Quantum Hidden Markov Chains

Jeroen Wouters

Supervisor:
Prof. Dr. M. Fannes

Jury:
Prof. Dr. D. Bollé, K.U.Leuven, president
Prof. Dr. M. Fannes, K.U.Leuven
Prof. Dr. J.O. Indekeu, K.U.Leuven
Prof. Dr. C. Maes, K.U.Leuven
Prof. Dr. J. Rogiers, K.U.Leuven
Prof. Dr. T.C. Dorlas,
Dublin Institute for Advanced Studies

Dissertation presented in
partial fulfillment of the
requirements for the
degree of Doctor of Science

February 2011

D/2011/10.705/10
ISBN 978-90-8649-392-0

Bedankt, thank you, спасибо,...

Maandenlang loop je met namen en halve zinnen rond in je hoofd. Namen van mensen die je zeker niet mag vergeten te bedanken en zinnen waarmee je je eindeloze dankbaarheid probeert uit te drukken.

Uiteindelijk komt dan het moment waarop je je realiseert dat de thesis naar de drukker moet (of beter gezegd, ze moest al bij de drukker zijn) en dat je het dankwoord nog steeds niet geschreven hebt. Rien ne vas plus! Hier gaan we dan.

In den beginne was er... Mark. Bij het begin van dit doctoraatsproject dan toch. Het eerste mailtje dat ik hem stuurde met de vraag of hij me niet kon helpen bij de zoektocht naar een interessant doctoraatsonderzoek werd prompt beantwoord met een joviaal "Kom maar langs en we bespreken het wel eens."(of toch iets in die strekking). Die openheid en vriendelijkheid zijn constantes gebleven in onze samenwerking. Mark, bedankt voor al je hulp en ondersteuning.

Christ, Désiré, Jos en Joseph, bedankt dat jullie deel wilden uitmaken van mijn doctoraatscommissie. Mijn bijzondere dank ook aan het extern jurylid, prof. Dorlas, voor het grondig nalezen van de thesis en ook voor het trotseren van een sneeuwstorm om op de verdediging aanwezig te zijn.

During the time that I spent at the institute in Leuven, I have had the opportunity to meet some truly wonderful people from around the world. Ismail, thank you for the great collaboration and the nice discussions, going from physics to religion and everything in between. To professor Alicki, thank you for a wonderful stay in Gdańsk. Monika and Stanley, thank you for showing me the true meaning of hospitality. Nick, Fernando and Duck, you're the best office mates one could wish for (although Duck could be a bit more talkative, sure).

Soms durft een fysicus (althans deze toch) al eens te voldoen aan het stereotype beeld en a) ongeorganiseerd zijn en b) een beetje een computernerd zijn. Gelukkig waren er Anita, Anneleen en Christine om me bij te staan bij de praktische beslommeringen van een doctoraat en alzo het leed op het vlak van a) te verzachten. Kwestie b) werd met glans verzorgd door Bavo, Filip en Rob. Zij zorgden voor alle

technische ondersteuning tijdens het werk en de nodige ”geek talk”naast de uren.

To all the friendly and fun colleagues I’ve had over the years (of which there are by now too many to recite them all here): you made life at the institute and even the regular visits to Alma a real joy. Thanks!

Aan de achterban uit Westerlo ben ik mijn dank verschuldigd voor het zorgen voor zeer effectieve en aangename verstrooiing tijdens de dagen buiten het instituut. Ik ben niet helemaal zeker of de uren aan de Westelse toog een positieve invloed hebben gehad op mijn productiviteit, maar toch, vrienden, merci!

Moeke, vake, Hanne, Bert en Saskia, jullie waren er letterlijk al van in de beginne bij (of toch zo goed als). Aan jullie heb ik dan ook zo veel te danken; jullie liefde, steun, aanmoedigingen,... Bedankt.

Чем больше благодарность, тем труднее ее выразить. Отблагодарить тебя за всё, солнышко моё, - почти невозможно. Позволь мне заимствовать слова известного бельгийского математика.

“C’est elle qui, collaboratrice sans le savoir, a constamment écarté de ma route les ronces et les épines, et rendu ma tâche plus facile et plus douce, elle qui a gardé toujours vivante la flamme du foyer où j’ai réchauffé mon coeur.”

Charles-Jean de la Vallée-Poussin

Спасибо тебе, мой цветочек. Без тебя это бы никогда не удалось!

Abstract

English abstract

In this thesis we study quantum mechanical processes with a Markovian character. We focus on matters that are of interest for quantum information theory. There are three main topics: capacity of memory channels, quantum correlations in bipartite systems and asymptotic entropy densities.

