a metric on $\mathcal{B}(\mathcal{H})$ which is called the **trace distance** of A and B. In particular, we have $D(\rho,\sigma) \in [0,1]$ for all density operators $\rho,\sigma \in \mathcal{D}(\mathcal{H})$.

We should convince ourselves that $\|\cdot\|_1$ defines indeed a norm. Clearly, for any $\alpha \in \mathbb{C}$ and $A \in \mathcal{B}(\mathcal{H})$ we get $\|\alpha A\|_1 = |\alpha|\|A\|_1$ and $\|A\| \geqslant 0$. The fact that $\|A\|_1 = 0$ if and only if $A = 0$ follows by noting that $\mathrm{tr}[|A|] = \sum_i \sigma_i$ where $\sigma_i \geqslant 0$ are the singular values of A. Since the singular values vanish if and only if the matrix is the zero matrix, we have shown that the first two norm properties hold true. The triangle inequality demands a little more effort. The key to the proof is lemma A.7 which implies

$$\|A\|_1 = \max_{U \in \mathcal{U}(d)} |\mathrm{tr}[AU]|$$

and thus

$$\begin{aligned}\|A+B\|_1 &= \max_{U\in\mathscr{U}(d)} |\operatorname{tr}[AU]+\operatorname{tr}[BU]| \\ &\leqslant \max_{U\in\mathscr{U}(d)} |\operatorname{tr}[AU]| + \max_{U\in\mathscr{U}(d)} |\operatorname{tr}[BU]| = \|A\|_1 + \|B\|_1.\end{aligned}$$

Indeed, the trace norm is a special case of the **Schatten p–norms** $\|A\|_p = \operatorname{tr}[|A|^p]^{1/p}$ for $p \in [1,\infty)$. As we have already seen in the case $p=1$, there is a very convenient connection between the Schatten p–norm of a matrix A and the Euclidean p–norm of the vector formed by the singular values of A.

Proposition 1.21. *For $p \in [1,\infty)$ the Schatten p–norm of a matrix $A \in \mathfrak{M}_n$ is given by*

$$\|A\|_p = \left(\sum_{i=1}^n \sigma_i(A)^p\right)^{1/p}$$

where $\{\sigma_1(A),\dots,\sigma_n(A)\} \subset [0,\infty)$ is the set of the singular values of A. Moreover, by taking $p\to\infty$, the Schatten p–norm converges to operator norm:

$$\|A\|_\infty := \lim_{p\to\infty}\|A\|_p = \sup_{v\neq 0}\frac{|Av|}{|v|} = \|A\|_{\mathrm{op}}.$$

Proof. By writing the singular value decomposition $A = U\Sigma V^\dagger$ in the bra–ket formalism, we obtain

$$A = \sum_{i=1}^r \sigma_i |u_i\rangle\langle v_i|$$

with $r = \operatorname{rank}(A)$ and $\{|u_i\rangle\}_{i=1}^n, \{|v_i\rangle\}_{i=1}^n \subset \mathbb{C}^n$ onbs. Note that in this representation the singular values $\sigma_1,\dots,\sigma_r$ are strictly positive. Moreover, a straight–forward computation yields

$$\begin{aligned}A^\dagger A &= \left(\sum_{i=1}^r \sigma_i |v_i\rangle\langle u_i|\right)\left(\sum_{j=1}^r \sigma_j |u_j\rangle\langle v_j|\right) \\ &= \sum_{i,j=1}^r \sigma_i\sigma_j |v_i\rangle\langle u_i|u_j\rangle\langle v_j| \\ &= \sum_{i=1}^r \sigma_i^2 |v_i\rangle\langle v_i|. \end{aligned} \tag{1.4}$$

Consequently, we conclude

$$\|A\|_p^p = \operatorname{tr}\left[|A|^p\right] = \operatorname{tr}\left[\sqrt{A^\dagger A}^p\right] = \operatorname{tr}\left[\sum_{i=1}^r \sqrt{\sigma_i^2}^p |v_i\rangle\langle v_i|\right] = \sum_{i=1}^r \sigma_i^p = \sum_{i=1}^n \sigma_i^p$$

since $\sigma_{r+1} = \cdots = \sigma_n = 0$.
The convergence behavior of the Schatten p–norm follows immediately from convergence of the Euclidean p–norm to the maximum norm and by noting that $\|A\|_{\text{op}} = \max\{\sigma_1(A), \ldots, \sigma_n(A)\}$. The latter statement is proven as follows. From equation (1.4), we easily deduce

$$\|A\|_{\text{op}}^2 = \sup_{v=1} |Av|^2 = \sup_{v=1} v^\dagger A^\dagger A v \leqslant (\max\{\sigma_1(A), \ldots, \sigma_n(A)\})^2 .$$

Clearly, equality is attained by picking v the normalized eigenvector corresponding to the largest eigenvalue of $A^\dagger A$. □

The Schatten 2–norm is called the **Hilbert–Schmidt norm** and is induced by the **Hilbert–Schmidt inner product**

$$\langle A \,|\, B \rangle_{HS} = \operatorname{tr}\left[A^\dagger B\right].$$

Actually, the Hilbert–Schmidt inner product is a quite natural choice of a scalar product on $\mathfrak{M}_n$. We can think of a complex $n \times n$ A matrix as column vector living in $\mathbb{C}^{n^2}$ endowed with the standard inner product. Then we find

$$\left\langle \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{nn} \end{pmatrix} \middle| \begin{pmatrix} b_{11} \\ b_{12} \\ \vdots \\ b_{nn} \end{pmatrix} \right\rangle_{\mathbb{C}^{n^2}} = \sum_{i,j=1}^{n} a_{ij}^* b_{ij} = \operatorname{tr}\left[A^\dagger B\right] = \langle A \,|\, B \rangle_{HS}$$

and thus the standard dot product of these associated vectors is exactly the Hilbert–Schmidt inner product of the matrices. The last two facts about the Schatten p–norms we want to state and prove are Hölder's inequality and the triangle inequality.

Theorem 1.22 (Hölder's Inequality). *Let $A, B \in \mathfrak{M}_n$ and $p, q \in [1, \infty]$ such that $p^{-1} + q^{-1} = 1$. In the case $p = 1$, we set $q = \infty$ and vice versa. Then the following bound on the trace norm holds:*

$$\|AB\|_1 \leqslant \|A\|_p \|B\|_q .$$

Proof. Unfortunately, the standard technique using Young's inequality is not applicable in this case. Our proof relies mainly on the singular value decomposition.
First, note that it suffices to prove $|\operatorname{tr}[AB]| \leqslant \|A\|_p \|B\|_q$ since the polar decomposition $AB = U|AB|$ implies

$$\|AB\|_1 = \operatorname{tr}[|AB|] = \operatorname{tr}[U^\dagger AB] \leqslant \|U^\dagger A\|_p \|B\|_q = \|A\|_p \|B\|_q .$$

In fact, the unitary invariance of the Schatten p–norm follows immediately from $|U^\dagger A| = \sqrt{A^\dagger U U^\dagger A} = |A|$. Consider the singular value decompositions

$$A = \sum_{i=1}^{n} \sigma_i(A) \left| u_i^{(A)} \right\rangle \left\langle v_i^{(A)} \right|, \qquad B = \sum_{j=1}^{n} \sigma_j(B) \left| u_j^{(B)} \right\rangle \left\langle v_j^{(B)} \right|$$

and note that the trace of their product can be computed via

$$\operatorname{tr}[AB] = \sum_{i=1}^{n} \left\langle u_i^{(A)} \middle| AB \middle| u_i^{(A)} \right\rangle = \sum_{i,j=1}^{n} \sigma_i(A)\sigma_j(B) \underbrace{\left\langle v_i^{(A)} \middle| u_j^{(B)} \right\rangle \left\langle v_j^{(B)} \middle| u_i^{(A)} \right\rangle}_{=:\xi_{ij}}.$$

Moreover by the ordinary Cauchy–Schwarz inequality for sums, we find

$$\sum_{i=1}^{n} |\xi_{ij}| \leqslant \underbrace{\sqrt{\sum_{i=1}^{n} \left|\left\langle v_i^{(A)} \middle| u_j^{(B)} \right\rangle\right|^2}}_{=\left|u_j^{(B)}\right|} \underbrace{\sqrt{\sum_{i=1}^{n} \left|\left\langle v_j^{(B)} \middle| u_i^{(A)} \right\rangle\right|^2}}_{=\left|v_j^{(B)}\right|} = 1.$$

Clearly, the same calculation shows $\sum_{j=1}^{n} |\xi_{ij}| \leqslant 1$. Let us now assume that $p, q \in (1,\infty)$. Due to the triangle inequality and Hölder's inequality for sums, we conclude

$$\begin{aligned}
|\operatorname{tr}[AB]| &= \left| \sum_{i,j=1}^{n} \sigma_i(A)\sigma_j(B)\xi_{ij} \right| \\
&\leqslant \sum_{i,j=1}^{n} \sigma_i(A) |\xi_{ij}|^{1/p} \sigma_j(B) |\xi_{ij}|^{1/q} \\
&\leqslant \left(\sum_{i,j=1}^{n} \left(\sigma_i(A) |\xi_{ij}|^{1/p} \right)^p \right)^{1/p} \left(\sum_{i,j=1}^{n} \left(\sigma_j(B) |\xi_{ij}|^{1/q} \right)^q \right)^{1/q} \\
&= \left(\sum_{i=1}^{n} \sigma_i(A)^p \sum_{j=1}^{n} |\xi_{ij}| \right)^{1/p} \left(\sum_{j=1}^{n} \sigma_j(B)^q \sum_{i=1}^{n} |\xi_{ij}| \right)^{1/q} \\
&\leqslant \|A\|_p \|B\|_q
\end{aligned}$$

In the case $p = 1, q = \infty$ or vice versa, we find

$$\begin{aligned}
|\operatorname{tr}[AB]| &\leqslant \sum_{i,j=1}^{n} \sigma_i(A)\sigma_j(B) |\xi_{ij}| \\
&\leqslant \left(\max_{1 \leqslant \ell \leqslant n} \sigma_\ell(B) \right) \sum_{i=1}^{n} \sigma_i(A) \sum_{j=1}^{n} |\xi_{ij}| \leqslant \|A\|_1 \|B\|_{\mathrm{op}}.
\end{aligned}$$

□

In order to derive the triangle inequality of the Schatten p–norm, we present the following lemma which reduces the problem to the triangle inequality of the $\|\cdot\|_1$–norm proven on page 13.

Lemma 1.23. *For $p \in (1,\infty)$ and q the Hölder conjugated exponent of p, the Schatten p–norm of $A \in \mathfrak{M}_n$ is given by*

$$\|A\|_p = \max_{\|B\|_q = 1} \|AB\|_1.$$

Proof. The case $\|A\|_p = 0$ is trivial and hence, let $A \neq 0$. For $\|B\|_q = 1$, Hölder's inequality implies $\|AB\|_1 \leqslant \|A\|_p$. Note that equality is clearly attained if $B = \left(|A|/\|A\|_p\right)^{p-1}$:

$$\|AB\|_1 = \frac{1}{\|A\|_p^{p-1}} \operatorname{tr}\left[\sqrt{|A|^{2p}}\right] = \|A\|_p.$$

Moreover, this matrix B fulfills $\|B\|_q = 1$ since

$$\|B\|_q = \left(\operatorname{tr}\left[|B|^q\right]\right)^{1/q} = \frac{1}{\|A\|_p^{p-1}} \left(\operatorname{tr}\left[|A|^{(p-1)q}\right]\right)^{1/q} = \frac{\left(\|A\|_p^p\right)^{1/q}}{\|A\|_p^{p-1}} = 1$$

and is hence a valid candidate for the maximum in the assertion. □

Eventually, the triangle inequality of the Schatten p–norms ($1 < p < \infty$) follows from a straight–forward calculation based on lemma 1.23:

$$\begin{aligned}\|A+B\|_p &= \max_{\|C\|_q=1} \|(A+B)C\|_1 \\ &\leqslant \max_{\|C\|_q=1} \|AC\|_1 + \max_{\|C\|_q=1} \|BC\|_1 \\ &= \|A\|_p + \|B\|_p.\end{aligned}$$

Observe that the triangle inequality of the operator norm follows immediately from the triangle inequality of the Euclidean norm.

After this mathematical excursion, we want to give an operational interpretation of the trace norm. In order to make the argument more vivid, we present the following thought experiment. Let us consider an apparatus which produces two different states ρ_1 and ρ_2 with a–priori probabilities p_1 and $p_2 = 1 - p_1$. Furthermore, we have a measuring instrument with two light bulbs. This analyzer tries to identify the incoming state and will light up the corresponding bulb. We are now interested in the success probability of our experiment, i.e. the probability of seeing light bulb one lighting up given the incoming state was ρ_1 or vice versa. For a pictorial illustration of the experiment, we refer to figure 1.2. Carrying out the calculation which mainly relies on Born's rule (the interested reader might find further details in appendix A.1.2), it turns out that the success probability reads

$$\mathbb{P}_{\text{success}} = \frac{1}{2}\left(1 + \|p_1\rho_1 - p_2\rho_2\|_1\right).$$

This result is also known as **Helstrom's formula of ambiguous state discrimination**. Thus, the trace norm represents how much better than random guessing our apparatus can perform.

As we see by the following theorem, there is no quantum operation which can enhance the distinguishability of two quantum states.

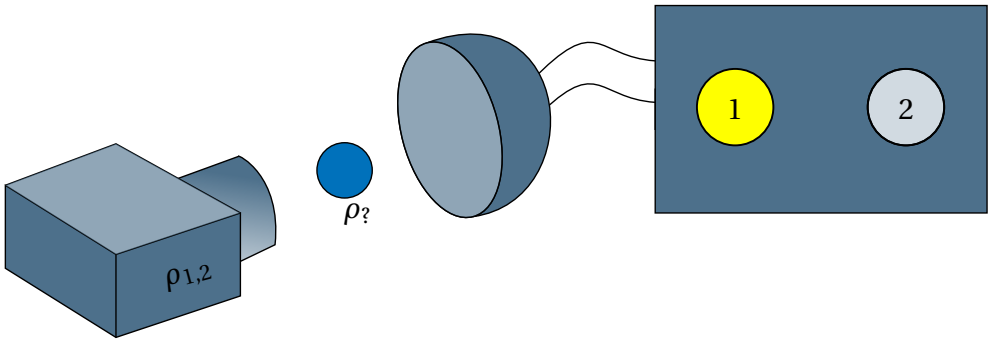


Figure 1.2: The setting of ambiguous state discrimination. The apparatus on the left produces the quantum states ρ_1 and ρ_2 with fixed a–priori probabilities p_1 and p_2. The unknown quantum state $\rho_?$ is analyzed and the light bulb corresponding to the state identified by the analyzer lights up. As theorem A.6 shows, this identification is defective in general. Figure adapted from [HZ11].

As we see by the following theorem, there is no quantum operation which can enhance the distinguishability of two quantum states.

Theorem 1.24. *Let $\rho, \sigma \in \mathscr{D}(\mathscr{H}_A)$ be two density matrices on the d_A–dimensional Hilbert space $\mathscr{H}_A$. Furthermore $\mathcal{N} : \mathscr{D}(\mathscr{H}_A) \to \mathscr{D}(\mathscr{H}_B)$ denotes quantum operation. Then the following inequality for the trace distance holds:*

$$D(\rho, \sigma) \geqslant D\big(\mathcal{N}(\rho), \mathcal{N}(\sigma)\big).$$

Proof. Since $\rho - \sigma$ is Hermitian, its spectral decomposition exists and reads

$$\rho - \sigma = \sum_{i=1}^{d_A} \lambda_i |i\rangle\langle i| = \sum_{i:\lambda \geqslant 0} \lambda_i |i\rangle\langle i| - \sum_{i:\lambda_i < 0} |\lambda_i| |i\rangle\langle i| =: P - N \tag{1.5}$$

where $\{|i\rangle\}_{i=1}^{d_A}$ forms an onb of the d_A–dimensional Hilbert space $\mathscr{H}$. Equation (1.5) immediately implies that the trace of the absolute value of $\rho - \sigma$ reads

$$\operatorname{tr}[|\rho - \sigma|] = \sum_{i:\lambda_i \geqslant 0} \lambda_i + \sum_{i:\lambda_i < 0} |\lambda_i| = 2 \sum_{i:\lambda_i \geqslant 0} \lambda_i.$$

Based on the spectral decomposition of $\mathcal{N}(\rho - \sigma)$

$$\mathcal{N}(\rho - \sigma) = \sum_{j=1}^{d_B} \eta_j |j\rangle\langle j|,$$

we define the projector onto the non–negative eigenvalues of $\mathcal{N}(\rho - \sigma)$ by

$$\Pi := \sum_{j:\eta_j \geqslant 0} |j\rangle\langle j|.$$

From here only a small effort is needed to finish the proof. The trace–preserving property of $\mathcal{N}$ together with the Cauchy Schwarz inequality implies

$$\operatorname{tr}\left[|\rho-\sigma|\right] = 2\operatorname{tr}\left[\mathcal{N}(P)\right] \geqslant 2\operatorname{tr}\left[\Pi\mathcal{N}(P)\right] \geqslant 2\operatorname{tr}\left[\Pi\mathcal{N}(P-N)\right] = \operatorname{tr}\left[\left|\mathcal{N}(\rho-\sigma)\right|\right].$$

□

1.2.2 The Fidelity of Quantum States

We continue our brief introduction into some of the most important notions of quantum theory needed in this thesis by introducing the fidelity of two density operators.

Definition 1.25. The mapping

$$F : \mathcal{D}(\mathcal{H}) \times \mathcal{D}(\mathcal{H}) \to [0,\infty)$$
$$(\rho,\sigma) \mapsto \|\sqrt{\sigma}\sqrt{\rho}\|_1 = \operatorname{tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]$$

defines the **fidelity** of the two quantum states ρ and σ.

In order to make the notion of fidelity more vivid, we firstly note that for a pure state $\rho = |\psi\rangle\langle\psi|$ the fidelity reads $F(\rho,\sigma) = \sqrt{\langle\psi|\sigma|\psi\rangle}$. Hence, we may think of the fidelity as a quantity which represents the overlap of the states ρ and σ. In 1976, Uhlmann proved that this interpretation remains valid even when considering more general density operators.

Theorem 1.26 (Uhlmann's Theorem [Uhl76]). *Let $\rho,\sigma \in \mathcal{D}(\mathcal{H})$ be two density operators. Then their fidelity is given by*

$$F(\rho,\sigma) = \max_{|\psi\rangle,|\varphi\rangle} |\langle\psi|\varphi\rangle|$$

where the maximum is taken over all purifications $|\psi\rangle$ of ρ and $|\varphi\rangle$ of σ into $\mathcal{H}\otimes\mathcal{H}$.

For the sake of completeness, the interested reader may find a proof of theorem 1.26 in appendix A.2. Uhlmann's theorem will be the main ingredient in proving some important properties of the fidelity in the following.

Proposition 1.27. *Suppose ρ and σ denote two density operators. Then the fidelity between them fulfills:*

(i) $F(\rho,\sigma) \in [0,1]$ and $F(\rho,\sigma) = 1$ if and only if $\rho = \sigma$.

(ii) $F(\rho,\sigma) = F(\sigma,\rho)$.

(iii) *Applying a similarity transform on the density operators* ρ *and* σ *does not change their fidelity, i.e.* $F(U\rho U^\dagger, U\sigma U^\dagger) = F(\rho,\sigma)$ *for any unitary* U.

Proof. The first claim follows immediately from the application of the Cauchy Schwarz inequality onto theorem 1.26. Recalling that equality in Cauchy Schwarz occurs if and only if the vectors $|\varphi\rangle$ and $|\psi\rangle$ are collinear establishes the second assertion under (i).
Clearly, point (ii) does not require any proof.
Finally, we note that the spectral decomposition of a matrix $A \geqslant 0$ implies $\sqrt{UAU^\dagger} = U\sqrt{A}U^\dagger$ from which we easily deduce (iii). □

In analogy to the trace distance, we try to establish a bound on the fidelity of two states to which the same quantum operation was applied in terms of their original fidelity. Contrary to the preceding results, the fidelity never decreases; that means the overlap of the two density operators increases or at least remains the same.

Theorem 1.28. *Let* $\mathcal{N}$ *be a trace–preserving quantum operation acting on the density operators* ρ *and* σ. *Then we can lower bound the fidelity between* $\mathcal{N}(\rho)$ *and* $\mathcal{N}(\sigma)$ *by*

$$F\big(\mathcal{N}(\rho),\mathcal{N}(\sigma)\big) \geqslant F(\rho,\sigma).$$

Especially, tracing out an ancillary system never shrinks the fidelity of the two quantum states, that is, $F(\rho_{AB},\sigma_{AB}) \leqslant F(\rho_A,\sigma_A)$.