Capacity is the amount of information that can reliably be sent through a communication channel. In practical applications, it is an important problem to understand the influence of noise on capacity. The simplest assumption on the type of noise is that at each use of the channel it is independent of previous transmissions. Here, however, we have studied a model channel that does have noise correlations governed by a Markov process. The capacity of our channel turns out to be related to the entropy density of a classical hidden Markov process. We introduce a well-known method for calculating this density and describe an efficient numerical method for determining the capacity. Finally, we analytically study the effect of increasing noise correlations on the channel capacity.

A long term goal is to extend the method used for computing entropy densities of classical hidden Markov process to quantum mechanical systems. To this end we explore what could be understood by "quantum Markov processes". Similarly to classical Markov processes this can be done by limiting the dependence of the process on its history. To do this for quantum systems however, one needs to adapt the notion of conditioning. We describe the concept of conditional state spaces and explore its use to generate quantum correlations.

Finally, the calculation of the asymptotic entropy density that was introduced for classical Markov processes is extended to certain quantum Markov processes. We perform this calculation explicitly for free Fermionic Markov chains. For this purpose we extend a well-known theorem by Szegö concerning the asymptotic behaviour of the eigenvalues of Toeplitz matrices.

Nederlandse abstract

In deze thesis bestuderen we kwantumprocessen met een Markoviaans karakter. We leggen de nadruk op thema's die van belang zijn in kwantuminformatietheorie. Er zijn drie hoofdonderwerpen: de capaciteit van geheugenkanalen, kwantumcorrelaties in tweeledige systemen en asymptotische entropiedichtheden.

Capaciteit is de hoeveelheid informatie die op een betrouwbare manier kan verzonden worden door een communicatiekanaal. In de praktische toepassingen is het van belang te weten hoe de capaciteit zich gedraagt onder invloed van ruis. De eenvoudigste aanname over de aard van de ruis is dat deze bij elk gebruik van het kanaal onafhankelijk is van eerdere verzendingen. Hier hebben we echter een modelkannal bestudeerd dat wel ruiscorrelaties heeft die worden aangestuurd door een Markovproces. De capaciteit van ons kanaal blijkt verbonden te zijn aan de entropiedichtheid van een klassiek Markovproces. We introduceren een bekende methode om deze dichtheid te berekenen en beschrijven een efficiënte numerische methode om deze capaciteit te bepalen. Uiteindelijk bestuderen we op een analytische wijze het effect van toenemende ruiscorrelaties op de capaciteit van het kanaal.

Op de lange termijn is het een doel om de methode die gebruikt werd voor het berekenen van de entropiedichtheid uit te breiden naar kwantummechanische systemen. Om deze reden verkennen we wat men kan verstaan onder "kwantummarkovprocessen". Gelijkaardig aan klassieke Markovprocessen kan dit gedaan worden door het beperken van de afhankelijkheid van het proces van zijn geschiedenis te beperken. Om dit te kunnen doen voor kwantumsystemen, moeten we echter eerst de notie van conditionering aanpassen. We beschrijven het concept van conditionele toestandsruimtes en verkennen het nut hiervan om kwantumcorrelaties te genereren.

Uiteindelijk breiden we de berekening van asymptotische entropiedichtheid die we hebben ingevoerd voor klassieke Markovprocessen uit naar bepaalde kwantummarkovprocessen. We voeren deze berekening expliciet uit voor Fermionische Markovketens. Hiertoe breiden we een bekend theorema van Szegö over het asymptotisch gedrag van de eigenwaarden van Toeplitz matrices uit.

Nomenclature

$< . >$	expectation value
$\mathcal{A}(\mathcal{H})$	CAR algebra on $\mathcal{H}$
$\mathcal{E}$	depolarizing channel
Ω	classical configuration space
ω	quantum state
$\|.\|$	operator norm
$\|.\|_{cb}$	norm of complete boundedness
ρ	density matrix
$\mathsf{H}(X)$	Shannon entropy of X
$\mathsf{H}(X,Y)$	joint Shannon entropy of X and Y
$\mathsf{I}(X,Y)$	mutual information of X and Y
s	entropy density
$\mathsf{S}(\rho)$	von Neumann entropy of ρ
C	channel capacity
p, μ, ν	probability distribution
CAR	canonical anti-commutation relations (abbr.)
CP	completely positive (abbr.)
GICAR	guage-invariant canonical anti-commutation relations (abbr.)

Contents

List of Figures

Chapter 1

Introduction

The invention of electronic means of information storage and transmission has produced an entirely new branch of science during the second half of the twentieth century. This branch of science is called information theory and it brings together scientists from many different backgrounds, including computer science, physics, mathematics and recently even biology.

The foundations of information theory were laid out by Claude Shannon in his 1948 paper "A Mathematical Theory of Communication" [61]. However, as he noted in this paper, this theory did not come into existence out of thin air. The example he uses gives us a clear idea of one of the main goals of information theory: the optimization of information transmission.