Proof. We want to apply Uhlmann's theorem in some fruitful manner. Thus, we consider the tensor product space $\mathscr{H}\otimes\mathscr{H}_A$ with the ancillary system $\mathscr{H}_A$. Moreover, let $|\varphi\rangle$ and $|\psi\rangle$ denote the maximizing purifications from Uhlmann's theorem of ρ and σ respectively, i.e. the fidelity reads $F(\rho,\sigma) = |\langle\varphi|\psi\rangle|$. According to Stinespring's dilation in the open system representation (theorem 1.19), there are another ancillary Hilbert space $\mathscr{H}_E$, a pure state $|\eta\rangle\langle\eta| \in \mathscr{D}(\mathscr{H}_E)$ and a unitary $U \in \mathscr{B}(\mathscr{H}\otimes\mathscr{H}_E)$ such that

$$\mathcal{N}(\rho) = \operatorname{tr}_E\Big[U\big(\rho\otimes|\eta\rangle\langle\eta|\big)U^\dagger\Big]$$

holds for all density operators ρ. Hence, $U(|\varphi\rangle\otimes|\eta\rangle)$ and $U(|\psi\rangle\otimes|\eta\rangle)$ denote purifications of $\mathcal{N}(\rho)$ and $\mathcal{N}(\sigma)$ into $\mathscr{H}\otimes\mathscr{H}_A\otimes\mathscr{H}_E$. Using Uhlmann's theorem, the following computation establishes the claim:

$$F\big(\mathcal{N}(\rho),\mathcal{N}(\sigma)\big) \geqslant \Big|\big(\langle\varphi|\otimes\langle\eta|\big)U^\dagger U\big(|\psi\rangle\otimes|\eta\rangle\big)\Big| = |\langle\varphi|\psi\rangle| = F(\rho,\sigma).$$

□

We want to use the last lines of this section to investigate the relationship between fidelity and trace distance. As we have already seen for pure states $\rho = |\varphi\rangle\langle\varphi|$ and $\sigma = |\psi\rangle\langle\psi|$, the fidelity reads $F(\rho,\sigma) = |\langle\varphi|\psi\rangle|$. If we apply Gram-Schmidt's procedure on the two Hilbert space vectors $|\varphi\rangle$ and $|\psi\rangle$, we can write them as

$$\begin{aligned} |\varphi\rangle &= |1\rangle \\ |\psi\rangle &= \cos\theta\,|1\rangle + \sin\theta\,|2\rangle \end{aligned}$$

where $|1\rangle$ and $|2\rangle$ denote to orthonormal vectors. There is no loss of generality if we choose real coefficients since a complex phase can be absorbed in the unit vector $|2\rangle$. Hence, we find $F(\rho,\sigma) = |\cos\theta|$. Furthermore, the density matrix σ is given by

$$\sigma = |\psi\rangle\langle\psi| = \cos^2\theta\,|1\rangle\langle 1| + \sin\theta\cos\theta\,|2\rangle\langle 1| + \cos\theta\sin\theta\,|1\rangle\langle 2| + \sin^2\theta\,|2\rangle\langle 2|.$$

The difference of the two density matrices has the following matrix representation (in the basis $\{|1\rangle, |2\rangle\}$)

$$\rho - \sigma = \begin{pmatrix} 1-\cos^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & -\sin^2\theta \end{pmatrix}$$

which has eigenvalues $\pm|\sin\theta|$. Therefore, we obtain

$$D(\rho,\sigma) = \frac{1}{2}\operatorname{tr}\left[\left|\begin{pmatrix} 1-\cos^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & -\sin^2\theta \end{pmatrix}\right|\right] = |\sin\theta| = \sqrt{1-F(\rho,\sigma)^2}.$$

For general density matrices ρ and σ we consider their purifications $|\varphi\rangle$ and $|\psi\rangle$. Using both, Uhlmann's theorem and the monotonicity of the trace distance (theorem 1.24), we find

$$D(\rho,\sigma) \leqslant \sqrt{1-F(\rho,\sigma)^2}.$$

By this computation we have proven the following proposition:

Proposition 1.29 (First Fuchs–van de Graaf Inequality). *For $\rho,\sigma \in \mathscr{D}(\mathscr{H})$ the trace distance is bounded from above by*

$$D(\rho,\sigma) \leqslant \sqrt{1-F(\rho,\sigma)^2}.$$

Chapter 2

From Classical to Quantum Information Theory

There is a whole bunch of different quantities in the field of information theory. We restrict to the most important notions for which we discuss their quantum analogues in section § 2.2. The interested reader may find a comprehensive survey of the classical quantities in the very accessible book by Cover and Thomas [CT12]. For further information about the quantum realm we refer to the standard reference of the subject, namely the excellent textbook of Nielsen and Chuang [NC10].

2.1 Quantities of Classical Information Theory

Arguably the most famous quantity in information theory is the **entropy** H introduced by Shannon in 1948. Actually, the name originated in statistical mechanics, more precisely, it was adopted from Boltzmann's entropy representing the logarithm of the number of microstates of a system. We only consider the case of a random variable X taking values in the finite set S. The generalization to continuous random variables is pretty obvious. The reader has to substitute the sums in the subsequent definitions by an integral w.r.t. the distribution of the random variable solely.
The entropy of X is given by

$$\begin{aligned} H(X) &= -\sum_{x\in S} \mathbb{P}(X=x)\log\left[\mathbb{P}(X=x)\right] \\ &= -\sum_{x\in S} p(x)\log\left[p(x)\right] = \mathbb{E}\left[\log\frac{1}{p(X)}\right] \geqslant 0 \end{aligned}$$

where $p : S \to [0,1]$, $p(x) = \mathbb{P}(X = x)$, denotes the probability mass function of X. Note that in this thesis log indicates the logarithm to base 2. We use the convention that $0\log 0 = 0$ which seems natural by noting that $\lim_{x\to 0} x\log x = 0$. In particular, this convention ensures that $X = Y$ a.s. implies $H(X) = H(Y)$.

Example 2.1. Let us compute the entropy of the random variable $X_p \sim \text{Bernoulli}(p)$. We find

$$H(X_p) = -p\log p - (1-p)\log(1-p)$$

which is maximized for $p = 1/2$ (*c.f.* figure 2.1). Hence, we might guess that a random variable $X \sim \text{Uniform}(S)$ maximizes the entropy. To prove this claim, we introduce the **relative entropy** or **Kullback–Leibler divergence** between two random variables $X \sim p(x)$ and $Y \sim q(x)$ taking values in the same alphabet S via

$$D(X\|Y) = D(p\|q) = \sum_{x\in S} p(x)\log\left[\frac{p(x)}{q(x)}\right].$$

We use the conventions that $0\log(0/0) = 0$, $0\log(0/q) = 0$ and $p\log(p/0) = \infty$. The importance of the relative entropy becomes clearer by observing that we have $D(p\|q) \geqslant 0$. In fact, by letting $A = \operatorname{supp} p$, an easy application of Jensen's inequality to the strictly concave function $x \mapsto \log x$ reveals that

$$-D(p\|q) = \sum_{x\in A} p(x)\log\left[\frac{q(x)}{p(x)}\right] \leqslant \log\left[\sum_{x\in A} q(x)\right] \leqslant \log\left[\sum_{x\in S} q(x)\right] = 0. \tag{2.1}$$

From this computation, we can easily deduce the necessary and sufficient condition for $D(p\|q) = 0$. Namely, the equality condition in Jensen's inequality requires $q(x)/p(x) = c = \text{const}$. Observe that

$$\sum_{x\in A} qx = c\sum_{x\in A} p(x) = c$$

and the second bound in (2.1) requires $\sum_{x\in A} q(x) = \sum_{x\in S} q(x) = 1$. Hence, we find $c = 1$ and consequently, $D(p\|q) = 0$ if and only if $p(x) = q(x)$ for all $x \in S$.

Eventually, we conclude that the uniform distribution maximizes the entropy by the following argument. Let $X \sim p(x)$ and $Y \sim \text{Uniform}(S)$. Then we have

$$0 \leqslant D(X\|Y) = \sum_{x\in A} p(x)\log\big[p(x)|S|\big] = \log(|S|) - H(X) = H(Y) - H(X)$$

from which the claim follows.

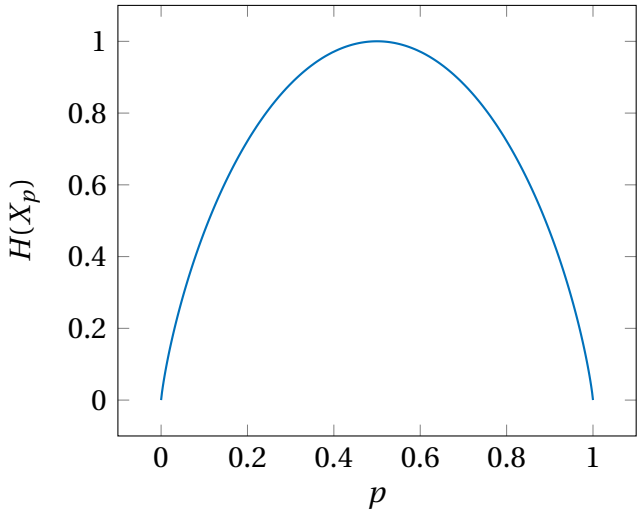


Figure 2.1: The plot depicts the function $[0,1] \ni p \mapsto H(X_p)$ with $X_p \sim \text{Bernouilli}(p)$.

The definition of the entropy can be easily generalized to a pair of random variables $(X,Y) : \omega \to S_X \times S_Y$ with a joint probability mass function $p(x,y)$. We define the **joint entropy** by

$$H(X,Y) = -\sum_{x\in S_X}\sum_{y\in S_Y} p(x,y)\log\big[p(x,y)\big] = \mathbb{E}\left[\log\frac{1}{p(X,Y)}\right].$$

If the random variable X and Y are independent their joint entropy reads $H(X,Y) = H(X)+H(Y)$. In general this does not hold true but one can prove that

$$H(X,Y) = H(X) + H(Y|X) \tag{2.2}$$

where $H(Y|X)$ denotes the **conditional entropy** which is given by

$$H(Y|X) = -\sum_{x\in S_X} p(x)\sum_{y\in S_Y} p(y|x)\log\big[p(y|x)\big] = \sum_{x\in S_X} p(x)H(Y|X=x).$$

In the previous equation we used the suggestive labeling $p(x)$ for the marginal distribution of (X,Y) with respect to X as well as $p(y|x)$ for the condition marginal mass function of Y given X. Note that $H(Y|X) \geqslant 0$ since $H(Y|X=x) \geqslant 0$.

The last quantity we want to introduce is the **mutual information** $I(X:Y)$ of two random variables (X,Y) with joint distribution $p(x,y)$ and marginals $p(x)$ and $p(y)$ respectively. Its definition reads

$$I(X:Y) = \sum_{x\in S_X}\sum_{y\in S_Y} p(x,y)\log\left[\frac{p(x,y)}{p(x)p(y)}\right] = I(Y:X).$$

It can be easily seen that $I(X:Y) = H(X) - H(X|Y)$. By this observation, the definition of the **conditional mutual information** via

$$I(X:Z|Y) = H(X|Y) - H(X|Z,Y)$$

seems quite natural. The connection between the mutual information and its conditional form is described by the following theorem.

Theorem 2.2 (Chain Rule for Mutual Information).

$$I(X_1,\dots,X_n:Y)=\sum_{i=1}^{n} I(X_i:Y|X_{i-1},\dots,X_1)$$

The proof of the theorem is done by straight–forward computation. Further details can be found in [CT12].
Eventually, we want to mention that $I(X:Y)\geqslant 0$ as well as $I(X:Z|Y)\geqslant 0$ which follows easily by writing the (conditional) mutual information in terms of the non–negative relative entropy. In fact, we have

$$\begin{aligned} I(X:Y)&=D\big(p(x,y)\|p(x)p(y)\big),\\ I(X:Z|Y)&=D\big(p(x,z|y)\|p(x|y)p(z|y)\big). \end{aligned}$$

The non–negativity of the conditional mutual information will be of importance in § 2.3.

2.2 Quantities of Quantum Information Theory

After the review of a selection of important quantities from classical information theory, we introduce their quantum counterparts in this section. Since the presented quantities are ubiquitous in the rest of the thesis, we take them into deeper consideration than the classical ones.
The central quantity of quantum information is the **von–Neumann entropy** (or just entropy for short) which generalizes the notion of Shannon entropy to quantum states. By recalling the convention $0\log 0=0$, we define the von–Neumann entropy of a quantum state ρ by

$$S(\rho)=-\operatorname{tr}[\rho\log\rho]=-\sum_{\lambda\in\sigma(\rho)}\lambda\log\lambda.$$

Since the eigenvalues of ρ sum up to one, they represent a discrete probability distribution. Consequently, the von–Neumann entropy of ρ coincides with the Shannon entropy corresponding to the random variable X distributed according to ρ's spectrum. Hence, the entropy is maximized by the **maxillymally mixed state** $\frac{\mathbb{1}}{n}=\frac{1}{n}\sum_{i=1}^{n}|i\rangle\langle i|$ which correspondents to the uniform distribution (*c.f.* example 2.1). Analogously to the classical quantity, we define the **quantum relative entropy** by

$$S(\rho\|\sigma)=\begin{cases}\operatorname{tr}\left[\rho(\log\rho-\log\sigma)\right] & \text{if } \operatorname{supp}\rho\subset\operatorname{supp}\sigma,\\ +\infty & \text{otherwise.}\end{cases}$$

The non–negativity of the relative entropy still holds true as the following theorem shows.

Theorem 2.3 (Klein's Inequality). *Let A, B be positive matrices and $f : (0,\infty) \to \mathbb{R}$ convex and differentiable. Then*

$$\operatorname{tr}\left[f(A) - f(B) - (A-B)f'(B)\right] \geqslant 0.$$

Moreover, if f is strictly convex, equality holds if and only if $A = B$.
In particular: $S(\rho||\sigma) \geqslant 0$ and $S(\rho||\sigma) = 0$ if and only if $\rho = \sigma$.

Proof. Let $t \in [0,1]$ and define

$$\phi(t) = \operatorname{tr}\left[f(B + t(A-B))\right].$$

This function ϕ is convex since the trace is monotone and linear. Furthermore, ϕ is differentiable and $\phi'(t) = \operatorname{tr}[(A-B)f'(B+t(A-B))]$. Let $x, y \in [0,1]$, $y > x$, and $\lambda \in (0,1)$, then the definition of convexity reads

$$\phi(\lambda y + (1-\lambda)x) \leqslant \lambda\phi(y) + (1-\lambda)\phi(x).$$

Simple arithmetics shows that this condition is equivalent to

$$\frac{\phi(x + \lambda(y-x)) - \phi(x)}{\lambda} \leqslant \phi(y) - \phi(x).$$

Finally, we let $y = 1$ and $x = 0$ and take the limit $\lambda \searrow 0$ to conclude

$$\operatorname{tr}[(A-B)f'(B)] = \phi'(0) \leqslant \phi(1) - \phi(0) = \operatorname{tr}[f(A) - f(B)]. \tag{2.3}$$

In the case of a strictly convex function f and $A - B \neq 0$, ϕ is strictly convex, too. Hence, inequality 2.3 is strict.
The assertions about the quantum relative entropy follow by noting that $(0,\infty) \ni x \mapsto x\log x$ is strictly convex. □

The argument presented in the proof above is more general than the original one of Klein [Kle31] (see [NC10] for a modern version).
We collect some of the basic properties of the von–Neumann entropy in the subsequent proposition.

Proposition 2.4.

(i) The von–Neumann entropy $S(\rho)$ is non–negative and it vanishes if and only if ρ is pure.

(ii) Only the maximally mixed state $\frac{\mathbb{1}}{n}$ maximizes the entropy.

(iii) If $\rho_{AB} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is pure, then $S(\rho_A) = S(\rho_B)$.

Proof. The non–negativity of the von–Neumann entropy emerges immediately from the definition. Note that a quantum state ρ is pure if and only if

$\sigma(\rho) \subset \{0,1\}$ and there is only one non–vanishing eigenvalue from which the second assertion under (i) follows.
Imitating the classical proof, point (ii) can be immediately deduced from Klein's inequality (theorem 2.3):

$$0 \leqslant S\left(\rho \middle\| \frac{\mathbb{1}}{n}\right) = -S(\rho) + \log d.$$

Since the eigenvalues of a density operator determine its entropy completely, claim (iii) follows from the discussion subsequent to the Schmidt decomposition (theorem 1.10). □

As a notional convenience, we occasionally use the abbreviation $S(\rho_A) \equiv S(A)$ in the following. Moreover, we define the **joint entropy** in a natural way:

$$S(A,B) = -\operatorname{tr}\left[\rho_{AB} \log \rho_{AB}\right].$$

The three remaining quantities, which we want to consider in our brief discussion, are the **conditional entropy** $S(A|B) = S(A,B) - S(B)$, the **quantum mutual information** $I(A:B) = S(A) + S(B) - S(A,B)$ and the **quantum conditional mutual information** $I(A:C|B)_\rho = S(A,B) + S(B,C) - S(B) - S(A,B,C)$. At this point, we are able to present the first major difference between classical and quantum information theory. From equation (2.2) it immediately follows that $H(X,Y) \geqslant H(X)$ which seems to be a quite apodictic statement since the uncertainty of the random variable X should not exceed the vagueness of the random vector (X,Y). This fact does not hold true in the quantum realm as we show by the following example.

Example 2.5. Let $|\psi\rangle = \frac{1}{\sqrt{2}}(|1\rangle \otimes |1\rangle + |2\rangle \otimes |2\rangle) \in \mathbb{C}^2 \otimes \mathbb{C}^2$ be the EPR state. The state $\rho_{AB} = |\psi\rangle\langle\psi|$ is clearly pure and hence $S(A,B) = 0$. Nevertheless, according to example 1.4 (ii), we obtain the marginal

$$\rho_A = \frac{\mathbb{1}}{2}$$

by tracing out the latter Hilbert space. Consequently, $0 = S(A,B) \ngeqslant S(A) = 1$.

A legitimate question which arises by the consideration of example 2.5 is whether there exists a similar inequality to $H(X) \leqslant H(X,Y)$ in quantum information. The answer to this question is positive as the following theorem shows.

Theorem 2.6. *The von–Neumann entropy satisfies*

(i) subadditivity: $S(A,B) \leqslant S(A) + S(B)$ *with equality if and only if* $\rho_{AB} = \rho_A \otimes \rho_B$,

(ii) triangle inequality (Araki–Lieb inequality): $S(A,B) \geqslant |S(A) - S(B)|$.

Proof. The subadditivity of the von–Neumann entropy is essentially a corollary of Klein's inequality (theorem 2.3). In fact, we find for quantum states ρ and σ that $S(\rho) \leqslant -\operatorname{tr}[\rho \log \sigma]$. Consequently, we set $\rho = \rho_{AB}$ and $\sigma = \rho_A \otimes \rho_B$ to compute

$$\begin{aligned} S(A,B) &\leqslant -\operatorname{tr}\left[\rho_{AB}\left(\log(\rho_A \otimes \rho_B)\right)\right] \\ &= -\operatorname{tr}\left[\rho_{AB}\left(\log \rho_A + \log \rho_B\right)\right] \\ &= -\operatorname{tr}\left[\rho_A \log \rho_A\right] - \operatorname{tr}\left[\rho_B \log \rho_B\right] \\ &= S(A) + S(B). \end{aligned}$$

Note that the penultimate manipulation is due to the definition of the partial trace. Moreover, Klein's inequality implies that equality in subadditivity is attained if and only if ρ_{AB} is a product state.
Inequality (ii) is established by introducing an ancillary system $\mathscr{H}_E$ which purifies the state ρ_{AB}. The already proven subadditivity implies

$$S(A,E) \leqslant S(A) + S(E). \tag{2.4}$$

According to proposition 2.4 (iii) the entropies of the states ρ_{AE} and ρ_B coincide. Moreover, the same holds for ρ_E and ρ_{AB}. Eventually, equation (2.4) reads

$$S(B) - S(A) \leqslant S(A,B).$$

The same argument shows $S(A) - S(B) \leqslant S(A,B)$ and hence the Araki–Lieb inequality. □

The most famous inequality which the von–Neumann entropy satisfies is the strong subadditivity established by Lieb and Ruskai in 1973 [LR73]. There were several other proofs of strong subaddivity published in the subsequent years.

Theorem 2.7 (Strong Subadditivity). *For any state acting on the tripartite system* $\mathscr{H}_A \otimes \mathscr{H}_B \otimes \mathscr{H}_C$ *the inequality*

$$S(A,B,C) + S(B) \leqslant S(A,B) + S(B,C)$$

holds.

We do not want to give a proof of strong subadditivity in this thesis since all known arguments are quite long and usually there is some technicality involved. A proof based on Lieb's and Ruskai's original one is given in [NC10]. Moreover, the reader can find a proof requiring no knowledge beyond linear algebra in the paper of Ruskai [Rus07]. Note that the statement of strong

subadditivity is equivalent to the non–negativity of the quantum conditional mutual information $I(A:C|B)_\rho = S(A,B) + S(B,C) - S(B) - S(A,B,C) \geqslant 0$. This fact can also be easily derived from the recoverability theorem presented in chapter 4.

2.3 Markov Chains

We start this section by giving a definition of classical Markov chains. Afterwards we investigate some interesting examples from probability and information theory.
Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $(X_i)_{i\in\mathbb{N}_0}$ be a real–valued stochastic process in discrete time. The filtration the stochastic process is adapted to is of no particular interest and hence, we can conveniently work with the filtration generated by $(X_i)_{i\in\mathbb{N}_0}$, that is, $\mathscr{F}_i = \sigma(X_0,\dots,X_i)$. Since we deal with conditional probabilities in the following, all statement require that these are well–defined; in particular that the event conditioned on is not a null set. We assume that this is always the case and do not mention this prerequisite explicitly.