Imagine you are in the year 1845. You have just been employed as one of the first operators to send messages on the recently invented telegraph. In order to get the messages across, you have to agree on a system to use, such that the operator on the receiving side will understand the signals you send down the line. As you do not know beforehand exactly what messages you will be sending, you will have to devise a model of this information, based on what you do know about the content. Since you are sitting in a telegraph office in New York City, you can be fairly certain the messages will be written in English. As a first approximation, you can look at the sentence not as a sequence of words, but just as a sequence of individual letters. You will notice that the letters 'e' and 't' occur very often, whereas the letters 'q', 'x', 'y' and 'z' occur much less frequently. An efficient way of encoding the information into the telegraph's alphabet of dots and dashes would then be to assign a short code to the frequent letters and longer codes to the less frequent letters. This idea is indeed reflected in the well-known Morse code (see Table 1). The letters 'e' and 't' are encoded as the two single-letter codes 'dot' and 'dash' respectively, while 'q', 'x', 'y' and 'z' are all assigned four-letter codes.

A	. -	8.2	H	. . . .	6.1	O	- - -	7.5	V	. . . -	1.0
B	- . . .	1.5	I	. .	7.0	P	. - - - .	1.9	W	. - -	2.4
C	- . - .	2.8	J	. - - -	0.2	Q	- - . -	0.1	X	- . . -	0.2
D	- . .	4.3	K	- . -	0.8	R	. - .	6.0	Y	- . - -	2.0
E	.	12.7	L	. - . .	4.0	S	. . .	6.3	Z	- - . .	0.1
F	. . - .	2.2	M	- -	2.4	T	-	9.1			
G	- - .	2.0	N	- .	6.8	U	. . -	2.8			

Table 1.1: The Morse code for letters in the English alphabet and their frequency of occurrence in percentage.

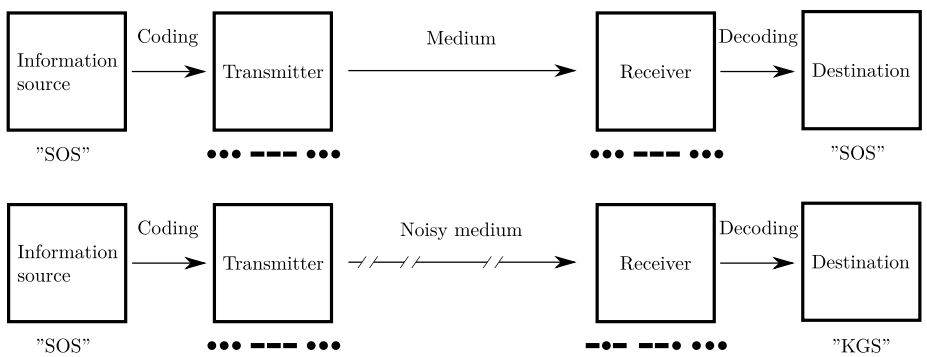

Figure 1.1: Noiseless (top) and noisy (bottom) communication

The Morse code example already illustrates many crucial ideas of information theory, although they were not explicitly described as such at the time. The incoming information is modelled as a probability distribution by assigning a probability of occurrence to the letters in the input alphabet (the English alphabet in this case). This information has to be efficiently encoded into the alphabet of the transmission device (dots and dashes) and finally be decoded by the receiver into the alphabet of the original message so it can be interpreted by the receiver. This situation is depicted in the upper part of Figure 1.1.

The Morse code system served its purpose well for about 100 years. With the advance of technology however, it became apparent that an important factor had to be taken into account. As the transmission of information became ever faster, noise on the communication lines became more and more noticeable. One can think for example of sending the letter 'a', which corresponds to '.-' in Morse code. If by a temporary glitch the long 'dash' signal gets interrupted, it may be interpreted as a short 'dot' signal instead and the receiver would read the letter 'i'. Such a scenario is shown in the lower part of Figure 1.1.

An intuitive solution to combating noise is known to everyone who ever had to get a message across over a bad phone line. One can simply repeat the message a certain number of times and hope that the person on the other side can after a while figure out what the message was. The number of times one has to repeat depends on the intensity of the noise. How many times do we have to repeat the message though? Repeat too few times and errors show up. Repeat too many times and your expensive transatlantic communication line will be bogged down by useless repeating messages. Besides, maybe there are better methods to combat noise, for example by taking into account the noise in the coding system itself? These questions are answered by Shannon's coding theorem and are discussed in more detail in Section 1.1.

During the 20th century, two important evolutions led to a new paradigm in information science. The first one was the realization that there is a strong connection between information and the physical system it is implemented on. An example of this connection is the contained in Landauer's principle, which he used in 1961 to resolve a famous paradox called Maxwell's demon [40, 6]. His principle states that each time a bit of information is erased, a certain amount of heat has to be generated. This shows a close connection between the rather abstract notion of information and a very concrete, physical quantity as heat.

The second important discovery was the advent of quantum mechanics. This was a new theory to explain experiments performed at the beginning of the 20th century that could not be explained by the so-called classical theories that existed before. The fact that information is connected to physical systems, combined with the discovery of radically new behaviour of physical systems led to the establishment of quantum information science.

One of the strange new ways in which quantum systems behave is encountered in the counterintuitive notion of quantum superposition. For physical systems in such a superposition of two outcomes ('on' and 'off', say) one cannot predict which outcome a measurement will give, only assign a probability to each outcome. Shannon's studies were based on the so-called 'classical' way of describing physical systems, where such superpositions are not possible. Either the current of electrons was flowing through the wire, or it was not. The photons of the signal lamp were either being emitted or not. This means quantum systems give us more liberty to encode information in these superpositions.

Then again there are other strange elements of quantum theory that restrict our possibilities, such as the uncertainty relations that tell us we cannot measure all properties of the system with high precision or the fact that measuring a system alters its configuration. As measuring the transmitted particles is crucial to receiving the message, how do these rules of quantum mechanics affect our capability to transmit information? Besides, the effects of quantum mechanics often only come into play when one looks at only a few particles. How does noise

affect the transmission of such fragile messengers? These questions are some of the questions quantum information scientists try to answer. We will review some of the basic principles of quantum mechanics in Section 1.2 and then we will consider the task of communication by means of quantum systems in Section 1.3.

1.1 Classical communication

Now we will have a closer look at the model of communication that Shannon used. The message that is to be transmitted is represented as a string of characters out of a certain alphabet. For our purposes it is not important what exactly the characters in this alphabet are, so we can just as well denote them by a range of numbers, e.g. an alphabet Ω consisting of d characters becomes

$$\Omega = \{1, \ldots, d\} .$$

1.1.1 Entropy

A crucial quantity in information theory is the entropy of the information source. In a general sense entropy describes the information content of the source. It describes the amount of strings that are likely to be generated out of all possible strings. If we want to encode the information the source sends out, less code words will be necessary to describe a source with lower entropy. Hence its information content is lower.

To make this more precise, we will use a derivation in the vein of the one made in [14]. Assume that the emission of characters is governed by a stochastic process X, where character $x \in \Omega$ is emitted with a probability $p_X(x)$. If we assume successive emissions are independent of each other, the probability of a string of k characters $w = (x_1, \ldots, x_k)$ (also called a word) being emitted is given by

$$p_X^{(k)}(w) = p_X(x_1) \ldots p_X(x_k) .$$

Rewriting this probability makes it possible to use the law of large numbers to see what probability will be typical for long strings:

$$p_X^{(k)}(w) = p_X(x_1) \ldots p_X(x_k) = 2^{-k(-\frac{1}{k}\sum_{l=1}^{k} \log p_X(x_l))} .$$

The law of large numbers tells us that an average of a random variable is very likely to be close to the expectation value, for large k. In the exponent, we have an

average, which converges (in probability) to the following quantity:

$$-\frac{1}{k}\sum_{l=1}^{k}\log p_X(x_l) \to -\sum_{x=1}^{d} p_X(x)\log p_X(x) =: \mathsf{H}(X) := \mathsf{H}(p_X)$$

This quantity is called the Shannon entropy of the random variable X.

Let us look at the set of words whose probabilities are close to this typical value of $2^{-k\mathsf{H}(X)}$. The closeness is determined by a positive real number δ and we denote this set as the δ-typical words of length k

$$T^{k,\delta} = \left\{ w \in \Omega^k \,\middle|\, 2^{-k(\mathsf{H}(X)+\delta)} < p_X^{(k)}(w) < 2^{-k(\mathsf{H}(X)-\delta)} \right\} .$$

We can now state the result of the law of large numbers more precisely. It tells us that for k large enough, the probability of being in $T^{k,\delta}$ gets high:

$$\forall \varepsilon > 0,\ \exists k_0 \ \text{ such that } \ \forall k > k_0 :\ P(w \in T^{k,\delta}) > 1 - \varepsilon \,.$$

We can use this outcome to get an estimate on the number of typical words:

$$1-\varepsilon < P(w \in T^{k,\delta}) = \sum_{w \in T^{k,\delta}} p_X^{(k)}(w) < \sum_{w \in T^{k,\delta}} 2^{-k(\mathsf{H}(X)-\delta)} = 2^{-k(\mathsf{H}(X)-\delta)}(\#T^{k,\delta})$$

so the number of words in $T^{k,\delta}$ is less than $(1-\epsilon)2^{k(\mathsf{H}(X)-\delta)}$. By using $P(T^{k,\delta}) < 1$ we also get that there are at most $2^{k(\mathsf{H}(X)+\delta)}$.

Now consider a code that only encodes the typical words and forgets about the non-typical ones. If we use codewords of $k(\mathsf{H}(X)+\delta)$ bits we are certain that we have enough room to encode all the $2^{k(\mathsf{H}(X)+\delta)}$ typical words. This code also has a low error, as the non-typical words that are left out only amount to a probability of ϵ.