Definition 2.8. A discrete stochastic process $(X_i)_{i\in\mathbb{N}_0}$ is called a **Markov chain** if for all $i \in \mathbb{N}_0$ and all $x_0,\dots,x_{i+1} \in \mathbb{R}$

$$\mathbb{P}(X_{i+1} = x_{i+1}|X_i = x_i,\dots,X_0 = x_0) = \mathbb{P}(X_{i+1} = x_{i+1}|X_i = x_i)$$

holds.

The first example a probability theorist might think of when hearing the term Markov Chain is the random walk on $\mathbb{Z}^d$:

Example 2.9. Let $x \in \mathbb{Z}^d$ be the starting point of our random walk and let X_i, $i \in \mathbb{N}$, be iid. $\mathbb{Z}^d$–valued random variables. For example, a signed Bernoulli distribution seems to be an appropriate choice. Then the stochastic process $(S_n)_{n\in\mathbb{N}_0}$,

$$S_n = x + \sum_{i=1}^{n} X_i,$$

forms a Markov Chain.

There are numerous examples of Markov Chains in probability theory and even more subtle ones as the Galton–Watson branching process which models the size of a population reproducing itself asexual. Nevertheless, this stochastic process is usually studied using martingale methods in order to have their very strong convergence results at hand.
The information theoretic part of this thesis requires only a small subset of the general notions from probability theory. As before, we only consider

random variables X, Y, Z taking values in countable alphabets S, T and U respectively. Boiling down the theory to this specific setting, we can give the following equivalent characterization of Markov chains.

Theorem 2.10. *The random variables X, Y, Z form a Markov chain if and only if the probability mass functions satisfy*

$$p(x, y, z) = p(x)\,p(y|x)\,p(z|y)$$

for all $(x, y, z) \in S \times T \times U$.
As a notational convenience, we write $X \to Y \to Z$ in this case.

Proof. For the «only if» part using the Markov property of definition 2.8, we easily compute

$$\begin{aligned} p(x, y, z) &= \mathbb{P}(X = x, Y = y, Z = z) \\ &= \mathbb{P}(X = x)\mathbb{P}(Y = y|X = x)\mathbb{P}(Z = z|Y = y, X = x) \\ &= p(x)\,p(y|x)\,p(z|y). \end{aligned}$$

The «if» part follows from

$$\begin{aligned} \mathbb{P}(Z = z|Y = y, X = x) &= \frac{p(x, y, z)}{p(x, y)} \\ &= \frac{p(x)\,p(y|x)\,p(z|y)}{p(x, y)} \\ &= \frac{p(z, y)}{p(y)} \\ &= \mathbb{P}(Z = z|Y = y). \end{aligned}$$

□

In order to understand Markov chains in information theory better, we give another equivalent characterization of them. To this end, we consider the conditional mutual information $I(X : Z|Y)$.

Theorem 2.11. *X, Y, Z form a Markov chain if and only if $I(X : Z|Y) = 0$.*

Proof. We start by assuming $X \to Y \to Z$. Due to the characterization from theorem 2.10, we find

$$p(x|z, y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x)\,p(y|x)\,p(z|y)}{p(y)} = \frac{p(x, y)\,p(z|y)}{p(y)} = p(x|y)\,p(z|y).$$

From the concluding remarks of § 2.1, we obtain $I(X : Z|Y) = 0$.
Contrarily, $I(X : Z|Y) = 0$ implies $p(x|z, y) = p(x|y)\,p(z|y)$ and hence $p(x, y, z) = p(x)\,p(y|x)\,p(z|y)$ which shows $X \to Y \to Z$. □

Note that $X \to Y \to Z$ implies $Z \to Y \to X$ wherefore some authors use $X \leftrightarrow Y \leftrightarrow Z$ to indicate a Markov chain. The notion of Markov chains in information theory can be enlivened further by investigating their operational consequences. The most famous of these is the so–called data processing inequality.

Theorem 2.12 (Data Processing Inequality). *If $X \to Y \to Z$, then $I(X : Y) \geqslant I(X : Z)$.*

Proof. Due to theorem 2.2, we can expand the mutual information as follows:

$$I(X : Y, Z) = I(X : Y) + I(X : Z|Y) = I(X : Y),$$
$$I(X : Y, Z) = I(X : Z) + I(X : Y|Z).$$

Since $I(X : Y|Z) \geqslant 0$, we find $I(X : Y) \geqslant I(X : Z)$. □

Observe that $X \to Y \to f(Y)$ forms a Markov chain. Therefore the preceding theorem implies $I(X : Y) \geqslant I(X : f(Y))$ which shows that post–processing Y cannot increase the mutual information between X and Y. Following the previous proof, another inequality can be derived. If $X \to Y \to Z$, then $I(X : Y|Z) \leqslant I(X : Y)$. Roughly spoken, the dependence of X and Y is reduced by conditioning on an auxiliary random variable Z which lays on the data road in between them.
In light of theorem 2.11, we say that a tripartite system $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ forms a **quantum Markov chain**, denoted by $A \to B \to C$, if the quantum conditional mutual information between A and C given B vanishes, that is, $I(A : C|B) = 0$. As we will see in § 4.2, this definition is equivalent to the characterization of quantum Markov chains as systems where for any given state $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ there exists a recovery map $\mathcal{R}_{B\to BC}$ which exactly recovers ρ_{ABC} from the marginal $\rho_{AB} = \mathrm{tr}_C[\rho_{ABC}]$, that is, $\rho_{ABC} = \mathcal{R}_{B\to BC}(\rho_{AB})$.
Finally, we want to state the necessary and sufficient condition on a tripartite system in order to form a quantum markov chain.

Theorem 2.13 (Equality in the Strong Subadditivity [Hay+04, Theorem 6]). *A state $\rho_{ABC} \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$ satisfies*

$$S(A, B, C) + S(B) = S(A, B) + S(B, C)$$

if and only if there exists a decomposition of the system B of the form

$$\mathcal{H}_B = \bigoplus_j \mathcal{H}_{b_j^L} \otimes \mathcal{H}_{b_j^R}$$

such that

$$\rho_{ABC} = \bigoplus_j q_j \left(\rho^{(A)}_{b^L_j} \otimes \rho^{(C)}_{b^R_j} \right)$$

for states $\rho^{(A)}_{b^L_j} \in \mathscr{D}\left(\mathscr{H}_A \otimes \mathscr{H}_{b^L_j}\right)$, $\rho^{(C)}_{b^R_j} \in \mathscr{D}\left(\mathscr{H}_{b^R_j} \otimes \mathscr{H}_C\right)$ *and* $q_j \geqslant 0$ *with* $\sum_j q_j = 1$.

Chapter 3

Matrix Product States in a Nutshell

In order to see how the notion of matrix product states (mps) arises naturally in quantum many body physics, we consider the following simple system in one spatial dimension. Suppose we have a chain of N quantum objects which are all described by the finite dimensional Hilbert space $\mathbb{C}^d$. This easy setting is depicted in figure 3.1. Every ground state of this system is of the form

$$|\Psi\rangle = \sum_{i_1,\dots,i_N=1}^{d} c_{i_1\dots i_N} |i_1,\dots,i_N\rangle$$

where the $c_{i_1\dots i_N}$ denote the complex coefficients of the expansion in the onb $\{|i_1,\dots,i_N\rangle\}_{i_1,\dots,i_N}$. Unfortunately, the number of these coefficients grows exponentially in the generic case, in particular $\sim d^N$, since the Hilbert space of the whole system is given by $\left(\mathbb{C}^d\right)^{\otimes N}$. Hence, it is impractical to apply this approach to typical systems condensed matter physicists are interested in. In fact, the reader might think of N in the order of Avogadro's number, that is, $N \sim 10^{23}$. Taking $d = 2$ for a simple spin–up and spin–down chain, the dimension of the total system's Hilbert space exceeds the number of atoms in the universe ($\sim 10^{82}$) by numerous orders of magnitude.

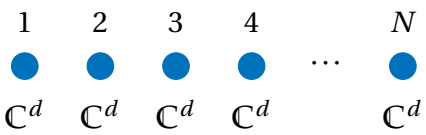


Figure 3.1: Chain of N quantum systems with associated Hilbert space $\mathbb{C}^d$.

3.1 Area Laws

Formally, we describe a n–dimensional lattice as a simple graph $G = (V, E)$ with a set of vertices V and set of edges E. The total Hilbert space $\mathscr{H}$ of the system is the tensor product of the local Hilbert spaces $\mathscr{H}_i$ associated to the

vertex i:

$$\mathscr{H} = \bigotimes_{i \in V} \mathscr{H}_i.$$

Imposing the assumption that the total Hamiltonian H acts locally on the lattice sites, *e.g.* by nearest or next–to–nearest neighbor interaction, the total Hamiltonian reads

$$H = \sum_{L \subset V} H_L$$

where the local Hamiltonians H_L have compact support L. Such approximations are ubiquitous in the models of condensed matter physicists. For instance, one may think of the famous Lennard–Jones potential which shows a $O(r^{-6})$ decay as $r \to \infty$. We use the ad–hoc notation $\mathrm{dist}(i, j)$ to indicate the natural graph distance between the two vertices $i, j \in V$. In the special case of a square lattice modeled by a set of vertices $V \subset \mathbb{Z}^d$, the natural graph distance $\mathrm{dist}(i, j)$ of two vertices $i, j \in V$ is given by $\|i - j\|_1$ where $\|\cdot\|_1$ indicates the 1–norm in $\mathbb{Z}^d$, that is,

$$\|i\|_1 = \sum_{k=1}^{d} |i_k|.$$

For a distinguished region $A \subset V$, as depicted in figure 3.2, we define the boundary ∂A by

$$\partial A = \left\{ i \in A \,\middle|\, \exists j \in V \setminus A : \mathrm{dist}(i, j) = 1 \right\}.$$

Tracing out the complement of this distinguished region A, *i.e.* forming $\rho_A = \mathrm{tr}_{V \setminus A}[\rho]$ for a state $\rho \in \mathscr{D}(\mathscr{H})$ acting on the total system, the entanglement between the region A on the rest of the lattice is given by the von–Neumann entropy of ρ_A. In the generic case of a lattice consisting of d–dimensional constituents, it can be shown that the expectation of the von–Neumann entropy fulfills $\mathbb{E}[S(\rho_A)] \sim |A| \log d$ [Pag93; FK94; Sen96]. Nevertheless, while studying the ground states of local Hamiltonians in further detail, it turns out that the entanglement between the distinguished region A and the remainder of the lattice $V \setminus A$ scales with the volume of the boundary $|\partial A|$ rather than with the volume of the whole region, that is, $S(\rho_A) \sim |\partial A|$. It turns out that this requirement is too strict in case of a gapples Hamiltonian (i.e. the ground state is degenerate) where we say that ρ_A fulfills an area law if $S(\rho_A) \sim |\partial A| \log |A|$. A rigorous mathematical prove is only known in the case of a one–dimensional system with a gapped Hamiltonian [Has07]. In this setting of a spin chain, the area law states that the von–Neumann entropy of the state ρ_A is bounded by a constant independent of the system size N and the length of the distingushed region L (*cf.* figure 3.2). Hence, the physically reasonable states of the many body system live in a small subset of the total Hilbert space and are of a specific structure which we will exploit in the subsequent section. Finally, we want to mention that the notion

of area laws are not solely of interest in the fields of quantum information and condensed matter physics. The survey of Eisert et al. [ECP10] provides further information about area laws and gives a short review of their various applications.

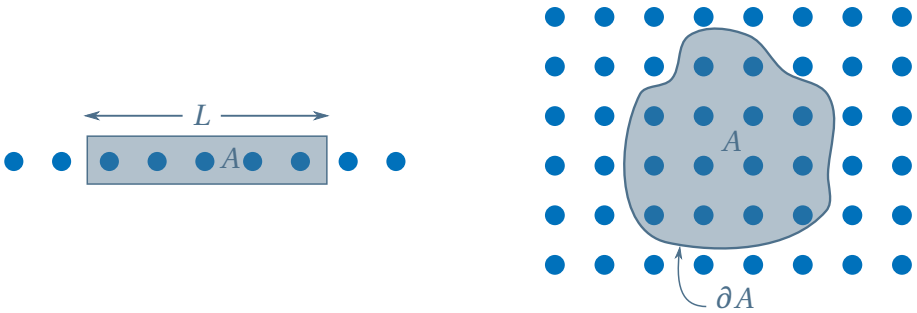


Figure 3.2: The distinguished region A and its boundary ∂A in the one and two spatial dimensions.

3.2 Matrix Product States

We restrict the upcoming discussion to the one–dimensional case only. The reader may find generalizations, in particular to two–dimensional systems, in [BC17; Sch13; Oru14].
Motivated by the area law presented in the previous section, we suppose that the entanglement is concentrated on the boundary of the distinguished region. Therefore, we decompose the vertices of our chain into to two D–dimensional, virtual subsystems. Their dimension may vary from site to site. Nevertheless, we can conveniently think of a single **bond dimension** D if we set

$$D = \max_{i \in V} D_i$$

where D_i denotes the dimension of the virtual subsystems associated with the lattice site i. We impose the assumption that the neighboring virtual subsystems of adjacent sites are in the maximally entangled state

$$|\omega_D\rangle = \sum_{i=1}^{D} |i\rangle \otimes |i\rangle .$$

This construction clearly satisfies an area law since by «cutting» the bonds at the boundary of a distinguished region A, we are left with two maximally entangled states and hence $2 \log D$ is an upper bound on the von–Neumann entropy $S(A)$. In order to recover the original state $|\Psi\rangle$ of the whole chain, we apply the site dependent linear map

$$\mathscr{P}_s : \mathbb{C}^D \otimes \mathbb{C}^D \to \mathbb{C}^d$$

$$\mathscr{P}_s |\psi\rangle = \left(\sum_{i=1}^{d} \sum_{\alpha,\beta=1}^{D} A^{[s]}_{i,\alpha\beta} |i\rangle \langle\alpha| \otimes \langle\beta| \right) |\psi\rangle , \qquad s = 1, \ldots, N \tag{3.1}$$

on the maximally entangled states $|\omega_D\rangle$. Note that equation (3.1) deals with three–index tensors $A^{[s]}_{i,\alpha\beta}$ which can be interpreted as $D \times D$ matrices $A^{[s]}_i$. The state of the whole one–dimensional system is given by

$$|\Psi\rangle = \left(\bigotimes_{s=1}^{N} \mathscr{P}_s\right) |\omega_D\rangle^{\otimes N} \tag{3.2}$$

with the linear maps $\mathscr{P}_s$ defined in equation (3.1). We want to mention the technicality that in (3.2) we assumed a system with closed boundary conditions where the first and the last site share a common bound. For a pictorial explanation, we refer to figure 3.3.

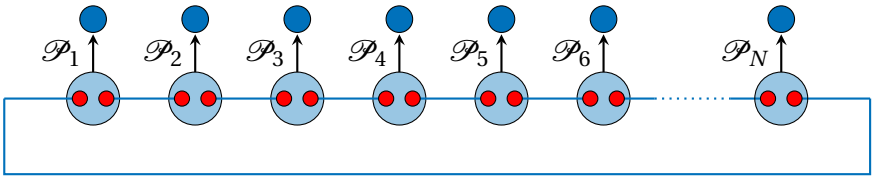


Figure 3.3: Our original spin chain, represented by the dark blue dots, is recovered by the application of the map $\mathscr{P}$ on the ancillary system containing two virtual subsystems on each site (light blue circles). Moreover, note that $|\omega_D\rangle =$ represents the maximally entangled state.

Eventually, we are able to show that the state $|\Psi\rangle$ has a **matrix product state** (mps) representation. We encapsulate this result in the following proposition.

Proposition 3.1. *The state $|\Psi\rangle$ has a representation of the form*

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N=1}^{d} \operatorname{tr}\left[A^{[1]}_{i_1} A^{[2]}_{i_2} \cdots A^{[N]}_{i_N}\right] |i_1,\ldots,i_N\rangle .$$

Proof. Let us investigate how the maps applied to the first and the second site act on the bond between them. Using $\mathbb{1} = \sum_{\alpha_1} |\alpha_1\rangle\langle\alpha_1|$ and $\mathbb{1} = \sum_{\beta_2} |\beta_2\rangle\langle\beta_2|$ (probably not of the same dimension), we find

$$\begin{aligned}
&(\mathscr{P}_1 \otimes \mathscr{P}_2)(\mathbb{1} \otimes |\omega_D\rangle \otimes \mathbb{1}) \\
&\quad = \sum_{\substack{i_1,i_2,j,\\ \alpha_1,\beta_1,\alpha_2,\beta_2}} A^{[1]}_{i_1,\alpha_1\beta_1} A^{[2]}_{i_2,\alpha_2\beta_2} |i_1,i_2\rangle\langle\alpha_1,\beta_1,\alpha_2,\beta_2| (\mathbb{1} \otimes |j,j\rangle \otimes \mathbb{1}) \\
&\quad = \sum_{i_1,i_2,\alpha_1,\beta_1,\beta_2} A^{[1]}_{i_1,\alpha_1,\beta_1} A^{[2]}_{i_2,\beta_1\beta_2} |i_1,i_2\rangle\langle\alpha_1,\beta_2| .
\end{aligned}$$

By iterating the argument, we obtain from (3.2) a final state of the form

$$|\Psi\rangle = \sum_{\substack{i_1,\dots,i_N,\\ \alpha,\beta,\dots,\omega}} A^{[1]}_{i_1,\alpha\beta}\cdots A^{[N]}_{i_N,\omega\alpha}\,|i_1,\dots,i_N\rangle = \sum_{i_1,\dots,i_N} \operatorname{tr}\left[A^{[1]}_{i_1}\cdots A^{[N]}_{i_N}\right]|i_1,\dots,i_N\rangle\,. \tag{3.3}$$

Clearly, the small finite number of Greek letters does not suffice to label all the tensors in the generic case. In order to avoid several technicalities (e.g. distinguishing between N even or odd), we prefer this lax notion since the reader is able to fill in the details in a more specific situation. □

Note that in the case of open boundary conditions, the tensors associated to the first and the last site are row and column vectors respectively. Hence in this setting, we find

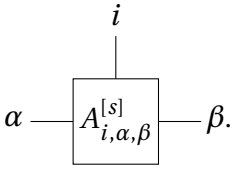
$$|\Psi\rangle = \sum_{i_1,\dots,i_N} A^{[1]}_{i_1}\cdots A^{[N]}_{i_N}\,|i_1,\dots,i_N\rangle \tag{3.4}$$

where we are allowed to omit the trace.
Occasionally, one encounters translational invariant mps for which $A_{i_j} := A^{[1]}_{i_j} = \cdots = A^{[N]}_{i_j}$ for any $j \in \{1,\dots,N\}$. Consequently, equation (3.3) simplifies as follows:

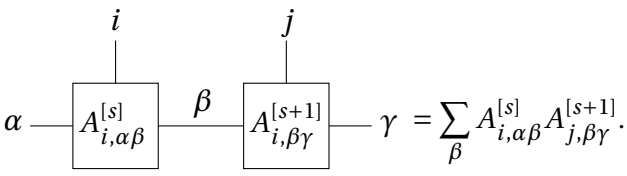
$$|\Psi\rangle = \sum_{i_1,\dots,i_N} \operatorname{tr}\left[A_{i_1}\cdots A_{i_N}\right]|i_1,\dots,i_N\rangle\,.$$

3.2.1 A Graphical Language for Matrix Product States

We want to introduce briefly a widespread and convenient notation for tensor networks, and especially mps, which is adapted from the Penrose notation used in multilinear algebra. Let us start with one of the three index tensors $A^{[s]}_{i,\alpha\beta}$ occurring in equation (3.4). We represent it graphically by a box with three adjacent «legs», each corresponding to one of the three indices:



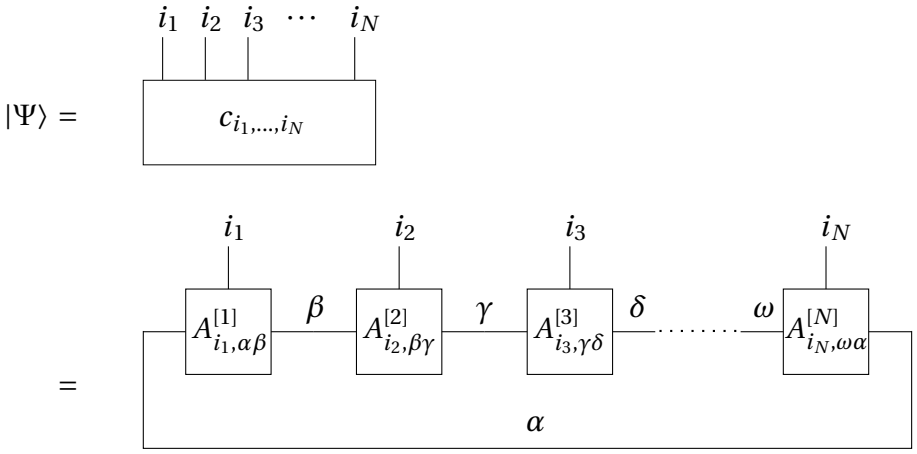
When two of the legs are connected, we sum over the corresponding common index. In graphical language:



With these illustrations at hand, we obtain for the full mps representation of the state $|\Psi\rangle$

$$|\Psi\rangle = \sum_{i_1,\ldots,i_N=1}^{d} c_{i_1,\ldots,i_N} |i_1,\ldots,i_N\rangle = \sum_{i_1,\ldots,i_N} \mathrm{tr}\left[A^{[1]}_{i_1}\cdots A^{[N]}_{i_N}\right] |i_1,\ldots,i_N\rangle$$

the diagrammatic expression



where we used closed boundary conditions.

3.2.2 Examples of Matrix Product States

In order to familiarize ourselves with the notion of mps, we will discuss several easy examples. In the presented settings, we will always be able to give the exact mps representation of the total system's state.

Product States

Arguably the easiest example we can consider is the case where the total state is of the simplest factorized form, that is for instance, $|\Psi\rangle = |1,\ldots,1\rangle$. This state can be represented by introducing the translational invariant mps formed by the 1×1–matrices

$$A_1 = (1), \qquad A_j = (0) \text{ for } j \geqslant 2.$$

An alternative representation is given by the $D\times D$–matrices

$$A_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad A_j = 0 \text{ for } j \geqslant 2$$

for some arbitrary bond dimension $D \in \mathbb{N}$. In order to achieve an efficient description, we choose the mps representation with the smallest bond dimension. Nevertheless, even the mps representation with the smallest bond

dimension is not unique since for B_s, $1 \leqslant s \leqslant N-1$, with right inverse B_s^{-1}, we can perform a gauge transformation of the form

$$A_i^{[s]} \to A_i^{[s]} B_s$$
$$A_i^{[s+1]} \to B_s^{-1} A_i^{[s+1]}$$

without changing the representation.

W State

Let us investigate the mps generated by the matrices

$$A_1^{[s]} = \frac{1}{\sqrt[2N]{N}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A_2^{[s]} = \frac{1}{\sqrt[2N]{N}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

for $s \in \{2, \ldots, N-1\}$ and

$$A_i^{[1]} = \begin{pmatrix} 1 & 0 \end{pmatrix} A_i^{[2]}, \qquad A_i^{[N]} = A_i^{[2]} \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

for $i \in \{1, 2\}$. Since $A_2^{[s]}$, $2 \leqslant s \leqslant N-1$, is nilpotent, we find the W state:

$$|W\rangle = \sum_{i=1}^{N} |1, \ldots, 1, \underset{\substack{\uparrow \\ i^{\text{th}} \text{ position}}}{2}, 1, \ldots, 1\rangle .$$

Greenberger-Horne-Zeilinger State

The Greenberger–Horne–Zeilinger state, which is of interest in the field of quantum cryptography, is given by

$$|\text{GHZ}\rangle = \frac{1}{\sqrt{2}} \left(|1, \ldots, 1\rangle + |2, \ldots, 2\rangle \right).$$

Clearly, the simplest translational invariant matrix product representation is obtained by setting

$$A_1 = \frac{1}{\sqrt[2N]{2}} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad A_2 = \frac{1}{\sqrt[2N]{2}} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

AKLT State

Eventually, we study a model of a spin chain introduced by Affleck, Kennedy, Lieb and Tasaki [Aff+87; Aff+88]. According to Perez–Garcia et al. [Per+07] the ground state of this system is the «father of all matrix product states». We start with a slightly different setup as under § 3.2. Let us impose the assumption that the site of the spin chain consist of two virtual spin–1/2 subsystems. In contrast to the construction of a mps, we put the bond in the spin–1/2 singlet state ($S = 0$)

$$|\xi\rangle = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle).$$

Note that by this construction, we still obtain a mps since the unitary matrix

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

which transforms $|\xi\rangle$ to

$$|\omega_2\rangle = \frac{1}{\sqrt{2}}(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle)$$

can be absorbed in the map $\mathscr{P}$. In contrast to § 3.2, we choose $\mathscr{P} = \Pi_{S=1}$ where $\Pi_{S=1}$ denotes the orthogonal projection on the spin−1 subspace. The whole model is depicted in figure 3.4. Using a group theoretic argument, it can be shown that the partial trace of two consecutive sites cannot be in the state $S = 2$. Since this is non−trivial constraint, we have found a local Hamiltonian for which the AKLT state forms the ground state. We do not want to study the construction of the Hamiltonian in further detail. The interested reader may find a detailed exposition in [Sch13].

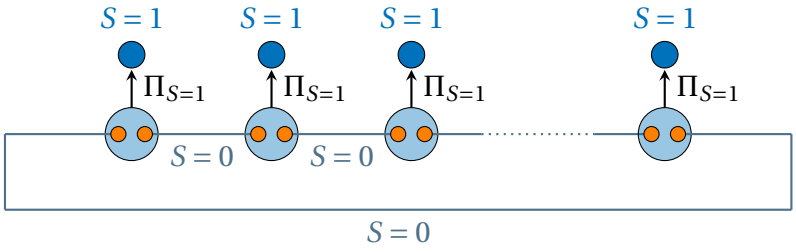


Figure 3.4: The AKLT model with closed boundary conditions.Analogue toto § 3.2, we obtain a mps by this construction. Note that $|\xi\rangle =$  represents the singlet state.

Summing up the local Hamiltonians yields

$$H = \sum_{i=1}^{N-1} \mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{3}(\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2 .$$

The vector $\mathbf{S}_i \in \mathbb{C}^3$ is the the spin−1 operator associated to site i. Note that the first term of the sum coincides with the ordinary one−dimensional Heisenberg spin model without a magnetic field (except the fact that the AKLT Hamiltonian deals with spin−1 operators).
The following matrices [Sch11] represent the ground state of the AKLT Hamiltonian as mps:

$$A^+ = \begin{pmatrix} 0 & \sqrt{\frac{2}{3}} \\ 0 & 0 \end{pmatrix}, \qquad A^0 = \begin{pmatrix} -\frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{pmatrix}, \qquad A^- = \begin{pmatrix} 0 & 0 \\ -\sqrt{\frac{2}{3}} & 0 \end{pmatrix}.$$

We changed the indexing of the matrices in order to emphasize the correspondence to the triplet states ($S = 1$)

$$\begin{aligned} |+\rangle &= |\uparrow\uparrow\rangle \\ |0\rangle &= \frac{|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle}{\sqrt{2}} \\ |-\rangle &= |\downarrow\downarrow\rangle\,. \end{aligned}$$

The preceding examples were only a selection from the models with an exact mps representation. Unfortunately, the most interesting physical system do not have such simple ground states. Nevertheless, mps often provide a powerful numerical tool for their approximation.

Chapter 4

Recovery Maps

In this chapter, we introduce and study the main topic of this thesis. The concept of recovery maps was adressed by Petz [Pet86; Pet88] for the first time. He investigated the equality conditions in the monotonicity of quantum relative entropy and established the so–called «transpose channel». The topic had been only of minor interest for several years before the groundbreaking paper of Fawzi and Renner [FR15] was published. Together with Sutter [SFR16], they were able to improve their result even further. We will consider their work in § 4.2. Fortunately, an explicit structure of a recovery map is known for finite–dimensional systems. It has a nice mathematical derivation, in particular it uses some basic notions of complex and harmonic analysis, and thus we want to sketch the proof given by Wilde [Wil15] (*c.f.* § 4.1 and appendix A.3).

Before we address the mathematical details of the recovery maps, we should motivate our interest in the concept of recoverability. Let us recall the simple setting of a one–dimensional chain of finite length from figure 3.1. We impose the assumption that we already know the state of the first two sites $\sigma_{1,2} = \mathrm{tr}_{3,\dots,N}[|\Psi\rangle\langle\Psi|]$. We aim to reconstruct the total state $\sigma_{1,\dots,N} = |\Psi\rangle\langle\Psi|$ by a successive application of linear maps on $\sigma_{1,2}$. To this end, we consider the action of a single recovery map at the first place. In the case of a vanishing quantum conditional mutual information $I(1:3|2)_\rho = 0$, it is well–known that the Petz recovery map ensures a perfect recoverability of $\sigma_{1,2,3}$. In particular, we have $I(1:3|2)_\rho = 0$ if $\sigma_{1,2,3}$ is a product state, that is, $\sigma_{1,2,3} = \sigma_1 \otimes \sigma_2 \otimes \sigma_3$.[*] However in general, we are not permitted to assume that a pure state $\sigma_{1,2,3}$ is of product form. We can define a natural probability measure on the unit sphere $S^{n-1} \subset \mathbb{C}^n$ by considering the Haar measure on the unitary group $\mathcal{U}(n)$. To this end, we pick a starting vector $|\psi_0\rangle \in S^{n-1}$ which we undertake a random unitary evolution drawn w.r.t. the Haar measure. Invoking Lévy's lemma [Led05], a result from the study of the concentration of measure phenomenon, it can be shown that a given state is almost maximally entangled with an overwhelming probability, i.e.

[*] Recall that the necessary and sufficient condition of a vanishing quantum conditional information is encapsulated in theorem 2.13.

exponential saturation to one as $n \to \infty$. Based on the work of Fawzi, Renner and Sutter, we are interested in states with a small but non–vanishing quantum conditional mutual information, that is, $0 < I(A:C|B)_\rho \leqslant \varepsilon$ for some small $\varepsilon > 0$. These states are said to form an **approximate Markov chain** (*cf.* § 2.3). Nevertheless, we start our exposition with studying the Petz recovery map and Wilde's recoverability theorem in § 4.1.

4.1 Recoverability Theorem

The exposition of this section mainly follows the textbook of Wilde [Wil17]. We start by introducing an explicit recovery map which will appear slightly modified in Wilde's recoverability theorem. Moreover, this map will play an important role in the subsequent sections.

Definition 4.1. Let $\rho \in \mathscr{D}(\mathscr{H}_A)$ and $\mathcal{N} : \mathscr{D}(\mathscr{H}_A) \to \mathscr{D}(\mathscr{H}_B)$ be a quantum channel. We define the **Petz recovery map** by

$$\begin{aligned} \mathscr{P}_{\rho,\mathcal{N}} &: \mathscr{B}(\mathscr{H}_B) \to \mathscr{B}(\mathscr{H}_A) \\ X &\mapsto \rho^{1/2} \mathcal{N}^\dagger \left(\left(\mathcal{N}(\rho)\right)^{-1/2} X \left(\mathcal{N}(\rho)\right)^{-1/2} \right) \rho^{1/2}. \end{aligned}$$

The **adjoint** $\mathcal{E}^\dagger : \mathscr{B}(\mathscr{H}_B) \to \mathscr{B}(\mathscr{H}_A)$ of a linear map $\mathcal{E} : \mathscr{B}(\mathscr{H}_A) \to \mathscr{B}(\mathscr{H}_B)$ is defined as the unique linear operator satisfying

$$\langle B | \mathcal{E}(A) \rangle_{HS} = \langle \mathcal{E}^\dagger(B) | A \rangle_{HS}$$

for all $A \in \mathscr{B}(\mathscr{H}_A)$, $B \in \mathscr{B}(\mathscr{H}_B)$. It can be shown that the adjoint of a quantum is unital and completely positive, but not necessarily trace preserving.

A legitimate question the reader may ask is whether the adjoint $\mathcal{N}^\dagger$ of a quantum channel $\mathcal{N}$ is a quantum operation itself. The answer is negative in the generic case. In fact using the Kraus representation of $\mathcal{N}$ (*cf.* theorem 1.17), we observe that

$$\langle B \,|\, \mathcal{N}(A) \rangle_{HS} = \operatorname{tr}\left[B^\dagger \sum_j V_j A V_j^\dagger \right] = \operatorname{tr}\left[\left(\sum_j V_j^\dagger B V_j \right)^\dagger A \right]$$

and hence, $\mathcal{N}^\dagger(B) = \sum_j V_j^\dagger B V_j$. Consequently, we deduce that the adjoint of a quantum channel is completely positive and unital. $\mathcal{N}^\dagger$ is in addition trace–preserving if and only if $\mathcal{N}$ is unital.

Since the Petz recovery map from definition 4.1 is a concatenation of three completely positive linear maps, it is linear and completely positive as well. Furthermore, if $\operatorname{supp} X \subset \operatorname{supp} \rho$, then the Petz recovery map $\mathscr{P}_{\rho,\mathcal{N}}$ is trace–

preserving. In fact, we have

$$\begin{aligned}\operatorname{tr}\left[\mathscr{P}_{\rho,\mathscr{N}}(X)\right] &= \operatorname{tr}\left[\rho\mathscr{N}^\dagger\left(\left(\mathscr{N}(\rho)\right)^{-1/2}X\left(\mathscr{N}(\rho)\right)^{1/2}\right)\right]\\ &= \operatorname{tr}\left[\mathscr{N}(\rho)\left(\mathscr{N}(\rho)\right)^{-1/2}X\left(\mathscr{N}(\rho)\right)^{1/2}\right]\\ &= \operatorname{tr}\left[\Pi_{\mathscr{N}(\rho)}X\right] = \operatorname{tr}[X].\end{aligned}$$

In the case $\operatorname{supp} X \not\subset \operatorname{supp}\rho$ following the lines of the computation above, we can deduce that the Petz recovery map $\mathscr{P}_{\rho,\mathscr{N}}$ is at least trace–nonincreasing, that is, $\operatorname{tr}\left[\mathscr{P}_{\rho,\mathscr{N}}(X)\right] \leqslant \operatorname{tr}[X]$, due to the fact $\Pi_{\mathscr{N}(\rho)}X \leqslant X$ and the monotonicity of the trace.
For $t \in \mathbb{R}$ and a density operator $\rho \in \mathscr{D}(\mathscr{H})$, we introduce a partial isometry via

$$\begin{aligned}\mathscr{U}_{\rho,t} : \mathscr{B}(\mathscr{H}) &\to \mathscr{B}(\mathscr{H})\\ X &\mapsto \rho^{it}X\rho^{-it}\end{aligned}$$

where we defined $\rho^{\pm it} = \sum_{j:\lambda_j\neq 0}\lambda_j^{\pm it}\left|\varphi_j\right\rangle\left\langle\varphi_j\right|$ using the spectral decomposition of ρ as usual. Observe that $\rho^{it}\rho^{-it} = \Pi_\rho$ and hence, this partial isometry has a rather convenient property:

$$\mathscr{U}_{\rho,t}(\rho) = \rho.$$

The fact that $\mathscr{U}_{\rho,t}$ leaves the state of its first index invariant is the impetus of the definition of the rotated Petz recovery map.

Definition 4.2. For $t \in \mathbb{R}$, a density operator $\rho \in \mathscr{D}(\mathscr{H})$ and a quantum channel $\mathscr{N} : \mathscr{D}(\mathscr{H}_A) \to \mathscr{D}(\mathscr{H}_B)$, we define the **rotated Petz recovery map** as follows:

$$\begin{aligned}{}^{\mathrm{rot}}\mathscr{P}^{\,t}_{\rho,\mathscr{N}} : \mathscr{D}(\mathscr{H}_B) &\to \mathscr{B}(\mathscr{H}_A)\\ X &\mapsto \left(\mathscr{U}_{\rho,-t}\circ\mathscr{P}_{\rho,\mathscr{N}}\circ\mathscr{U}_{\mathscr{N}(\rho),t}\right)(X).\end{aligned}$$

After having introduced and studied all the necessary ingredients, we are now able to state the main result of this section, that is, Wilde's recoverability theorem.

Theorem 4.3 (Wilde's Recoverability Theorem; Theorem 4 in [Wil15] and Chapter 12 in [Wil17]). *Let $\rho,\sigma \in \mathscr{D}(\mathscr{H}_A)$ such that $\operatorname{supp}\rho \subset \operatorname{supp}\sigma$ and let $\mathscr{N} : \mathscr{D}(\mathscr{H}_A) \to \mathscr{D}(\mathscr{H}_B)$ be a quantum channel.*

(i) Then we have

$$\begin{aligned}&S(\rho\|\sigma) - S(\mathscr{N}(\rho)\|\mathscr{N}(\sigma))\\ &\qquad\geqslant -2\int_{-\infty}^{\infty}\beta_0(t)\log\left[F\left(\rho,\left({}^{\mathrm{rot}}\mathscr{P}^{\,t/2}_{\sigma,\mathscr{N}}\circ\mathscr{N}\right)(\rho)\right)\right]\,dt\end{aligned}$$

where $\beta_0(t) = \lim_{\theta \searrow 0} \beta_\theta(t) = \pi/2\,(\cosh(\pi t) + 1)^{-1}$ *(cf. theorem A.10 and figure 4.1). Observe that* $\beta_0(t)$ *is a probability density.*

(ii) Then there exists a recovery map $\mathscr{R}_{\sigma,\mathcal{N}} : \mathscr{D}(\mathscr{H}_B) \to \mathscr{D}(\mathscr{H}_A)$ *such that*

$$S(\rho\|\sigma) - S(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geqslant -2\log F\big(\rho, (\mathscr{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\rho)\big)$$

and $(\mathscr{R}_{\sigma,\mathcal{N}} \circ \mathcal{N})(\sigma) = \sigma$*. Moreover, a possible recovery map* $\mathscr{R}_{\sigma,\mathcal{N}}$ *is given by*

$$\mathscr{R}_{\sigma,\mathcal{N}}(X) = \int_{-\infty}^{\infty} \beta_0(t)\, {}^{rot}\mathscr{P}^{t/2}_{\sigma,\mathcal{N}}(X)\, dt + \operatorname{tr}\big[(\mathbb{1} - \Pi_{\mathcal{N}(\sigma)})X\big]\eta$$

for some $\eta \in \mathscr{D}(\mathscr{H}_A)$*. The last summand ensures the trace–preserving property of* $\mathscr{R}_{\sigma,\mathcal{N}}$*.*

From this theorem, we want to draw a simple consequence. Observe that theorem 4.3 (ii) together with proposition 1.27 (i) implies the monotonicity of the quantum relative entropy under a quantum channel $\mathcal{N}$, that is, $S(\rho\|\sigma) \geqslant S(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$. This is a highly non–trivial and quite important result of quantum information theory first proven by Lindblad [Lin75] and Uhlmann [Uhl77] back in the 1970s.

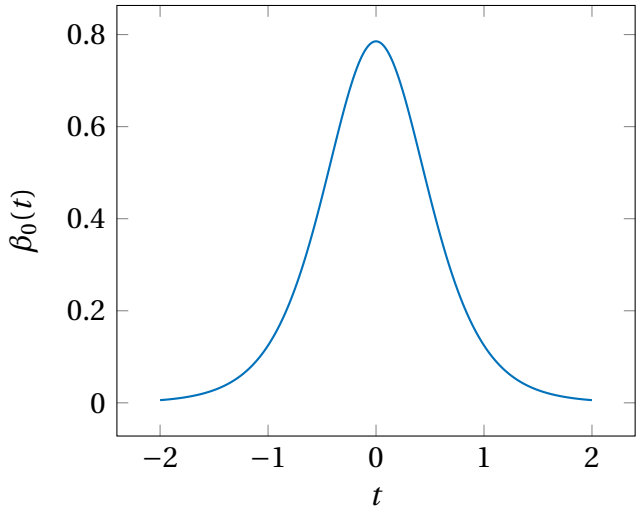


Figure 4.1: The plot depicts the probability density function $\mathbb{R} \ni t \mapsto \beta_0(t) = \pi/2\,(\cosh(\pi t) + 1)^{-1}$ from theorem 4.3. Observe that the peak at $t = 0$ corresponds to the ordinary Petz recovery map.

4.2 The Breakthrough Result of Fawzi and Renner

In 2014, Fawzi and Renner [FR15] proved a lower bound on the fidelity between the exact quantum state and the recovered one in terms of the quan-

tum mutual information. Unfortunately, their proof was non–constructive and did not reveal any information about a particular recovery map which satisfies this bound. Moreover, their argument is rather technical involved and uses several quantities from quantum information theory which we do not consider in this thesis. Together with Sutter [SFR16], they were able to generalize the previous result towards the existence of a universal recovery map which satisfies the bound for any extension $\rho_{ABC} \in \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B \otimes \mathscr{H}_C)$ of a given marginal $\rho_{BC} \in \mathscr{D}(\mathscr{H}_B \otimes \mathscr{H}_C)$, that is, $\mathrm{tr}_A[\rho_{ABC}] = \rho_{BC}$. Hence, this universal recovery map may only depend on the marginals ρ_B and ρ_{BC} but neither on ρ_{AB} nor ρ_{ABC}.
The main results of the paper of Fawzi, Renner and Sutter are the following two theorems forming the cornerstone of our subsequent work.

Theorem 4.4 (Theorem 2.1 in [SFR16]). *For any density operator ρ_{BC} on $B \otimes C$ there exists a trace–preserving completely positive map $\mathscr{R}_{B\to BC}$ such that for any extension $\rho_{ABC} \in \mathscr{D}(A \otimes B \otimes C)$*

$$I(A:C|B)_\rho \geqslant -2 \log F\left(\rho_{ABC}, \mathscr{R}_{B\to BC}(\rho_{AB})\right), \tag{4.1}$$

where A, B and C are separable Hilbert spaces.

Note that we adopt their notational convenience of writing $\mathscr{R}_{B\to BC}(\rho_{AB})$ instead of $(\mathbb{1}_A \otimes \mathscr{R}_{B\to BC})(\rho_{AB})$ whenever there is no possibility of confusion.