If we use a code C of length less than $k(\mathsf{H}(X)-\delta)$ however, the error probability gets close to one for large k. The error rate is $1 - P(C)$, where

$$\begin{aligned} P(C) &= P(C \cap T^{k,\delta/2}) + P(C \cap \overline{T^{k,\delta/2}}) \\ &\leq (\#C)2^{-k(\mathsf{H}(X)-\delta/2)} + P(\overline{T^{k,\delta/2}}) \\ &\leq 2^{-k\delta/2} + \epsilon \,. \end{aligned}$$

To avoid errors we should not use a coding system which size is smaller then $2^{k\mathsf{H}(X)}$ and we know there are codes that do work for all sizes larger than this number. This means the Shannon entropy $\mathsf{H}(X)$ is a measure of the size of an efficient

description of the data and hence of the information content of the random variable X.

It is easily seen that H takes values in $[0, \log d]$. The minimum value of 0 corresponds to the pure Dirac measures where there is no information content, the source always emits the same character. The maximum value is attained when every character has an equal probability of being emitted, i.e. the source is very unpredictable. The entropy function is also a concave function on the state space:

$$\mathsf{H}(\lambda_1 p_{X_1} + \lambda_2 p_{X_2}) \geq \lambda_1 \mathsf{H}(p_{X_1}) + \lambda_2 \mathsf{H}(p_{X_2}), \;\; \lambda_i \geq 0, \;\; \lambda_1 + \lambda_2 = 1.$$

We see that when we introduce more randomness by mixing two measures, the randomness and hence the information content increases.

1.1.2 Entropies of multiple variables

If we want to study more than one random variable, we can use entropy to study the information content of the random variables together and also the influence knowledge of one has on another. The concepts of joint entropy, conditional entropy and mutual information are important quantities in this respect.

The joint entropy is simply the entropy of the joint distribution of the random variables. Let's say we have two random variable X and Y taking values in Ω_X and Ω_Y, with joint distribution $p_{X,Y}$, then we denote the entropy of this distribution by $\mathsf{H}(X, Y)$:

$$\mathsf{H}(X, Y) = -\sum_{\substack{x \in \Omega_x \\ y \in \Omega_y}} p_{X,Y}(x, y) \log p_{X,Y}(x, y) \,.$$

This entropy is the joint entropy of X and Y and expresses the information content of the two random variables together. For example, it is easy to show that if X and Y are independent (i.e. $p_{X,Y}(x, y) = p_X(x) p_Y(y)$ the joint entropy $\mathsf{H}(X, Y)$ is the sum of the two entropies $\mathsf{H}(X)$ and $\mathsf{H}(Y)$. If Y is a deterministic function of X, Y does not add any randomness to X and thus the joint entropy of X and Y together is equal to the entropy of X alone.

Restricting a distribution $p_{X,Y}$ on a composite system $\Omega_{X,Y} = \Omega_X \times \Omega_Y$ to the subsystem Ω_X returns the first marginal of $p_{X,Y}$

$$p_X(x) = \sum_y p_{X,Y}(x, y).$$

The Shannon entropy behaves well with respect to restrictions

- monotonicity: $\mathsf{H}(X) \leq \mathsf{H}(X, Y)$,

- sub-additivity: $\mathsf{H}(X,Y) \leq \mathsf{H}(X) + \mathsf{H}(Y)$, and
- strong sub-additivity: $\mathsf{H}(X,Y,Z) + \mathsf{H}(Y) \leq \mathsf{H}(X,Y) + \mathsf{H}(Y,Z)$.

We can also determine how knowledge of one random variable influences the other. If we know the random variable Y to take some value y and if the variable X is not independent of Y, the probability distribution of X will be influenced by this knowledge. The probability distribution in this case is given by the conditional probabilities

$$p_X(x|Y=y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}\,.$$

The corresponding entropy we denote by $\mathsf{H}(X|Y=y)$:

$$\mathsf{H}(X|Y=y) = -\sum_x \frac{p_{X,Y}(x,y)}{p_Y(y)} \log \frac{p_{X,Y}(x,y)}{p_Y(y)}\,.$$

If the value of Y is known, but we do not specify a specific value, we have to take the average of the previous entropy:

$$\mathsf{H}(X|Y) = \sum_y p_Y(y)\mathsf{H}(X|Y=y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(y)} = \mathsf{H}(X,Y) - \mathsf{H}(Y)$$

This is the conditional entropy of X conditioned on Y.

We see that to get the information content of X separately we subtract $\mathsf{H}(Y)$ from the joint information content of X and Y. In a similar manner one can look at the amount of information that the two random variables share:

$$\mathsf{I}(X,Y) = \mathsf{H}(X) + \mathsf{H}(Y) - \mathsf{H}(X,Y)\,.$$

This quantity is called the mutual information of X and Y and plays an important role in the theory of communication.