Theorem 4.5 (Corollary 2.4 and Remark 2.5 in [SFR16]). *For any density operator ρ_{BC} on $B \otimes C$ there exists a trace–preserving completely positive map $\mathscr{R}_{B\to BC}$ such that for any extension $\rho_{ABC} \in \mathscr{D}(A \otimes B \otimes C)$*

$$F\left(\rho_{ABC}, \mathscr{R}_{B\to BC}(\rho_{AB})\right) \geqslant 1 - \frac{\ln 2}{2} I(A:C|B)_\rho.$$

The recovery map $\mathscr{R}_{B\to BC}$ has the form

$$X_B \mapsto \int V_{BC}^s \rho_{BC}^{1/2} \left(\rho_B^{-1/2} U_B^s X_B U_B^{s\dagger} \rho_B^{-1/2} \otimes \mathbb{1}_C\right) \rho_{BC}^{1/2} V_{BC}^{s\dagger} \, d\mu(s) \tag{4.2}$$

for some probability measure μ on some set S. The family of unitaries $\{V_{BC}^s\}_{s\in S}$ on $B \otimes C$ commutes with ρ_{BC} and the unitaries $\{U_B^s\}_{s\in S}$ on B commute with ρ_B. For μ–almost all density operators ρ_{BC}, we can replace the unitaries U_B^s and V_{BC}^s by complex matrix exponentials of the form ρ_B^{it} and ρ_{BC}^{it}, respectively, with $t \in \mathbb{R}$.

Actually, theorem 4.4 can be deduced from theorem 4.3 in the case of finite–dimensional Hilbert spaces. To this end, we set $\rho = \rho_{ABC}$, $\sigma = \mathbb{1}_A \otimes \rho_{BC}$ and $\mathscr{N} = \mathrm{tr}_C$. Hence, we obtain $\mathscr{N}(\rho) = \rho_{AB}$, $\mathscr{N}(\sigma) = \mathbb{1} \otimes \rho_B$ and $\mathscr{N}^\dagger(X) =$

$X \otimes \mathbb{1}_C$. Moreover, we have

$$\begin{aligned} S(\rho\|\sigma) - S(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) &= S(\rho_{ABC}\|\mathbb{1}_A \otimes \rho_{BC}) - S(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) \\ &= -S(A,B,C) + S(B,C) + S(A,B) - S(B) \\ &= I(A:C|B)_\rho \end{aligned}$$

where the defining property of the partial trace as well as $\log(\mathbb{1} \otimes A) = \mathbb{1} \otimes \log A$ was used. Thus, part (ii) of Wilde's recoverability theorem implies

$$I(A:C|B) \geqslant -2\log\big(F(\rho_{ABC}, \mathscr{R}_{B\to BC}(\rho_{AB}))\big)$$

for some recovery map $\mathscr{R}_{B\to BC} : \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B) \to \mathscr{D}(\mathscr{H}_A \otimes \mathscr{H}_B \otimes \mathscr{H}_C)$. Wilde's theorem provides even an explicit form of this map which we now going to compute. To this end, note that

$$\begin{aligned} \mathscr{P}_{\sigma,\mathcal{N}}(X_B) &= \sigma^{1/2}\mathcal{N}^\dagger\left((\mathcal{N}(\sigma))^{-1/2} X_B (\mathcal{N}(\sigma))^{-1/2}\right)\sigma^{1/2} \\ &= \mathbb{1}_A \otimes \left[\rho_{BC}^{1/2}\left(\rho_B^{-1/2} X \rho_B^{-1/2} \otimes \mathbb{1}_C\right)\rho_{BC}^{1/2}\right] \end{aligned}$$

and hence, the recovery map reads

$$\mathscr{R}_{B\to BC}(X_B) = \int_{-\infty}^{\infty} \beta_0(t)\rho_{BC}^{it/2}\rho_{BC}^{1/2}\left(\rho_B^{-1/2}\rho_B^{it/2} X_B \rho_B^{it/2}\rho_B^{-1/2} \otimes \mathbb{1}_C\right)\rho_{BC}^{1/2}\rho_{BC}^{-it/2}\, dt$$

since the additional summand vanishes ($\mathbb{1} - \Pi_{\mathcal{N}(\sigma)} = \Pi_{(\operatorname{supp}\mathcal{N}(\sigma))^\perp}$).
Let us finish this section with a brief summary. We have seen that Wilde's approach gives an explicit form of a universal recovery map. In particular, we now know a possible probability measure μ in equation (4.2): we have $\mu \ll \lambda$ and $\frac{d\mu}{d\lambda}(t) = \pi/2\,(\cosh(\pi t) + 1)^{-1}$. A recent paper of Berta, Sutter and Tomamichel [SBT17] used the interpolation techniques of Wilde to prove a generalization of Lieb's three–matrix inequality and they obtained a strengthening of Wilde's recoverability theorem in terms of the measured quantum relative entropy as side product. We do not want to consider their results further in thesis.

Chapter 5

Iterating the Recovery Map through the Spin Chain

Based on the results of Fawzi, Renner and Sutter, we are now able to present our own results from the investigation of the recovery procedure sketched in the introduction of this thesis. The following paragraphs are used to restate the considered setting and to fix some notation. Furthermore, we deduce a proposition from theorem 4.5 which provides an upper bound on the error we make by a successive application of the recovery map on the spin chain. Afterwards in § 5.1, we show that we can approximate the pure system's state $\sigma_{1,\dots,N}$ by a mps. As we will see below, an important cornerstone underlying our argument is lemma 5.3 which exploits the geometry of convex combinations of unit vectors in spaces equipped with a scalar product. Eventually, § 5.2 studies the scaling of the bond dimension of the obtained mps and deduces constraints on the saturation behavior of the von–Neumann entropy in order to achieve a polynomial, or at least quasi–polynomial*, ascent due to an increasing number of sites.

Recall the setup of the one–dimensional chain depicted in figure 3.1. We use the intuitive labeling $1,\dots,N$ for the respective systems. The exact, but unknown, state of a distinguished region, containing the j subsystems $i, i+1,\dots,i+j-1$, $1 \leqslant i \leqslant N-j+1$, is denoted by $\sigma_{i,i+1,\dots,i+j-1}$. Using the same notational convention, we label the recovered, and probably defective, state by $\rho_{i,i+1,\dots,i+j-1}$. In order to give an upper bound on the inevitable error we make by iterating the recovery map, we investigate its propagation through the process. We assume that the state of the whole system $\sigma_{1,\dots,N} = |\Psi\rangle\langle\Psi|$ is pure, start our reconstruction procedure with the exact state

$$\sigma_{1,2} = \mathrm{tr}_{3,\dots,N}\left[\sigma_{1,\dots,N}\right]$$

and apply the recovery map from theorem 4.5 successively. This process is illustrated in figure 5.1.

*The reader may note that the term «quasi–polynomial» has no consistent use throughout the literature. We require a scaling behavior comparable to $\exp(\mathrm{poly}(\ln(x)))$ to say that a given quantity scales quasi–polynomial in x.

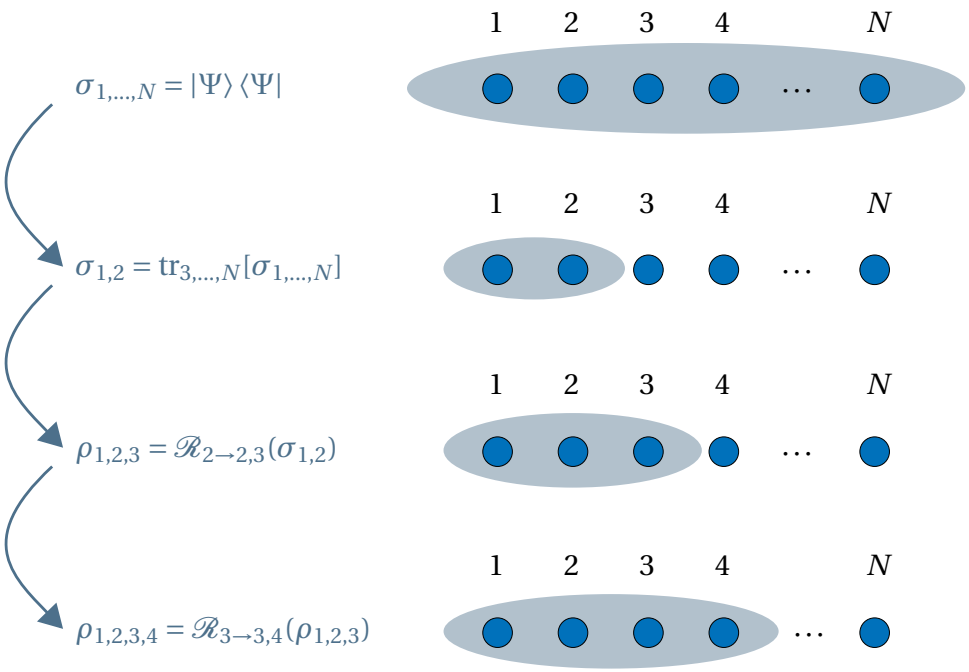


Figure 5.1: The figure depicts a sketch of the reconstruction procedure we consider in this thesis. Note that in the last step we let the recovery map only act on system three. Up to now, we do not known whether this is the optimal choice. We tackle this question in § 5.2.

It turns out that the error grows linear in the system size N as the following proposition shows.

Proposition 5.1. *Let $n \in \mathbb{N}$, $2 \leqslant n \leqslant N$, then the error generated by the procedure described above is upper bounded by*

$$\|\rho_{1,\dots,n} - \sigma_{1,\dots,n}\|_1 \leqslant 2\sqrt{\ln 2} \sum_{i=3}^{n} \sqrt{\varepsilon_i} \tag{5.1}$$

where we defined $\varepsilon_i = I(i-2 : i|i-1)_\rho$ for $i = 3, \dots, n$.
In particular, if the quantum conditional mutual information remains constant throughout the chain, that is, $\varepsilon_3 = \cdots = \varepsilon_n =: \varepsilon$, we have

$$\|\rho_{1,\dots,n} - \sigma_{1,\dots,n}\|_1 \leqslant 2\sqrt{\ln 2}(n-2)\sqrt{\varepsilon}.$$

Proof. We prove the statement by induction over $n \in \mathbb{N}$, $2 \leqslant n \leqslant N$. The basis is trivial since both sides of equation (5.1) vanish. In this case, we have $\rho_{1,2} = \sigma_{1,2}$.
For the inductive step, first observe that

$$\|\rho_{1,\dots,n+1} - \sigma_{1,\dots,n+1}\|_1 \leqslant \|\mathscr{R}(\rho_{1,\dots,n}) - \mathscr{R}(\sigma_{1,\dots,n})\|_1 + \|\mathscr{R}(\sigma_{1,\dots,n}) - \sigma_{1,\dots,n+1}\|_1$$

where we omitted the index of the recovery map to enhance readability. By theorem 1.24 and the induction hypothesis, we have

$$\|\mathscr{R}(\rho_{1,\dots,n}) - \mathscr{R}(\sigma_{1,\dots,n})\|_1 \leqslant \|\rho_{1,\dots,n} - \sigma_{1,\dots,n}\|_1 \leqslant 2\sqrt{\ln 2}\sum_{i=3}^{n}\sqrt{\varepsilon_i}. \tag{5.2}$$

Moreover, due to the first Fuchs–van de Graaf inequality (proposition 1.29) and the result of Fawzi et al. (theorem 4.5), the second summand can be bounded by

$$\begin{aligned}\|\mathscr{R}(\sigma_{1,\dots,n}) - \sigma_{1,\dots,n+1}\|_1 &\leqslant 2\sqrt{1 - F\left(\mathscr{R}(\sigma_{1,\dots,n}), \sigma_{1,\dots,n+1}\right)^2} \\ &\leqslant 2\sqrt{\ln(2)\varepsilon_{n+1} - \left(\frac{\ln(2)\varepsilon_{n+1}}{2}\right)^2} \\ &\leqslant 2\sqrt{\ln(2)\varepsilon_{n+1}}. \end{aligned} \tag{5.3}$$

Combining the equations (5.2) and (5.3) yields the desired result and proves the proposition. □

5.1 Constructing a Mps

In order to construct an explicit mps through iterating the recovery map, we consider the Kraus operators of the latter. Recall that the (non–universal) recovery maps has the form

$$\mathscr{R}_{B\to BC}(X_B) = V_{BC}\rho_{BC}^{1/2}\left(\rho_B^{-1/2}U_B X_B U_B^\dagger \rho_B^{-1/2} \otimes \mathbb{1}_C\right)\rho_{BC}^{1/2}V_{BC}^\dagger.$$

By introducing $M_{BC} = V_{BC}\rho_{BC}^{1/2}(\rho_B^{-1/2}U_B \otimes \mathbb{1}_C)$, we obtain the Kraus representation of the recovery map:

$$\mathscr{R}_{B\to BC}(X_B) = M_{BC}(X_B \otimes \mathbb{1}_C)M_{BC}^\dagger. \tag{5.4}$$

Let us investigate how the recovery maps act on our starting state $\sigma_{1,2} = \mathrm{tr}_{3,\dots,N}[\sigma_{1,\dots,N}]$. We find

$$\rho_{1,2,3} = (\mathbb{1}_1 \otimes \mathscr{R}_{2\to 2,3})(\sigma_{1,2}) = (\mathbb{1}_1 \otimes M_{2,3})(\sigma_{1,2} \otimes \mathbb{1}_3)(\mathbb{1}_1 \otimes M_{2,3})^\dagger. \tag{5.5}$$

To avoid confusing expressions in the following, we define $N_i = \mathbb{1}_{1,\dots,i-1} \otimes M_{i,i+1}$, $2 \leqslant i \leqslant N-1$, such that equation (5.5) simplifies towards

$$\rho_{1,2,3} = N_2(\sigma_{1,2} \otimes \mathbb{1}_3)N_2^\dagger.$$

A moment's thought reveals that after $n-2$ iterations, we obtain the state

$$\begin{aligned}\rho_{1,\dots,n} &= (\mathbb{1}_{1,\dots,n-2} \otimes \mathscr{R}_{n-1\to n-1,n})(\rho_{1,\dots,n-1}) \\ &= N_{n-1}(\rho_{1,\dots,n-1} \otimes \mathbb{1}_n)N_{n-1}^\dagger \\ &= N_{n-1}\left(\left\{N_{n-2}\left[\cdots N_3\left[\left(N_2(\sigma_{1,2} \otimes \mathbb{1}_3)N_2^\dagger\right) \otimes \mathbb{1}_4\right]N_3^\dagger \cdots\right]N_{n-2}^\dagger\right\} \otimes \mathbb{1}_n\right)N_{n-1}^\dagger \\ &= \left(N_{n-1} \circ \cdots \circ N_2\right)(\sigma_{1,2} \otimes \mathbb{1}_{3,\dots,n})\left(N_{n-1} \circ \cdots \circ N_2\right)^\dagger\end{aligned}$$

We now show that iterating the recovery map through the spin chain yields a mps. To this end, we first consider the special case where $\sigma_{1,2} = \mathrm{tr}_{3,\dots,N}[\sigma_{1,\dots,N}]$ is pure, that is, $\sigma_{1,2} = |\varphi_{1,2}\rangle\langle\varphi_{1,2}|$ and generalize the obtained results later on to a mixed starting state. Moreover, we fix for each site k, $k = 1,\dots N$, a computational onb $\{|i_k\rangle\}_{i_k}$. Let $\{|j_k\rangle\}_{j_k}$, $\{|\alpha_k\rangle\}_{\alpha_k}$ and $\{|\beta_k\rangle\}_{\beta_k}$ be further onbs of the Hilbert space associated to site k. Without loss of generality, we may assume that the latter two are identical, and hence $\langle\alpha_k|\beta_k\rangle = \delta_{\alpha_k,\beta_k}$, since otherwise we introduce the matrices $A^{[k]}_{\alpha_k,\beta_k} = \langle\beta_k|\alpha_k\rangle$ which represent the change

$$|\dots,\alpha_k,\dots\rangle \to |\dots,\beta_k,\dots\rangle$$

and can be absorbed in the matrix product operator representation of the Kraus operator $M_{k,k+1}$.
The mps representation of a pure state $|\varphi_{1,2}\rangle\langle\varphi_{1,2}| \in \mathscr{D}(\mathscr{H}_1 \otimes \mathscr{H}_2)$ is obtained by the standard procedure [Vido3]. From the Schmidt decomposition and a unitary change of basis, we easily see that

$$|\varphi_{1,2}\rangle = \sum_{i_1,\alpha_2,\beta_1} A^{[1]}_{i_1,\beta_1} A^{[2]}_{\beta_1,\alpha_2} |i_1,\alpha_2\rangle = \sum_{i_1,\alpha_2} \xi_{i_1,\alpha_2} |i_1,\alpha_2\rangle .$$

Observe that $\sum_{\beta_1} A^{[1]}_{i_1,\beta_1} A^{[2]}_{\beta_1,\alpha_2}$ is the product of a row and column vector which is abbreviated by ξ_{i_1,α_2} in order to enhance readability. The bond dimension of the mps is $D \leqslant \min\{\dim\mathscr{H}_1, \dim\mathscr{H}_2\} = d$ (*cf.* theorem 1.10).
The expansion of the Kraus operator $M_{2,3}$ in the chosen basis reads

$$M_{2,3} = \sum_{i_2,j_3,\alpha_2,\beta_3} {}^{[2,3]}M^{i_2,\beta_3}_{\alpha_2,j_3} |i_2,\beta_3\rangle\langle\alpha_2,j_3| .$$

Let $|j_3\rangle \in \{|j_3\rangle\}_{j_3}$ be an arbitrary member of the onb. The actual choice is of no particular interest since we will sum over j_3 later on. For $N_2(|\varphi_{1,2}\rangle \otimes |j_3\rangle)$ we obtain the expression

$$\begin{aligned} N_2(|\varphi_{1,2}\rangle \otimes |j_3\rangle) &= (\mathbb{1}_1 \otimes M_{2,3})(|\varphi_{1,2}\rangle \otimes |j_3\rangle) \\ &= \sum_{i_1,i_2,\alpha_2,\beta_3} \xi_{i_1,\alpha_2} {}^{[2,3]}M^{i_2,\beta_3}_{\alpha_2,j_3} |i_1,i_2,\beta_3\rangle . \end{aligned}$$

Let us investigate the next step in detail such that afterwards we got familiarized with this expansion and are able to write down the whole mps. Again, we start with the expansion of the Kraus operator M_{34}

$$M_{3,4} = \sum_{i_3,j_4,\alpha_3,\beta_4} {}^{[3,4]}M^{i_3,\beta_4}_{\alpha_3,j_4} |i_3,\beta_4\rangle\langle\alpha_3,j_4|$$

and pick $|j_4\rangle \in \{|j_4\rangle\}_{j_4}$ to find

$$\begin{aligned} N_3(N_2(|\varphi_{1,2}\rangle \otimes |j_3\rangle \otimes |j_4\rangle)) &= (\mathbb{1}_{1,2} \otimes M_{3,4})\big((\mathbb{1}_1 \otimes M_{2,3})(|\varphi_{1,2}\rangle \otimes |j_3\rangle \otimes |j_4\rangle)\big) \\ &= \sum_{\substack{i_1,i_2,i_3,\\ \alpha_2,\beta_3,\beta_4}} \xi_{i_1,\alpha_2} {}^{[2,3]}M^{i_2,\beta_3}_{\alpha_2,j_3} {}^{[3,4]}M^{i_3,\beta_4}_{\beta_3,j_4} |i_1,i_2,i_3,\beta_4\rangle . \end{aligned}$$

By iterating the procedure, we finally arrive at the total system's state

$$\begin{aligned}
&(N_{N-1}\circ N_{N-2}\circ\cdots\circ N_3\circ N_2)(|\varphi_{1,2}\rangle\otimes|j_3\rangle\otimes\cdots\otimes|j_N\rangle)\\
&=\sum_{\substack{i_1,\ldots,i_{N-1},\\ \alpha_2,\beta_3,\cdots,\beta_N}} \xi_{i_1,\alpha_2}\,{}^{[2,3]}M^{i_2,\beta_3}_{\alpha_2,j_3}\cdots{}^{[N-1,N]}M^{i_{N-1},\beta_N}_{\beta_{N-1},j_N}\,|i_1,\ldots,i_{N-1},\beta_N\rangle\\
&=\sum_{\substack{i_1,\ldots,i_{N},\\ \alpha_2,\beta_3,\cdots,\beta_N}} \xi_{i_1,\alpha_2}\,{}^{[2,3]}M^{i_2,\beta_3}_{\alpha_2,j_3}\cdots{}^{[N-1,N]}M^{i_{N-1},\beta_N}_{\beta_{N-1},j_N}A^{[N]}_{\beta_N,i_N}\,|i_1,\ldots,i_{N}\rangle\,. \qquad (5.6)
\end{aligned}$$

The last equality follows by introducing the matrix $A^{[N]}_{\beta_N,i_N}=\langle i_N|\beta_N\rangle$ which transforms the onb $\{\beta_N\}_{\beta_N}$ to the computational basis. Translating the construction of the mps into the graphical notation yields the diagram depicted in figure 5.2. Abbreviating $N_{\text{tot}}=(N_{N-1}\circ\cdots\circ N_2)$ and decomposing $\mathbb{1}_{3,\ldots,N}=\sum_{j_3,\ldots,j_N}|j_3,\ldots,j_N\rangle\langle j_3,\ldots,j_N|$, we find

$$\begin{aligned}
\rho_{1,\ldots,N}&=N_{\text{tot}}\left(\left|\varphi_{1,2}\right\rangle\left\langle\varphi_{1,2}\right|\otimes\mathbb{1}_{3,\ldots,N}\right)N^{\dagger}_{\text{tot}}\\
&=\sum_{j_3,\ldots,j_N}N_{\text{tot}}\left(\left|\varphi_{1,2},j_3,\ldots,j_N\right\rangle\left\langle\varphi'_{1,2}j_3,\ldots,j_N\right|\right)N^{\dagger}_{\text{tot}}. \qquad (5.7)
\end{aligned}$$

Unfortunately, as we have seen under example 1.4, forming the partial trace of a pure state yields a mixed one in general. Consequently, we are required to generalize our results obtained under the simplifying assumption $\sigma_{1,2}=|\varphi_{1,2}\rangle\langle\varphi_{1,2}|$. In the generic case, our starting state $\sigma_{1,2}$ is of the form

$$\sigma_{1,2}=\sum_i\lambda_i\left|\varphi^{(i)}_{1,2}\right\rangle\left\langle\varphi^{(i)}_{1,2}\right|.$$

Nevertheless exploiting the linearity of the recovery map, we easily arrive at

$$\begin{aligned}
\rho_{1,\ldots,N}&=\sum_i\lambda_i N_{\text{tot}}\left(\left|\varphi^{(i)}_{1,2}\right\rangle\left\langle\varphi^{(i)}_{1,2}\right|\otimes\mathbb{1}_{3,\ldots,N}\right)N^{\dagger}_{\text{tot}}\\
&=\sum_{i,j_3,\ldots,j_N}\lambda_i N_{\text{tot}}\left(\left|\varphi^{(i)}_{1,2},j_3,\ldots,j_N\right\rangle\left\langle\varphi^{(i)}_{1,2},j_3,\ldots,j_N\right|\right)N^{\dagger}_{\text{tot}}. \qquad (5.8)
\end{aligned}$$

instead of equation (5.7).

It is worthwhile noting that the preceding construction can also be carried out with a three index tensor representation of the Kraus operators. To this end, observe that the Kraus representation of the recovery map from equation (5.4) can be reformulated as follows:

$$\mathscr{R}_{B\to BC}(X_B)=M_{BC}(X_B\otimes\mathbb{1}_C)M^{\dagger}_{BC}=\sum_{j=1}^{\dim(\mathscr{H}_C)}M^{(j)}_{BC}X_BM^{(j)\dagger}_{BC}$$

where we defined $M^{(j)}_{BC}:\mathscr{B}(\mathscr{H}_B)\to\mathscr{B}(\mathscr{H}_B\otimes\mathscr{H}_C)$, $M^{(j)}_{BC}=M_{BC}(\mathbb{1}_B\otimes|j\rangle)$ for some onb $\{|j\rangle\}_j\subset\mathscr{H}_C$. Consequently, the expansion of the Kraus operator $M^{(j_s)}_{s-1,s}$, $s=3,\ldots,N$, reads

$$M^{(j_s)}_{s-1,s}=\sum_{i_{s-1},\alpha_{s-1},\beta_s}{}^{[s-1,s],(j_s)}M^{i_{s-1},\beta_s}_{\alpha_{s-1}}\,|i_{s-1},\beta_s\rangle\langle\alpha_{s-1}|\,.$$

Since the Kraus operators depend on the index j_s, our notation needs to be adapted. We define $N_{s-1}^{(j_s)} = \mathbb{1}_{1,\dots,s-2} \otimes M_{s-1,s}^{(j_s)}$, $s = 3,\dots,N$, and obtain the three index analogue of equation (5.6):

$$\begin{aligned}&\left(N_{N-1}^{(j_N)} \circ \cdots \circ N_2^{(j_3)}\right)(|\varphi_{1,2}\rangle)\\ &= \sum_{i_1,\dots,i_N} \xi_{i_1,\alpha_2}\, {}^{[2,3],(j_3)}M_{\alpha_2}^{i_2,\beta_3} \dots {}^{[N-1,N],(j_N)}M_{\beta_{N-1}}^{i_{N-1},\beta_N} A_{\beta_N,i_N}^{[N]} |i_1,\dots,i_N\rangle\,.\end{aligned}$$

Consequently, abbreviating $N_{\text{tot}}^{(j_3,\dots,j_N)} = N_{N-1}^{(j_N)} \circ \cdots \circ N_2^{(j_3)}$ yields

$$\rho_{1,\dots,N} = \sum_{i,j_3,\dots,j_N} \lambda_i N_{\text{tot}}^{(j_3,\dots,j_N)} \left|\varphi_{1,2}^{(i)}\right\rangle \left\langle\varphi_{1,2}^{(i)}\right| N_{\text{tot}}^{(j_3,\dots,j_N)\dagger},$$

instead of equation (5.8).

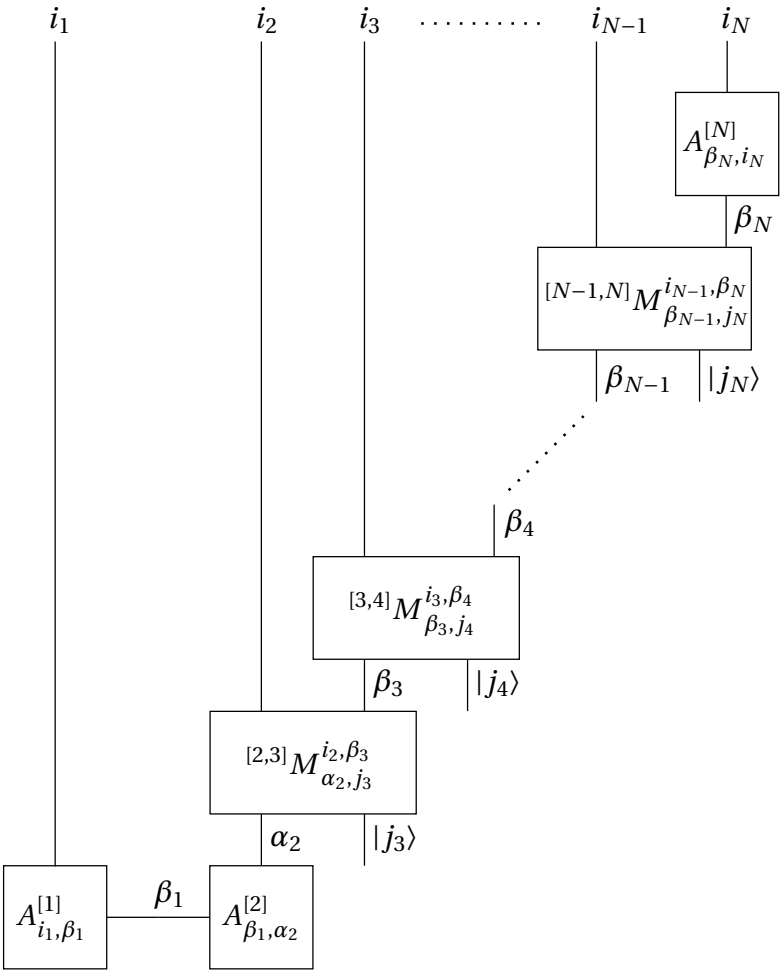


Figure 5.2: The figure translates the preceding calculation in the four index formulation into the graphical language introduced in § 3.2.1. The notation $|j_k\rangle$, $k = 3,\dots,N$, for the second incoming index of the tensor ${}^{[k-1,k]}M^{i_{k-1}\beta_k}_{\alpha_{k-1},j_k}$ indicates that this index takes only a single value fixed by $|j_k\rangle$ and over which we sum up in equation (5.8).

The reader might have noticed that the two approaches show only minor differences. In particular, the two mps representations have the same bond dimension given by

$$D = d^{\gamma} \tag{5.9}$$

where γ denotes the largest number of sites blocked together before the recovery map was applied. For instance, let us suppose we arrived at the state $\rho_{1,\dots,10}$. Since the result of Fawzi and Renner does not impose any constraint on the system $\mathscr{H}_B$ and $\mathscr{H}_C$, we are free to block several sites together. We choose $\mathscr{H}_B = \mathscr{H}_4 \otimes \cdots \otimes \mathscr{H}_{10}$ and $\mathscr{H}_C = \mathscr{H}_{11} \otimes \mathscr{H}_{12}$ to find $\gamma = 7$. Clearly, in this case the construction has to be slightly adapted since we only considered the reconstruction of one site by acting on the adjacent one in order to ensure the readability of our expressions. In § 5.2 we further investigate the dependence between bond dimension D and system size N.
In the further discussion, we use the four index formulation. Recall from (5.6) that we can represent $N_{\text{tot}}\big(|\varphi^{(i)}_{1,2}, j_3, \dots, j_N\rangle\big)$ as mps. However, both, (5.7) and (5.8), provide a mixed approximation of the pure state $\sigma_{1,\dots,N} = |\Psi\rangle\langle\Psi|$ and are consequently not of the required mps structure. To resolve this dilemma, we would like to pick one of the pure states in the decomposition (5.8) (in the generic case) which lays nearby the state we aimed to approximate. As the following two lemmas show, there is at least one choice of $(i, j_3, \dots, j_N)$ such that the corresponding vector has the desired property. The first lemma connects the trace norm and the Hilbert–Schmidt norm of pure states and is needed in order to prove the second one.

Lemma 5.2. *Let $\rho, \sigma \in \mathscr{D}(\mathscr{H})$ be pure states, then*

$$\|\rho - \sigma\|_1 = \sqrt{2}\|\rho - \sigma\|_2.$$

Proof. Let $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\varphi\rangle\langle\varphi|$. Hence, the Hilbert–Schmidt norm reads

$$\begin{aligned}\|\rho - \sigma\|_2^2 &= \| \, |\psi\rangle\langle\psi| - |\varphi\rangle\langle\varphi| \, \|_2^2 \\ &= \operatorname{tr}\left[|\psi\rangle\langle\psi| + |\varphi\rangle\langle\varphi| - \langle\psi|\varphi\rangle\,|\psi\rangle\langle\varphi| - \langle\varphi|\psi\rangle\,|\varphi\rangle\langle\psi|\right] \\ &= 2\left(1 - |\langle\psi|\varphi\rangle|^2\right)\end{aligned}$$

where the last equality follows immediately by a suitable chosen onb in order to calculate the trace. For the trace distance, we have (*cf.* the derivation of the first Fuchs–van de Graaf inequality (proposition 1.29)):

$$\|\rho - \sigma\|_1 = 2\sqrt{1 - |\langle\psi|\varphi\rangle|^2} = \sqrt{2}\|\rho - \sigma\|_2.$$

□

Lemma 5.3. *Let $\rho, \sigma \in \mathscr{D}(\mathscr{H})$ be two density operators where σ is a pure state and $\rho = \sum_i \lambda_i \rho_i$ is a mixture of pure states ρ_i. Moreover, let $\varepsilon > 0$ such that $\|\sigma - \rho\|_1 \leqslant \varepsilon$, then there exists at least one ρ_k from the convex decomposition of ρ such that*

$$\|\sigma - \rho_k\|_1 \leqslant 2\sqrt{\varepsilon}.$$

Proof. The lemma is proven by deducing a contradiction. Assume that for all i, we have that $\|\sigma - \rho_i\|_1 > 2\sqrt{\varepsilon}$ and hence, $\|\sigma - \rho_i\|_2 > \sqrt{2\varepsilon}$ according to lemma 5.2. Restricted to the set of Hermitian operator, the Hilbert–Schmidt inner product is symmetric. Indeed, we find for $A, B \in \mathfrak{M}_n$ Hermitian that

$$\langle A \,|\, B \rangle_{HS} = \operatorname{tr}\left[A^\dagger B\right] = \operatorname{tr}\left[B^\dagger A\right] = \langle B \,|\, A \rangle_{HS}.$$

Hence, the identity $\langle A|B\rangle_{HS} = \frac{1}{2}\left(\|A\|_2^2 + \|B\|_2^2 - \|A - B\|_2^2\right)$, well–known from real Euclidean vector spaces, holds true in the case of interest. Plugging the pure states ρ_i and σ in, we obtain $\langle \sigma|\rho_i\rangle < 1 - \varepsilon$ for any i. Invoking the linearity of the scalar product, we deduce

$$\langle \sigma|\rho\rangle = \sum_i p_i \langle \sigma|\rho_i\rangle < 1 - \varepsilon.$$

Observe that the characterization of the operator norm using singular values (*cf.* the proof of proposition 1.21) immediately implies $\|\sigma\|_{\mathrm{op}} \leqslant 1$. Due to Hölder's inequality (*cf.* the proof of theorem 1.22) for inner products

$$|\langle A|B\rangle_{HS}| = \left|\operatorname{tr}\left[A^\dagger B\right]\right| \leqslant \|A\|_{\mathrm{op}} \|B\|_1,$$

we eventually conclude that

$$\|\sigma - \rho\|_1 \geqslant \|\sigma\|_{\mathrm{op}} \|\sigma - \rho\|_1 \geqslant \langle \sigma|\sigma - \rho\rangle_{HS} > 1 - (1 - \varepsilon) = \varepsilon.$$

Consequently, we have showed that the assumption $\|\sigma - \rho_i\|_1 > 2\sqrt{\varepsilon}$ for any i leads to a contradiction. □

We are now able to sum the error up which was made during the mps approximation of the total system's state $|\Psi\rangle$ through successive application of the recovery map. For simplicity, let us assume that the quantum conditional mutual information remains constant throughout the spin chain, that is, $I(s : s+2|s+1) = \varepsilon$, $s = 1, \ldots, N-2$. Proposition 5.1 yields an upper bound on the inevitable error of

$$\|\rho_{1,\ldots,N} - \sigma_{1,\ldots,N}\|_1 \leqslant 2\sqrt{\ln 2}(N-2)\sqrt{\varepsilon}.$$

Due to lemma 5.3 and equation (5.8), there exists a vector $|\varphi_{1,2}^{(i)}, j_3, \ldots, j_N\rangle$ in the convex combination of

$$\sigma_{1,2} \otimes \mathbb{1}_{3,\ldots,N} = \sum_{i, j_3, \ldots, j_N} \lambda_i \left|\varphi_{1,2}^{(i)}, j_3, \ldots, j_N\right\rangle \left\langle \varphi_{1,2}^{(i)}, j_3, \ldots, j_N\right|$$

such that

$$\begin{aligned}&\left\| N_{\text{tot}} \left(\left| \varphi_{1,2}^{(i)}, j_3, \ldots, j_N \right\rangle \left\langle \varphi_{1,2}^{(i)}, j_3, \ldots, j_N \right| \right) N_{\text{tot}}^{\dagger} - |\Psi\rangle\langle\Psi| \right\|_1 \\ &\leqslant 2^{3/2} \sqrt{\sqrt{\ln 2}(N-2)\sqrt{\varepsilon}}.\end{aligned}$$

Since the vector $N_{\text{tot}}\big(|\varphi_{1,2}^{(i)}, j_3, \ldots, j_N\rangle\big)$ is represented exactly by (5.6), we have constructed a mps approximation $|\tilde{\Psi}\rangle = \big(N_{\text{tot}}\,|\varphi_{1,2}^{(i)}, j_3, \ldots, j_N\rangle\big)$ of the system's state $\sigma_{1,\ldots,N} = |\Psi\rangle\langle\Psi|$ fulfilling

$$\| \, |\Psi\rangle\langle\Psi| - |\tilde{\Psi}\rangle\langle\tilde{\Psi}| \, \|_1 \leqslant 2^{3/2} \sqrt{\sqrt{\ln 2}(N-2)\sqrt{\varepsilon}}. \tag{5.10}$$

Eventually, we want to note that the scaling behavior of the error in the system size coincides with the result of Verstraete and Cirac [VC06, Lemma 1]. Observe that they used the Euclidean norm to express the distance between $|\Psi\rangle$ and $|\tilde{\Psi}\rangle$ which upper bounds the trace norm via

$$\| \, |\Psi\rangle\langle\Psi| - |\tilde{\Psi}\rangle\langle\tilde{\Psi}| \, \|_1 \leqslant \sqrt{2} \, \| \, |\Psi\rangle - |\tilde{\Psi}\rangle \, \|.$$

5.2 Connection between Mutual Information and Bond Dimension

As we have already mentioned in § 3.1, Hastings [Has07] proved an one–dimensional area law for ground states of gapped Hamiltonians. In this setting, the area law boils down to the statement that the entanglement entropy $S(A)$ between a distinguished region A and the rest of the chain $V \setminus A$ is uniformly bounded by a constant independent of the size of this region and the length of the whole chain. In his paper, Hastings conjectured that a similar behavior holds true for the ground states of gapless Hamiltonians. Brandao and Horodecki were able to show that an exponential decay of correlations implies an one–dimensional area law [BH13; BH15]. Since Hastings proved that in any dimension the ground state of a gapped Hamiltonian exhibits exponential decay, their result give another proof of the one–dimensional area law. Nevertheless, we lack a rigorous proof of a two– or even higher–dimensional area and also in one dimension nothing is known about how the entanglement entropy of the distinguished region saturates towards the proven bound. Furthermore, it is unclear whether this saturation is monotonically. We present two different settings, exponential and polynomial saturation, and investigate the dependence between the quantum conditional mutual information and the size of the subsystem.

Exponential Saturation

Let A be a distinguished region of the spin chain and L_A denote its length. We make the ansatz

$$S(A) = S_{\max} - \eta e^{-L_A/\xi}, \qquad \xi > 0 \tag{5.11}$$

for the entropy of the reduced density operator ρ_A. Observe that the constant $\eta > 0$ in (5.11) only encodes information about the unit and is of no particular interest. Doing the algebra, one obtains the quantum conditional mutual information

$$\begin{aligned} I(A:C|B)_\rho &= S(A,B) + S(B,C) - S(B) - S(A,B,C) \\ &= \eta e^{-\frac{L_A+L_B+L_C}{\xi}} \left(e^{L_A/\xi} - 1\right)\left(e^{1/\xi} - 1\right) \\ &\leqslant \tilde{\eta} e^{-2\ell/\xi} \end{aligned}$$

where we defined $\tilde{\eta} = \eta\left(e^{1/\xi} - 1\right)$ and assumed $\ell = L_B = L_C$ for simplicity. Hence, we deduce from (5.10)

$$\| \left|\tilde{\Psi}\right\rangle \left\langle\tilde{\Psi}\right| - |\Psi\rangle \langle\Psi| \|_1 \leqslant 2\sqrt{\sqrt{\ln 2}\tilde{\eta} N} e^{-\frac{\ell}{2\xi}} =: \varepsilon_{\text{tot}} \tag{5.12}$$

as upper bound on the error of the approximation. Inserting the bond dimension of the mps representation of $|\tilde{\Psi}\rangle$ from equation (5.9) into (5.12), we find

$$D \sim \left(\frac{N}{\varepsilon_{\text{tot}}^2}\right)^{\xi \ln d}$$

and hence, $D \sim \text{poly}(N)$.

Polynomial Saturation

Again, we assume $L_B = L_C = \ell$ and let

$$S(A) = S_{\max} - \frac{\eta}{L_A^\xi}, \qquad \xi > 0$$

where $\eta > 0$ contains the information about the unit. We find a quantum conditional mutual information of

$$I(A:C|B)_\rho = \eta\left(\frac{1}{(L_A+\ell)^\xi} + \frac{1}{(2\ell)^\xi} - \frac{1}{\ell^\xi} - \frac{1}{(L_A+2\ell)^\xi}\right) \leqslant \frac{\eta}{(2\ell)^\xi}$$

and whence

$$\| \left|\tilde{\Psi}\right\rangle \left\langle\tilde{\Psi}\right| - |\Psi\rangle \langle\Psi| \|_1 \leqslant 2\sqrt{\sqrt{\ln 2}\eta N} \frac{1}{(2\ell)^{\xi/2}} = \varepsilon_{\text{tot}}.$$

Consequently, the bond dimension scales exponentially in the system size, namely

$$D \sim \exp\left(\ln d \frac{N^{2/\xi}}{\varepsilon_{\text{tot}}^{4/\xi}}\right).$$

To conclude this section, we turn the question upside down. Which saturation behavior of the entropy is necessary in order to achieve a quasi–polynomial scaling of the bond dimension with the system size, that is, $D \sim$

$\exp\big(\mathrm{poly}(\ln N)\big)$? From equation (5.12), we easily see that $I(A:C|B)_\rho \sim e^{-\ell^{1/\zeta}}$, $\zeta > 0$, is required to ensure

$$D \sim \exp\left(\ln(d)\ln^{\zeta}\left(\frac{\sqrt{N}}{\varepsilon_{\mathrm{tot}}}\right)\right).$$

Since an efficient description of a state in the matrix product formalism requires a polynomial, or at most quasi–polynomial, scaling behavior of the bond dimension, it is crucial that the quantum conditional mutual information satisfies $I(A:C|B)_\rho \lesssim e^{-\ell^{1/\zeta}}$ for some reasonably small ζ.

Conclusion

In the course of this thesis, the reader hopefully became aware of the highly non–trivial interplay of two major areas from the apparently different fields of physics and computer science, namely quantum many body physics and information theory. Their connection is essentially established by one of the most important theorems of modern analysis, the spectral theorem.

We aimed at giving a comprehensive introduction to both areas in the first two chapters and brought them together in order to study matrix product states in the third one. Chapter 4 gave a review of a selection of related work on the recovery map acting on a tripartite system.