In a very loose way of speaking, one can think of the information of different random variables as collections of data that have a certain overlap as well as information unique to that random variable. The information content of different subsets is given by the entropies described in this section. Figure 1.2 depicts the relation between these different quantities.

Let us look at some examples to illustrate the quantities introduced in this section.

Example 1. *As random variables X and Y we take the input and output of a noisy communication channel with an alphabet of two symbols. At the input of the channel the character* 0 *is generated with a probability μ and* 1 *is generated with probability $1-\mu$. During the transmission the characters are interchanged with*

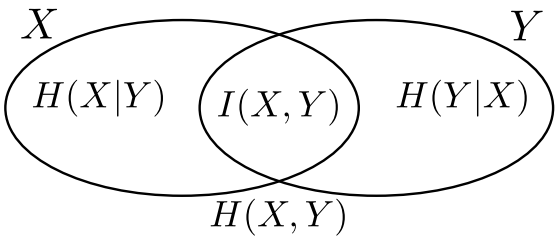

Figure 1.2: Diagram showing the relations between different entropies of multiple random variables.

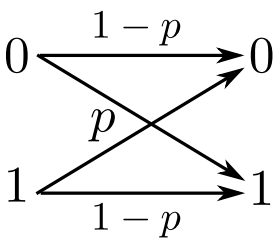

Figure 1.3: Schematic representation of the binary symmetric channel.

probability p, i.e. 0 becomes 1 or 1 becomes 0. The process is represented in Figure 1.3.

The probabilities associated with the input process X are μ and $1-\mu$, so the entropy of this random variable is

$$\mathsf{H}(X) = -\mu \log \mu - (1-\mu)\log(1-\mu) =: \mathrm{h}(\mu)\,.$$

This entropy function h *is called the binary entropy and is plotted in Figure 1.4. When $\mu = 0$ or $\mu = 1$ there is no randomness at all and the entropy is 0. When $\mu = 1/2$ the string of ones and zeros is completely random with no bias towards one or the other symbol. This is when the entropy reaches its maximum.*

The joint probabilities of the variables X and Y are easily seen to be:

$$p(0,0) = \mu(1-p) \qquad p(0,1) = \mu p$$

$$p(1,0) = (1-\mu)p \qquad p(1,1) = (1-\mu)(1-p)\,.$$

Using these probabilities we find that the joint entropy $\mathsf{H}(X,Y)$ becomes a sum of two terms:

$$\mathsf{H}(X,Y) = \mathrm{h}(p) + \mathrm{h}(\mu)\,.$$

This is easily explained by looking at Y as the result of two independent processes, namely the input process X and an error process Z which determines whether the

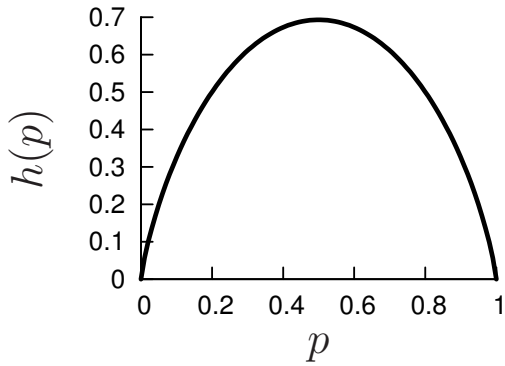

Figure 1.4: The entropy of a binary random variable with parameter p.

input is altered or not. As in this special example the probability of error is the same whether we input 0 *or* 1*, X and Z are independent random variable and their joint entropies is simply the sum of their entropies. The entropy of Y conditioned on X then becomes the entropy of the error process:*

$$\mathsf{H}(Y|X) = \mathsf{H}(X,Y) - \mathsf{H}(X) = \mathsf{H}(Z)\,.$$

This shows that the information that is unique to Y is the randomness added to it by the noisy channel.

Example 2. *Instead of limiting ourselves to two random variables, we can extend to more variables. Going further we can even think of infinite chains of variables that may be correlated. For example we can think of the subsequent characters emitted by a communication channel. If the input characters are correlated, so will the output (unless the noise completely destroys the correlations). But even when the inputs are statistically uncorrelated the output can be correlated by correlations in the noise that is applied. We will study this case in more detail in the next chapter.*