These works motivated this thesis and provided the groundwork our own results, presented in chapter 5, are based upon. The latter consist of an explicit construction of a mps approximation of the state $\sigma_{1,\dots,N}$ describing a spin chain of N sites provided $\sigma_{1,\dots,N}$ is pure. We applied the results of Fawzi, Renner and Sutter to an iterative reconstruction process starting with the state of the first two sites $\sigma_{1,2} = \mathrm{tr}_{3,\dots,N}[\sigma_{1,\dots,N}]$ and we were able to rigorously establish an upper bound on the defect our mps representation may carry. To this end, we proved proposition 5.1 which investigated the error propagation through the recovery process based on the bound of Fawzi et al. (theorem 4.5).

In order to achieve a good approximation of the system's state $\sigma_{1,\dots,N}$, we exploited the geometric properties of convex combinations of unit vectors in inner product spaces. The resulting lemma 5.3 was the key to overcome the issues arisen in the equations (5.7) and (5.8) respectively, and formed the decisive step of our argument. This lemma ensures that there exists at least one index combination $(i, j_3, \dots, j_N)$ such that $N_{\mathrm{tot}}\big(|\varphi^{(i)}_{1,2}, j_3, \dots, j_N\rangle\big)$ from the mixed approximation of the pure state $\sigma_{1,\dots,N}$ is a mps representation of the latter which may only carry a minor error.

In § 5.2 we have seen that a rapid saturation of the von–Neumann entropy towards $S_{\max}$, namely fast enough to ensure the bound $I(A:C|B)_\rho \lesssim \exp\left(-\ell^{1/\zeta}\right)$ for some $\zeta > 0$, is crucial to achieve a quasi–polynomial scaling of the bond dimension D in the system size N and thus in case of a reasonably small ζ, an efficient approximation of the system's state.

An interesting starting point of further work might be a generalization of lemma 5.3 in the following way: If we interpreted the the coefficients $\{\lambda_i\}$ of ρ's convex combination in pure states as probability distribution, we may draw the ρ_k according to this mass function. In this case, a probabilistic

upper bound on the error would be an appealing result since it provides further information about the construction of the mps representation $|\tilde{\Psi}\rangle$ of $\sigma_{1,\ldots,N} = |\Psi\rangle\langle\Psi|$.

Appendix A

Miscellanea

A.1 Measurements in Quantum Mechanics

We start this chapter with a brief introduction to positive operator valued measures which are used in appendix A.1.2 to revisit the ambiguous state discrimination mentioned in § 1.1 and to derive the result of Helstrom in full mathematical rigor. For further information about the measurement in quantum mechanics we refer the reader to the textbook of Heinosaari and Ziman [HZ11].

A.1.1 Positive Operator Valued Measures

The mathematical concept used for the proper description of the measurement in quantum mechanics is a generalization of abstract measure theory which we want to develop in the following. We denote for an arbitrary set Ω whose elements are the possible measurement outcomes ($\leftrightarrow$ sample space in probability theory) the Borel sets over Ω by $\mathfrak{B}(\Omega)$ ($\leftrightarrow$ set of events). We choose $\mathscr{F} = \mathscr{P}(\Omega)$ for a countable Ω and $\mathscr{F} = \mathfrak{B}(\Omega)$ for $\Omega \subset \mathbb{R}^n$ as $\sigma-$algebra.

Definition A.1. Every measurement is described by a **positive operator valued measure** (POVM) which is a mapping $M : \mathscr{F} \to \mathscr{B}(\mathscr{H})$ such that:

(i) $\forall B \in \mathscr{F} : M(B) \geqslant 0$.

(ii) Suppose $(B_n)_{n\in\mathbb{N}} \subset \mathscr{F}$ with $\bigcup_{n\in\mathbb{N}} B_n = \Omega$ and $B_k \cap B_\ell = \emptyset$ for all $k \neq \ell$ (note that non-emptiness is not required and hence $M(\emptyset) = 0$), then $\sum_{n\in\mathbb{N}} M(B_n) = \mathbb{1}$.

In analogy to classical measure theory, we are able to prove monotonicity and finite (sub-)additivity for POVM, too. It is convenient to introduce the following partial order on the set of bounded operators. For $A, B \in \mathscr{B}(\mathscr{H})$ we write $A \geqslant B$ if $A - B \geqslant 0$.

Lemma A.2. *For a POVM $M : \mathscr{F} \to \mathscr{B}(\mathscr{H})$ and $A, B \in \mathscr{F}$ the following statements hold:*

(i) $A \subset B$ *implies* $M(B \setminus A) = M(B) - M(A)$ *and* $M(A) \leqslant M(B)$.

(ii) $M(A \cup B) \leqslant M(A) + M(B)$ *and* $M(A \cup B) = M(A) + M(B)$ *if* $A \cap B = \emptyset$.

(iii) $M(A \cup B) + M(A \cap B) = M(A) + M(B)$.

Proof. Ad (i): Since $\Omega = (B \setminus A) \dot\cup A$ we get $\mathbb{1} = M(\Omega \setminus B) + M(B)$. The same argument yields $\mathbb{1} = M(A) + M(B \setminus A) + M(\Omega \setminus B)$. By subtracting these equations, we obtain $M(B) = M(B \setminus A) + M(A)$ and especially $M(B) - M(A) = M(B \setminus A) \geqslant 0$.
Ad (ii): We use exactly the same technique as in classical measure theory:

$$M(A \cup B) = M(A) + M((A \cup B) \setminus A) \overset{(i)}{\leqslant} M(A) + M(B).$$

Ad (iii): This point is also proved analogously as in basic measure theory. Note that $A \cup B = A \dot\cup [(A \cup B) \setminus A]$ and therefore we get $M(A \cup B) = M(A) + M((A \cup B) \setminus A)$. Since $[(A \cup B) \setminus A] \dot\cup (A \cap B) = B$, we conclude

$$\begin{aligned} M(A \cup B) + M(A \cap B) &= M(A) + M((A \cup B) \setminus A) + M(A \cap B) \\ &= M(A) + M(B). \end{aligned}$$

□

The reason why we introduced the abstract notion of POVM is the fact that we identify them with the **observables**.

Definition A.3. An observable M is called **sharp** if $M(B)^2 = M(B)$ (i.e. $M(B)$ is a projection) for every $B \in \mathscr{F}$.

Only the sharp observables are relevant in physical applications. Hence, some authors introduce POVM as projector valued measures but we choose a more general approach. Nevertheless, the rest of this chapter only deals with sharp observables.

Lemma A.4. *For an observable M the conditions*

(i) M *is sharp,*

(ii) $M(A)M(B) = M(A \cap B)\ \forall A, B \in \mathscr{F}$,

(iii) $M(A)M(A^c) = 0\ \forall A \in \mathscr{F}$

are equivalent.

Proof. In first place, we prove the equivalence of (i) and (iii). Let $A \in \mathscr{F}$, then we immediately get

$$M(A) - M(A)^2 = M(A)(\mathbb{1} - M(A)) = M(A)M(A^c)$$

and thus, $M(A) = M(A)^2$ if $M(A)M(A^c) = 0$ or $M(A)M(A^c) = 0$ if $M(A)^2 = M(A)$.
Now assume that M is sharp. Then we get $M(A)M(A \cap B) = [M(A \setminus (A \cap B)) + M(A \cap B)]M(A \cap B) = M(A \cap B)$ since $0 \leqslant M(A \setminus (A \cap B))M(A \cap B) \leqslant M((A \cap B)^c)M(A \cap B) = 0$. Analogously we find that $M(A)M(A \cup B) = M(A)$. Multiplying the inclusion-exclusion formula from lemma A.2 (iii) with $M(A)$ yields

$$M(A) + M(A \cap B) = M(A) + M(A)M(B)$$

which proves the assertion.
For the last implication consider $B = A^c$ and note that

$$M(A \cap A^c) = A(\emptyset) = A(\Omega \setminus \Omega) = \mathbb{1} - \mathbb{1} = 0.$$

□

The last lines of this section are intended to connect our abstract mathematical definitions with the physics of the quantum system. As mentioned at the beginning of § 1.1, we want to predict the outcomes which our experimenter may measure (at least probabilistically). To this end, we need to connect our concepts of observables with the probability of an outcome $B \in \mathscr{F}$. This connection is established by Born's rule which we present in the following postulate.

Postulate A.5 (Born's rule). Let $(\Omega, \mathscr{F})$ denote an outcome space. We describe the measurement with a sharp observable $M : \mathscr{F} \to \mathscr{B}(\mathscr{H})$ and the preparation with a density operator $\rho \in \mathscr{D}(\mathscr{H})$. Then for $B \in \mathscr{F}$ the conditional probability $\mathbb{P}(B|M, \rho)$ is given by

$$\mathbb{P}(B|M, \rho) = \operatorname{tr}[M(B)\rho].$$

To convince ourselves that the preceding formula defines indeed a probability measure, we check the Kolmogorov axioms briefly:

- $\mathbb{P}(B) \geqslant 0$ for all $B \in \mathscr{F}$ since $M(B), \rho \geqslant 0$ and the trace is monotone.
- $\mathbb{P}(\Omega) = \operatorname{tr}[M(\Omega)\rho] = \operatorname{tr}[\rho] = 1$.
- Let $(B_n)_{n\in\mathbb{N}} \subset \mathscr{F}$ pairwise disjoint sets. Then due to the continuity of the trace:

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \operatorname{tr}\left[M\left(\bigcup_{n\in\mathbb{N}} B_n\right)\rho\right] = \sum_{n\in\mathbb{N}} \operatorname{tr}[M(B_n)\rho] = \sum_{n\in\mathbb{N}} \mathbb{P}(B_n).$$

A.1.2 Binary Hypothesis Testing

To keep the following discussion simple, we only consider a finite sample space Ω of possible outcomes and use its power set $\mathscr{P}(\Omega)$ as σ–algebra. Moreover, the Hilbert space $\mathscr{H}$ has dimension d.
The first approach we make is trying to transfer our concepts from macroscopic world into quantum world. To this end, we consider a generalization of the thought experiment presented in § 1.2.1. Suppose we have an apparatus producing the states $\rho_1, \dots, \rho_n$ and our analyzer is able to identify the right state with full probability. The reader might think of drawing balls with different colors from an urn. Consequently, we choose the outcome space $\Omega = \{1, \dots, n\}$. Then we say the states are **perfectly discriminated** if there is an observable M such that

$$\operatorname{tr}[M(\{1\})\rho_1] = \cdots = \operatorname{tr}[M(\{n\})\rho_n] = 1.$$

Hence, in case of a perfectly discriminated set we are able to identify the states in a single-shot experiment.
A set of orthogonal states $\{\rho_1, \dots, \rho_n\}$ (i.e. $\operatorname{supp}\rho_i \perp \operatorname{supp}\rho_j$ for $i \neq j$ and hence, $\rho_i\rho_j = 0$) is always perfectly discriminated. To see this, we consider the spectral decompositions

$$\rho_1 = \sum_{j=1}^{d} \lambda_1^{(j)} \left|e_1^{(j)}\right\rangle \left\langle e_1^{(j)}\right|, \dots, \ \rho_n = \sum_{j=1}^{d} \lambda_n^{(j)} \left|e_n^{(j)}\right\rangle \left\langle e_n^{(j)}\right|.$$

Since the ρ_j's are orthogonal, the eigenvectors of the set of the non-vanishing eigenvalues

$$A = \left\{ \left|e_i^{(j)}\right\rangle \,\middle|\, \lambda_i^{(j)} > 0, i \in \{1, \dots, n\}, j \in \{1, \dots, d\} \right\}$$

is orthonormal. We expand this set by a collection of appropriate unit vectors B to an onb $A \cup B$ of $\mathscr{H}$. Now we can define the sharp observable

$$\begin{gathered} M : \mathscr{P}(\Omega) \to \mathscr{B}(\mathscr{H}) \\ \{1\} \mapsto \sum_{j:\lambda_1^{(j)}>0} \left|e_1^{(j)}\right\rangle \left\langle e_1^{(j)}\right| + \sum_{e\in B} |e\rangle\langle e| \\ \text{and for } i \in \{2, \dots, n\} : \{i\} \mapsto \sum_{j:\lambda_i^{(j)}>0} \left|e_i^{(j)}\right\rangle \left\langle e_i^{(j)}\right|. \end{gathered}$$

Eventually, for all $i \in \{1, \dots, n\}$:

$$\operatorname{tr}[M(\{i\})\rho_i] = \sum_{j:\lambda_i^{(j)}>0} \lambda_i^{(j)} = 1.$$

Theorem A.6. *Consider the pure states ρ_1 and ρ_2. They are perfectly discriminated if and only if they are orthogonal.*

Proof. The first implication is only special case of what we have already seen. Thus, we only need to prove the second assertion.
Therefore, let $\Omega = \{1,2\}$. Since ρ_1 and ρ_2 are perfectly discriminated, there is an observable M such that $\operatorname{tr}[M(\{1\})\rho_1] = \operatorname{tr}[M(\{2\})\rho_2] = 1$. Furthermore, the purity of the states implies that there are unit vector ψ_1, ψ_2 such that $\rho_1 = |\psi_1\rangle\langle\psi_1|$ and $\rho_2 = |\psi_2\rangle\langle\psi_2|$. Applying the Cauchy Schwarz inequality gives

$$1 = \operatorname{tr}[M(\{i\})\rho_i] \overset{(\star)}{=} \langle \psi_1 \,|\, M(\{i\})\psi_i \rangle \leqslant \|M(\{i\})\psi_i\| \leqslant \|M(\{i\})\| = 1.$$

Thus, we require equality in Cauchy Schwarz and hence $\psi = cM(\{i\})\psi$ for some $c \in \mathbb{C}$. The equality $(\star)$ shows $c = 1$ and hence $M(\{i\})\rho_i = \rho_i$. Note that $M(\{2\}) = \mathbb{1} - M(\{1\})$ and thus $M(\{2\})\rho_2 = \rho_2 - M(\{1\})\rho_2$ from which we conclude $M(\{1\})\rho_2 = 0$. Consequently, we compute

$$|\langle\psi_1|\psi_2\rangle|^2 = \operatorname{tr}[\rho_1\rho_2] = \operatorname{tr}[M(\{1\})\rho_1\rho_2] = \operatorname{tr}[\rho_1 M(\{1\})\rho_2] = 0$$

and thus $\psi_1 \perp \psi_2$. □

The preceding theorem shows that our naïve approach from macroscopic world fails when considering non-orthogonal states. Therefore, we have to develop weaker concepts to handle the issues caused by quantum mechanics.

Ambiguous State Discrimination

Recall the setup of the ambiguous discrimination of two states ρ_1 and ρ_2 occurring with a priori probabilities p_1 respectively $p_2 = 1 - p_1$ from § 1.2.1 and figure 1.2. We are not really free in the choice of our observable. Obviously, we have $M(\{2\}) = \mathbb{1} - M(\{1\})$ and the success probability reads

$$\mathbb{P}_{\text{success}} = p_1 \operatorname{tr}[M(\{1\})\rho_1] + p_2 \operatorname{tr}[M(\{2\})\rho_2] = \frac{1}{2}\underbrace{(1 - \operatorname{tr}[\Lambda])}_{=2p_2} + \operatorname{tr}[M(\{1\})\Lambda],$$

with $\Lambda := p_1\rho_1 - p_2\rho_2$. Clearly, Λ is Hermitian and thus the spectral decomposition implies that choosing

$$M(\{1\}) := \sum_{j:\lambda_j>0} \lambda_j \,|j\rangle\langle j|,$$

where λ_j denote the eigenvalues of Λ and $\{|j\rangle\}_j$ forms an orthonormal set of corresponding eigenvectors, maximizes the success probability. Eventually, the success probability is given by the following formula, known as **Helstrom's result**:

$$\mathbb{P}_{\text{success}} = \frac{1}{2}(1 + \operatorname{tr}[|\Lambda|]) = \frac{1}{2}\left(1 + \|p_1\rho_1 - p_2\rho_2\|_1\right).$$

A.2 Proof of Uhlmann's Theorem

In order to prove Uhlmann's theorem, we need the following lemma which provides a convenient characterization of the trace norm. We mainly follow the proof presented in the textbook of Nielsen and Chuang [NC10].

Lemma A.7. *Let $A \in \mathfrak{M}_n$ be a square complex matrix and $U \in \mathfrak{M}_n$ be unitary. Then*

$$|\operatorname{tr}[AU]| \leqslant \operatorname{tr}[|A|]$$

with equality being attained by choosing $U = V^\dagger$, where $A = V|A|$ is the polar decomposition of A.

Proof. First note that

$$|\operatorname{tr}[AU]| = |\operatorname{tr}[|A|VU]| = \left|\operatorname{tr}\left[\sqrt{|A|}\sqrt{|A|}VU\right]\right| = \left|\left\langle \sqrt{|A|}\middle|\sqrt{|A|}VU\right\rangle_{HS}\right|.$$

Thus due to Cauchy Schwarz, we find

$$|\operatorname{tr}[AU]| \leqslant \sqrt{\operatorname{tr}[|A|]\operatorname{tr}[U^\dagger V^\dagger |A| VU]} = \operatorname{tr}[|A|].$$

Clearly, equality is attained for the adjoint of the unitary matrix from the polar decomposition of A. □

Now we are able to prove Uhlmann's theorem.

Proof of Uhlmann's theorem. For onbs $\{e_j\}_{j=1}^d, \{e'_j\}_{j=1}^d$ of the d–dimensional Hilbert space $\mathscr{H}$ we define the vector $|\eta\rangle := \sum_{j=1}^d |e_j\rangle \otimes |e'_j\rangle$.The Schmidt decomposition reveals that the purifications of ρ and σ are of the form

$$|\psi\rangle = (\sqrt{\rho}U_1 \otimes U_2)\,|\eta\rangle$$
$$|\varphi\rangle = (\sqrt{\sigma}V_1 \otimes V_2)\,|\eta\rangle$$

where $U_1, U_2, V_1, V_2 \in \mathfrak{M}_d$ denote unitary matrices which arise from the change of basis. Let us check briefly that $|\psi\rangle$ is in fact a purification of ρ. First notice that we trace out the correct density operator

$$\begin{aligned}\operatorname{tr}_B[|\psi\rangle\langle\psi|] &= \operatorname{tr}_B\left[\sum_{i,j=1}^d (\sqrt{\rho}U|e_i\rangle\langle e_j|U^\dagger\sqrt{\rho}) \otimes (|e'_i\rangle\langle e'_j|)\right] \\ &= \sum_{i=1}^d \sqrt{\rho}U|e_i\rangle\langle e_i|U^\dagger\sqrt{\rho} = \rho.\end{aligned}$$

Furthermore, we get $\operatorname{tr}[|\psi\rangle\langle\psi|] = \operatorname{tr}[\operatorname{tr}_B[|\psi\rangle\langle\psi|]] = \operatorname{tr}[\rho] = 1$ due to proposition 1.5.

Taking the scalar product of the two purifications gives

$$|\langle\psi|\varphi\rangle| = \left|\left\langle \eta \middle| \left(U_1^\dagger \sqrt{\rho}\sqrt{\sigma} V_1 \otimes U_2^\dagger V_2\right)\eta \right\rangle\right|. \tag{A.1}$$

Here we use the quite helpful formula $\langle\eta|(A\otimes B)\eta\rangle = \mathrm{tr}\left[AB^t\right]$ which follows with the unitary change of basis $Ue'_j = e_j$:

$$\langle\eta|(A\otimes B)\eta\rangle = \sum_{i,j=1}^{d} \langle e_i\otimes e'_i|(A\otimes B)|e_j\otimes e'_j\rangle = \sum_{i,j=1}^{d} \langle e_i|A|e_j\rangle\,\langle e'_i|B|e'_j\rangle\,.$$

Whilst the other side yields

$$\begin{aligned}\mathrm{tr}[AB^t] = \mathrm{tr}[AUB^tU^\dagger] &= \sum_{i,j=1}^{d} \langle e_i|A|e_j\rangle\,\langle e_j|UB^tU^\dagger|e_i\rangle \\ &= \sum_{i,j=1}^{d} \langle e_i|A|e_j\rangle\,\langle e'_j|B^t|e'_i\rangle\,.\end{aligned}$$

Since $\langle e'_i|B|e'_j\rangle$ is the ij-th element of the matrix B in the Basis $\{e'_j\}_{j=1}^d$, both sides coincide and the identity is proven.
The last step in the proof of Uhlmann's theorem is the application of this identity onto equation (A.1). We find

$$|\langle\psi|\varphi\rangle| = \mathrm{tr}\left[U_1^\dagger\sqrt{\rho}\sqrt{\sigma}V_1V_2^t\overline{U_2}\right] \leqslant \mathrm{tr}\left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right].$$

Note that $U := U_1^\dagger V_1 V_2^t \overline{U}_2$ ($U^\dagger = U_2^t\overline{V}_2V_1^\dagger U_1$ and $\overline{V}_2V_2^t = (V_2V_2^\dagger)^t = \mathbb{1}$) is unitary and hence we are allowed to apply lemma A.7. Equality is attained by choosing $U_1 = U_2 = V_2 = \mathbb{1}$ and $V_1 = V$ where $\sqrt{\rho}\sqrt{\sigma} = V|\sqrt{\rho}\sqrt{\sigma}|$. □

Remark A.8. Uhlmann's theorem also holds true if we fix a purification $|\psi\rangle$ of ρ and take the maximum over all purifications $|\varphi\rangle$ of σ.

A.3 Sketch of the Proof of Wilde's Recoverability Theorem

The proof of Wilde's recoverability theorem is basically obtained as corollary to an interpolation result of the Schatten p–norms. We start our discussion with a brief review of some concepts from complex and harmonic analysis and derive afterwards the aforementioned interpolation theorem. In appendix A.3.2, we give a sketch of the actual proof of Wilde's recoverability theorem which requires a simple result about the Rényi generalization of the relative entropy. Nevertheless, this lemma has a rather technical proof and hence, we refer the reader to the original work at the respective place of the discussion.

A.3.1 Some Result from Harmonic Analysis

We denote the standard strip in the complex plane by

$$S = \{z \in \mathbb{C} \,|\, 0 < \operatorname{Re} z < 1\}.$$

On S, the maximum modulus principle takes the following form.

Proposition A.9 (Hadamard Three–Lines Theorem). *Let $f : \overline{S} \to \mathbb{C}$ be a bounded and continuous function, which is holomorphic on S, such that $M(x) = \sup_{y\in\mathbb{R}} |f(x+iy)| > 0$ for all $x \in [0,1]$. Then we have*

$$\ln M(x) \leqslant (1-x)\ln M(0) + x \ln M(1).$$

Proof. We follow the classical proof presented in [RS75].
The function $g : \overline{S} \to \mathbb{C}$,

$$g(z) = f(z) M(0)^{z-1} M(1)^{-z}$$

is holomorphic on S and satisfies $|g(z)| \leqslant 1$ for $z \in \partial S$. If $g(z) \to 0$ as $|\operatorname{Im}(z)| \to \infty$, the maximum modulus principle immediately implies $|g(z)| \leqslant 1$ for all $z \in \overline{S}$. Since this does not necessarily hold in the generic case, we consider for $n \in \mathbb{N}$ the functions

$$g_n(z) = g(z) \exp\left(\frac{z^2-1}{n}\right).$$

For all $n \in \mathbb{N}$, they fulfill $g_n(z) \to 0$ as $|\operatorname{Im}(z)| \to \infty$. Hence, by the maximum modulus principle, we have $|g_n(z)| \leqslant 1$ for all $z \in \overline{S}$. Moreover, $\lim_{n\to\infty} g_n(z) = g(z)$ and thus $|g(z)| \leqslant 1$ for all $z \in \overline{S}$.
Eventually, the claim follows by taking the logarithm on both sides. □

In order to prove an interpolation result of the Schatten p–norms, the following extension of the Hadamard three–lines theorem is needed.

Theorem A.10 (Hirschman [Hir52]). *Let $f : \overline{S} \to \mathbb{C}$ be bounded, holomorphic on S and continuous on the boundary ∂S. Then for all $\theta \in (0,1)$*

$$\ln|f(\theta)| \leqslant \int_{-\infty}^{\infty} \alpha_\theta(t) \ln\left(f|(it)|^{1-\theta}\right) + \beta_\theta(t) \ln\left(|f(1+it)|^{\theta}\right) dt$$

holds where

$$\alpha_\theta(t) = \frac{\sin(\pi\theta)}{2(1-\theta)(\cosh(\pi t) - \cos(\pi\theta))},$$
$$\beta_\theta(t) = \frac{\sin(\pi\theta)}{2\theta(\cosh(\pi t) + \cos(\pi\theta))}.$$

Rather than presenting a proof of the theorem (an amenable exposition can be found in [Gra14]), we want to show that the functions α_θ and β_θ emerging in theorem A.10 represent in fact probability density functions. The non–negativity of the two maps is obvious and it remains to show that for $\theta \in (0,1)$

$$\int_{-\infty}^{\infty} \alpha_\theta(t)\, dt = \int_{-\infty}^{\infty} \beta_\theta(t)\, dt = 1.$$

By substituting $u = e^{\pi t}$, we obtain

$$\begin{aligned}
\frac{2\theta}{\sin(\pi\theta)} \int_{-\infty}^{\infty} \beta_\theta(t)\, dt &= \frac{2}{\pi} \int_0^{\infty} \frac{du}{u^2 + 2u\cos(\pi\theta) + 1} \\
&= \frac{2}{\pi} \int_0^{\infty} \frac{du}{(u+\cos(\pi\theta))^2 + \sin^2(\pi\theta)} \\
&= \frac{2}{\pi} \left[\frac{1}{\sin(\pi\theta)} \arctan\left(\frac{u+\cos(\pi\theta)}{\sin(\pi\theta)} \right) \right]_0^{\infty} \\
&= \frac{2\theta}{\sin(\pi\theta)}.
\end{aligned}$$

Exactly the same computation shows that α_θ is normalized, too. With these integrals at hand, Hadamard's three–lines theorem can be easily deduced from theorem A.10. The main ingredient of the proof of Wilde's recoverability theorem is an application of theorem A.10 in order to interpolate the Schatten p–norms.

Theorem A.11 (Interpolation of Schatten p–Norms). *Let $G : \overline{S} \to \mathfrak{M}_n$ be a matrix–valued, bounded function which is holomorphic on S and continuous on the boundary ∂S.*
For $\theta \in (0,1)$ and $p_0, p_1 \in [1,\infty]$ we define p_θ by

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}.$$

Then we have

$$\ln \|G(\theta)\|_{p_\theta} \leqslant \int_{-\infty}^{\infty} \alpha_\theta(t) \ln\left(\|G(it)\|_{p_0}^{1-\theta} \right) + \beta_\theta(t) \ln\left(\|G(1+it)\|_{p_1}^{\theta} \right) dt$$

with the probability density functions defined in theorem A.10.

Note that we say a matrix–valued function $z \mapsto G(z) = (g_{ij}(z))_{i,j}$ is holomorphic if this holds for any component function $z \mapsto g_{i,j}(z)$.

Proof of Theorem A.11. For $\theta \in (0,1)$ fixed, we denote by q_0, q_1, q_θ the Hölder conjudated exponents of p_0, p_1 and p_θ respectively. Due to the characterization of the Schatten p–norms from lemma 1.23, there exists a matrix $A \in \mathfrak{M}_n$, $\|A\|_{q_\theta} = 1$, such that $\mathrm{tr}[G(\theta)A] = \|G(\theta)\|_{p_\theta}$. Without loss of generality are allowed to omit the absolute value in the argument of the trace since we

can absorb the unitary emerging in the polar decomposition into A without changing its Schatten q_θ–norm. Let $A = U\Sigma V$ be the singular value decomposition of A and define the map

$$\begin{aligned} X : \overline{S} &\to \mathfrak{M}_n \\ z &\mapsto U\Sigma^{q_\theta\left(\frac{1-z}{q_0}+\frac{z}{q_1}\right)}V. \end{aligned}$$

Moreover, we introduce the bounded map

$$\begin{aligned} g : \overline{S} &\to \mathbb{C} \\ z &\mapsto \operatorname{tr}[G(z)X(z)] \end{aligned}$$

which is in addition holomorphic on S and continuous on the boundary ∂S. By theorem A.10, we get

$$\begin{aligned} \ln \|G(\theta)\|_{p_\theta} &= \ln |g(\theta)| && \text{(A.2)} \\ &\leqslant \int_{-\infty}^{\infty} \alpha_\theta(t) \ln\left(|g(it)|^{1-\theta}\right) + \beta_\theta(t) \ln\left(|g(1+it)|^{\theta}\right) dt. && \text{(A.3)} \end{aligned}$$

From Hölder's inequality (theorem 1.22), we have

$$\begin{aligned} |g(it)| &= |\operatorname{tr}[G(it)X(it)]| \leqslant \|X(it)\|_{q_0} \|G(it)\|_{p_0} = \|G(it)\|_{p_0}, \\ |g(1+it)| &= |\operatorname{tr}[G(1+it)X(1+it)]| \\ &\leqslant \|X(1+it)\|_{q_1} \|G(1+it)\|_{p_1} \\ &= \|G(1+it)\|_{p_1} \end{aligned}$$

where we used that $\|X(it)\|_{q_0} = \|X(1+it)\|_{q_1} = 1$ for all $t \in \mathbb{R}$. Inserting these bounds in equation (A.2) proves the claim. □

A.3.2 Sketch of the Proof

With these results from harmonic analysis at hand, it requires only a minor effort to sketch a proof of Wilde's theorem 4.3.

To start the proof, we let $U : \mathscr{H}_A \to \mathscr{H}_B \otimes \mathscr{H}_E$ denote the Stinespring dilation (*cf.* theorem 1.18) of the quantum channel $\mathcal{N} : \mathscr{H}_A \to \mathscr{H}_B$. In order to apply the interpolation theorem of the Schatten p–norms A.11, we define the bounded function

$$\begin{aligned} G : \overline{S} &\to \mathscr{B}(\mathscr{H}_B \otimes \mathscr{H}_E) \\ z &\mapsto \left(\left(\mathcal{N}(\rho)\right)^{z/2} \left(\mathcal{N}(\sigma)\right)^{-z/2} \otimes \mathbb{1}_E\right) U \sigma^{z/2} \rho^{1/2} \end{aligned}$$

which is holomorphic on S and continuous on the boundary ∂S. We now fix $\theta \in (0,1)$ and choose $p_0 = 2$, $p_1 = 1$ to obtain $p_\theta = \frac{2}{1+\theta}$ in theorem A.11.

Evaluating the operator–valued function G at the points given in the interpolation theorem yields

$$\begin{aligned}
\|G(\theta)\|_{2/1+\theta} &= \left\| \left(\left(\mathcal{N}(\rho)\right)^{\theta/2} \left(\mathcal{N}(\sigma)\right)^{-\theta/2} \otimes \mathbb{1}_E \right) U \sigma^{\theta/2} \rho^{1/2} \right\|_{2/1+\theta} \\
\|G(it)\|_2 &= \left\| \left(\left(\mathcal{N}(\rho)\right)^{it/2} \left(\mathcal{N}(\sigma)\right)^{-it/2} \otimes \mathbb{1}_E \right) U \sigma^{it/2} \rho^{1/2} \right\|_2 \\
&\leqslant \|\rho^{1/2}\|_2 \\
&= 1 \\
\|G(1+it)\|_1 &= \left\| \left(\left(\mathcal{N}(\rho)\right)^{(1+it)/2} \left(\mathcal{N}(\sigma)\right)^{-(1+it)/2} \otimes \mathbb{1}_E \right) U \sigma^{(1+it)/2} \rho^{1/2} \right\|_1 \\
&= \left\| \left(\left(\mathcal{N}(\rho)\right)^{1/2} \left(\mathcal{N}(\sigma)\right)^{-it/2} \left(\mathcal{N}(\sigma)\right)^{-1/2} \otimes \mathbb{1}_E \right) U \sigma^{1/2} \sigma^{it/2} \rho^{1/2} \right\|_1 \\
&= F\left(\rho, \left(\mathcal{U}_{\sigma,-t/2} \circ \mathcal{P}_{\sigma,\mathcal{N}} \circ \mathcal{U}_{\mathcal{N}(\sigma),t/2}\right)\left(\mathcal{N}(\rho)\right)\right) \\
&= F\left(\rho, {}^{\mathrm{rot}}\mathcal{P}^{t/2}_{\sigma,\mathcal{N}}\right)
\end{aligned}$$

and consequently we obtain

$$\begin{aligned}
&\ln \left\| \left(\left(\mathcal{N}(\rho)\right)^{\theta/2} \left(\mathcal{N}(\sigma)\right)^{-\theta/2} \otimes \mathbb{1}_E \right) U \sigma^{\theta/2} \rho^{1/2} \right\|_{2/1+\theta} \\
&\leqslant \theta \int_{-\infty}^{\infty} \beta_\theta(t) \ln\left(F\left(\rho, {}^{\mathrm{rot}}\mathcal{P}^{t/2}_{\sigma,\mathcal{N}}\right)\right) dt. \qquad \text{(A.4)}
\end{aligned}$$

The left–hand site of the previous equation is (apart from a prefactor) known as Rényi generalization of the relative entropy [BSW15; SBW15]. More precisely, this quantity is defined for $\alpha \in (0,1) \cup (1,\infty)$ via

$$\tilde{\Delta}_\alpha(\rho,\sigma,\mathcal{N}) = \frac{2\alpha}{\alpha-1} \ln \left\| \left(\left(N(\rho)\right)^{(1-\alpha)/2\alpha} \left(\mathcal{N}(\sigma)\right)^{-(1-\alpha)/2\alpha} \otimes \mathbb{1}_E \right) U \sigma^{(1-\alpha)/2\alpha} \rho^{1/2} \right\|_{2\alpha}.$$

It can be shown (*cf.* appendix A of the original work of Wilde [Wil15]) that

$$\lim_{\alpha\to 1} \tilde{\Delta}_\alpha(\rho,\sigma,\mathcal{N}) = S(\rho\|\sigma) - S\left(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)\right).$$

Hence, letting $\theta = (1-\alpha)/\alpha$ for $\alpha \in (1/2, 1)$ and using dominated convergence in (A.4) proves part (i) of the theorem.
Assertion (ii) follows by inserting the given recovery map.

Bibliography

[Aff+87] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki: "Rigorous results on valence-bond ground states in antiferromagnets". In: *Physical review letters* 59 (1987), p. 799.

[Aff+88] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki: "Valence bond ground states in isotropic quantum antiferromagnets". In: *Communications in Mathematical Physics* 115 (1988), pp. 253–304.

[BC17] J. C. Bridgeman and C. T. Chubb: "Hand-waving and Interpretive Dance: An Introductory Course on Tensor Networks". In: *J. Phys. A: Math. Theor.* 50 (2017).

[BH13] Fernando GSL Brandão and Michal Horodecki: "An area law for entanglement from exponential decay of correlations". In: *Nature Physics* 721 (2013).

[BH15] Fernando GSL Brandão and Michał Horodecki: "Exponential decay of correlations implies area law". In: *Communications in mathematical physics* 333.2 (2015), pp. 761–798.

[Bha09] R. Bhatia: *Positive Definite Matrices*. Princeton Series in Applied Mathematics. Princeton University Press, 2009.

[Bha96] R. Bhatia: *Matrix Analysis*. Graduate Texts in Mathematics. Springer New York, 1996.

[BSW15] M. Berta, K. P. Seshadreesan, and M. M. Wilde: "Rényi generalizations of quantum information measures". In: *Physical Review A* 91 (2015).

[Car10] E. Carlen: "Trace inequalities and quantum entropy: an introductory course". In: *Entropy and the quantum* 529 (2010), pp. 73–140.

[Cho75] M.-D. Choi: "Completely positive linear maps on complex matrices". In: *Linear Algebra and its Applications* 10.3 (1975), pp. 285–290.

[CT12] T.M. Cover and J.A. Thomas: *Elements of Information Theory*. Wiley, 2012.

[Dir25] P. A. M. Dirac: "The fundamental equations of quantum mechanics". In: *Proceedings of the Royal Society A* 109 (1925), pp. 642–653.

[ECP10] J. EISERT, M. CRAMER, and M. B. PLENIO: "Area laws for the entanglement entropy - a review". In: *Rev. Mod. Phys.* 82 (2010).

[Eis13] J. EISERT: "Entanglement and tensor network states". In: *Modeling and Simulation 3* (2013).

[EPR35] ALBERT EINSTEIN, BORIS PODOLSKY, and NATHAN ROSEN: "Can quantum-mechanical description of physical reality be considered complete?" In: *Physical review* 47.10 (1935), p. 777.

[FK94] SK FOONG and S KANNO: "Proof of Page's conjecture on the average entropy of a subsystem". In: *Physical review letters* 72.8 (1994), p. 1148.

[FR15] O. FAWZI and R. RENNER: "Quantum conditional mutual information and approximate Markov chains". In: *Communications in Mathematical Physics* 340 (2015), pp. 575–611.

[Gra14] L. GRAFAKOS: *Classical Fourier Analysis*. Graduate Texts in Mathematics. Springer New York, 2014.

[Has07] M. B. HASTINGS: "An Area Law for One Dimensional Quantum Systems". In: *JSTAT* (2007).

[Hay+04] P. HAYDEN, R. JOZSA, D. PETZ, and A. WINTER: "Structure of states which satisfy strong subadditivity of quantum entropy with equality". In: *Communications in mathematical physics* 246 (2004), pp. 359–374.

[Hir52] I. I. HIRSCHMAN: "A convexity theorem for certain groups of transformations". In: *Journal d'Analyse Mathématique* 2 (1952), pp. 209–218.

[HZ11] T. HEINOSAARI and M. ZIMAN: *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press, 2011.

[JW28] P. JORDAN and E. P. WIGNER: "Über das Paulische Äquivalenzverbot". In: Zeitschrift für Physik 9 (1928), pp. 631–651.

[Kle13] A. KLENKE: *Probability Theory: A Comprehensive Course*. Universitext. Springer London, 2013.

[Kle31] O. KLEIN: "Zur quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre". In: *Z. Phys.* 72 (1931).

[Led05] M. LEDOUX: *The Concentration of Measure Phenomenon*. Mathematical surveys and monographs. American Mathematical Society, 2005.

[Lin75] G. LINDBLAD: "Completely positive maps and entropy inequalities". In: *Communications in Mathematical Physics* 40 (1975), pp. 147–151.

[LR73] E. H. LIEB and M. B. RUSKAI: "Proof of the strong subadditivity of quantum-mechanical entropy". In: *Journal of Mathematical Physics* 14 (1973), pp. 1938–1941.

[NC10] M.A. Nielsen and I.L. Chuang: *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[Oru14] R. Orus: "A Practical Introduction to Tensor Networks: Matrix Product States and Projected Entangled Pair States". In: *Annals of Physics* 349 (2014), pp. 117–158.

[Pag93] Don N Page: "Average entropy of a subsystem". In: *Physical review letters* 71.9 (1993), p. 1291.

[Per+07] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac: "Matrix Product State Representations". In: *Quantum Inf. Comput.* 401 (2007).

[Pet86] D. Petz: "Sufficient subalgebras and the relative entropy of states of a von Neumann algebra". In: *Communications in mathematical physics* 105.1 (1986), pp. 123–131.

[Pet88] D. Petz: "Sufficiency of channels over von Neumann algebras". In: *The Quarterly Journal of Mathematics* 39.1 (1988), pp. 97–108.

[Pla33] M. Planck: *Wege zur physikalischen Erkenntnis: Reden und Vorträge*. S. Hirzel, 1933.

[RS75] M. Reed and B. Simon: *II: Fourier Analysis, Self-Adjointness*. Methods of Modern Mathematical Physics. Elsevier Science, 1975.

[RS81] M. Reed and B. Simon: *I: Functional Analysis*. Methods of Modern Mathematical Physics. Elsevier Science, 1981.

[Rus07] M. B. Ruskai: "Another Short and Elementary Proof of Strong Subadditivity of Quantum Entropy". In: *Reports on Mathematical Physics* 60 (2007), pp. 1–12.

[SBT17] D. Sutter, M. Berta, and M. Tomamichel: "Multivariate Trace Inequalities". In: *Communications in Mathematical Physics* 352 (2017), pp. 37–58.

[SBW15] K. P Seshadreesan, M. Berta, and M. M. Wilde: "Rényi squashed entanglement, discord, and relative entropy differences". In: *Journal of Physics A: Mathematical and Theoretical* 48 (2015).

[Sch11] U. Schollwöck: "The density-matrix renormalization group in the age of matrix product states". In: *Annals of Physics* 326.1 (2011), pp. 96–192.

[Sch13] N.t Schuch: "Condensed Matter Applications of Entanglement Theory". In: *Quantum Information Processing. Lecture Notes of the 44th IFF Spring School 2013* (2013). arXiv: 1306.5551v1 [quant-ph].

[Sch35] Erwin Schrödinger: "Discussion of probability relations between separated systems". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 31 (1935), pp. 555–563.

[Sch36] ERWIN SCHRÖDINGER: "Probability relations between separated systems". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 32 (1936), pp. 446–452.

[Sen96] SIDDHARTHA SEN: "Average entropy of a quantum subsystem". In: *Physical review letters* 77.1 (1996), p. 1.

[SFR16] D. SUTTER, O. FAWZI, and R. RENNER: "Universal recovery map for approximate Markov chains". In: *Proc. R. Soc. A* 472 (2016).

[Sim15] B. SIMON: *Operator Theory:* American Mathematical Society, 2015.

[Uhl76] A. UHLMANN: "The "Transition Probability" in the State Space of a *-Algebra". In: *Rep. Math. Phys. 9* (1976).

[Uhl77] A. UHLMANN: "Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory". In: *Communications in Mathematical Physics* 54 (1977), pp. 21–32.

[VC06] FRANK VERSTRAETE and J IGNACIO CIRAC: "Matrix product states represent ground states faithfully". In: *Physical Review B* 73.9 (2006), p. 094423.

[Vid03] G. VIDAL: "Efficient classical simulation of slightly entangled quantum computations". In: *Physical Review Letters* 91 (2003).

[Von32] J. VON NEUMANN: *Mathematische Grundlagen der Quantenmechanik.* Grundlehren der mathematischen Wissenschaften. Dover, 1932.

[Wil15] M. M. WILDE: "Recoverability in quantum information theory". In: *Proc. R. Soc. A* 471 (2015).

[Wil17] M. M. WILDE: *Quantum Information Theory.* Cambridge University Press, 2017.

[Wol12] M. M. WOLF: *Quantum Channels & Operations.* 2012. URL: https://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MichaelWolf/QChannelLecture.pdf.

[Wol14] M. M. WOLF: *Lecture Notes: Quantum Effects.* 2014. URL: http://www-m5.ma.tum.de/foswiki/pub/M5/Allgemeines/MA5047_2013W/qe_(4).pdf.