The most basic example of a correlated process is the stationary Markov process. The outcomes at each time step lie in a configuration space Ω. The probability of n subsequent outcomes is then given by a probability measure p_n on $\Omega^{\times n} = \Omega \times \ldots \times \Omega$. The Markov process has the characteristic of being correlated to the past only through the most recent outcome:

$$p(x_n|x_1,\ldots,x_{n-1}) = p(x_n|x_{n-1})\,.$$

This ensures that the probability of a sequence $(x_1,\ldots,x_n)$ can be decomposed into transition probabilities:

$$p_n(x_1,\ldots,x_n) = p_1(x_1)p(x_2|x_1)\ldots p(x_n|x_{n-1})\,.$$

In the case of an infinite Markov chain the entropy diverges to infinity:

$$\lim_{n\to\infty} \mathsf{H}_n = \infty\,,$$

where H_n *is the entropy of* p_n*. Hence we calculate the entropy density* $\overline{h}$ *instead:*

$$\overline{h} := \lim_{n\to\infty} \frac{\mathsf{H}_n}{n}\,.$$

Using the properties of the Markov process we can find its entropy density:

$$\overline{h} = \lim_{n\to\infty} \frac{1}{n}\left(\sum_x h(p(x)) + (n-1)\sum_{x,y} p(x)h(p(y|x))\right) \tag{1.1}$$

$$= \sum_{x,y} p(x)h(p(y|x))\,, \tag{1.2}$$

so we see that the entropy density of the Markov process is an average of the entropies of the conditional probabilities.

As a side note we remark that the concepts of entropy and mutual information have spread far beyond the subject of communication theory. As an example we mention the use of mutual information in image processing, more precisely image registration. Image registration is the task of deforming one image such that is matches up as well as possible with a given target image.

Of course one needs to define what "as well as possible" means. For this purpose the mutual information between the intensity distributions of the images is often used. This quantity has a major advantage over simple distances between intensities as it can also align images made using different devices or under different conditions. In such images, intensities may differ while still being correlated. Take for example medical images made with X-ray computed tomography (CT) and magnetic resonance imaging (MRI). Bone has a high intensity on CT, but a low intensity on MRI. Even if the images are perfectly aligned, simply subtracting the high and low intensities will result in a large distance. However, when the images are aligned there is a high correlation between the high and low intensities of bone pixels, resulting in high mutual information.

Figure 1.5 shows a patient image that is deformed to match up with a standard target image. If in the target image an expert has delineated different organs or regions, this information can now be automatically pulled back to the new image, without requiring the expert to do this task again for the new image. For more details see [69, 44, 43].

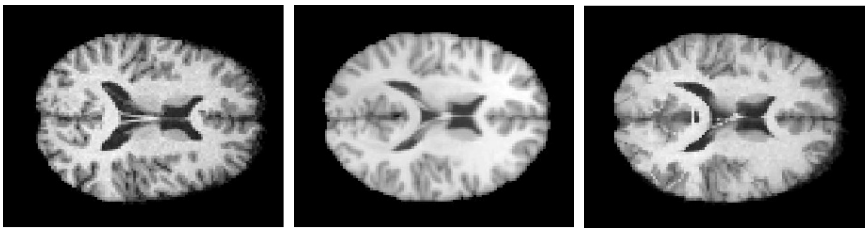

Figure 1.5: The registration of 3D medical images. Left: slice of the original image. Middle: slice of the deformed image. Right: slice of the target image

1.1.3 Shannon's classical coding theorem

In the previous section we have seen an example of noisy communication. A character was sent through a noisy channel, where it had a certain probability of being altered by the noise during transmission. In this section we will generalize this example and describe the results of Shannon concerning the information capacity of such noisy channels.

Instead of the alphabet consisting of only two letters in the example, we can now have any number of characters at the input or output. As the channel we are sending information through is noisy, the characters on the input are not mapped one-to-one to the output. There is a certain probability that an input character is mapped to an output character. If x is on the input, we denote by $p(y|x)$ the probability of receiving y. Furthermore it is assumed that the noise is not influenced by previous uses of the channel, i.e.:

$$p(y_1, \ldots, y_n|x_1, \ldots, x_n) = p(y_1|x_1) \ldots p(y_n|x_n)\,.$$

Such channels are called memoryless channels.

Of course there is no reason for us to just send our message and hope it comes through unharmed. As noted before, we can repeat the message or use more advanced tricks. This is called coding the message. On the receiving end we need to decode the message accordingly. The coding is done by choosing a set of N codewords $V = \{v^{(1)}, \ldots, v^{(N)}\}$ of length n. When these codewords are sent through the channel, an output word of n characters in the output alphabet is received. The receiver now has to guess which of the N codewords the sender wanted to send to him. This is done by partitioning the set of all possible output words into N sets $\{W^{(1)}, \ldots, W^{(N)}\}$. If the output falls into the set $W^{(j)}$, it is decided that the input should have been j. The choice of $v^{(j)}$ and $W^{(j)}$ is called code of length n encoding N words and is denoted by $\mathcal{C}_{n,N}$.

The probability of decoding correctly is then given by $p(W^{(j)}|v^{(j)})$. The maximal error probability is thus given by

$$P_e(\mathcal{C}_{n,N}) = \max_{1\leq j\leq N}(1 - p(W^{(j)}|v^{(j)}))\,.$$

In order to communicate successfully, this maximal error probability should be kept low. If there exists a code of length n such that we can send N characters with low probability of error, we can successfully communicate $\log(N)$ bits per n uses of the channel (as k bits require 2^k codewords). The rate of communication for this code is then given by $R = \log(N)/n$, the number of bits sent per use of the channel.

Now we can define the capacity of the channel. It is simply the highest rate that can be successfully attained over the noisy channel. More precisely it is the maximal rate for which the probability of error vanishes when we allow for arbitrarily long codes.

Definition 1. *A rate R is called achievable if there exists a series of codes $\mathcal{C}_{n,\lceil 2^{nR}\rceil}$ for $n \in \mathbb{N}$ such that $\lim_{n\to\infty} P_e(\mathcal{C}_{n,\lceil 2^{nR}\rceil}) = 0$. The capacity C of a noisy memoryless channel is the maximal achievable rate.*

The result Shannon obtained is a way to calculate this capacity for a given channel. The Shannon coding theorem states that the capacity of a noisy memoryless channel is given by

$$C = \max_{p_x} \mathsf{I}(X, Y)\,.$$

The maximum is taken over all the input probability distributions. For a proof of this theorem, see for example the original proof by Shannon [61] or some more recent books [13, 14].

Example 3. *Let us look again at the binary symmetric channel. We use the Shannon coding theorem to calculate its capacity:*

$$\begin{aligned}
\mathsf{I}(X, Y) &= \mathsf{H}(Y) - \mathsf{H}(Y|X) \\
&= \mathsf{H}(Y) - \sum_x p(x)\mathsf{H}(Y|Y = x) \\
&= \mathsf{H}(Y) - \sum_x p(x)h(p) \\
&= \mathsf{H}(Y) - h(p) \\
&\leq 1 - h(p)
\end{aligned}$$

Equality is reached in the inequality when the input distribution is uniform, since then the output is also uniform and has maximal entropy $\log(2) = 1$*. We see that the capacity is maximal when* $p = 0$ *or when* $p = 1$*. In these cases there is no noise. When the noise is maximal (*$p = 1/2$*) we get a uniform distribution at the output, irrespective of the input. We can get no information through and the capacity is zero. For other values of* p *the capacity behaves as one would expect, i.e. the more noise, the lower the capacity. For more examples, see [13].*

1.2 Quantum mechanics

The previous section was concerned only with the communication of classical information. When we are dealing with quantum systems, the way we describe systems and measurements of the system needs to be changed.

Previously we described the input and output of a communication channel by probability distributions over a set of possible configurations. Let's denote this configuration space by X and the probability distribution by p_x. Then imagine we want to measure a certain property f of the system. For every configuration x there is a corresponding outcome of the measurement $f(x)$. Hence, the expected value of the measurement would be

$$\langle f \rangle_{p_x} = \sum_{x \in X} p(x) f(x) \,.$$

The distribution p_x is what is called the state of the system, while the function f is called an observable.

In the 1920's physicist realized that this description in terms of probability distributions did not fit some of the new experimental results of the time. Quantum mechanics was proposed as a more fitting description. In this section we will briefly describe the principles of quantum mechanics. For a detailed overview of the quantum mechanics and its origins, see for example [26, 65, 49].

1.2.1 States and observables

The classical observables, which are functions over the configuration space, are replaced in quantum mechanics by an algebra of operators, say $\mathcal{A}$. A finite dimensional algebra of observables can always be decomposed into a direct sum of matrix algebras, so we will focus for now on the case where $\mathcal{A}$ is the algebra of bounded operators on a finite-dimensional Hilbert space, i.e.

$$\mathcal{A} = \mathcal{B}(\mathcal{H}) \,.$$

A state is a mathematical object describing the properties of a system under study. It is a function which takes in an observable and returns its expectation value under this state of the system. In the classical case, a state was given by a probability distribution. In the quantum setting a state ω is a linear functional on the algebra of observables:

$$\omega : \mathcal{A} \to \mathbb{C}\,,$$

such that $\omega(A) \geq 0$ for $A \geq 0$ and $\omega(\mathbb{1}) = 1$.

The set of quantum states is a compact convex set, therefore there are distinct states on its boundary that cannot be written as a mix of other states. These are called pure states. Taking a convex combination of states physically corresponds to statistically mixing different possible configurations. Hence the pure states correspond to physical systems that are in a certain well-described configuration. A pure quantum state on a matrix algebra corresponds to a normalized vector $|\phi\rangle$ in a vector space $\mathcal{H}$ such that

$$\langle\phi|\phi\rangle = 1\,.$$

Example 4. *In a two-dimensional system, the pure states are linear combinations of two basis vectors:*

$$|\phi\rangle = \alpha_0|0\rangle + \alpha_1|1\rangle\,.$$

The requirement that $|\phi\rangle$ is normalized boils down to

$$|\alpha_0|^2 + |\alpha_1|^2 = 1\,.$$

Up to an unimportant global phase factor these states are determined by two angles θ and φ:

$$|\psi\rangle = \begin{bmatrix} \cos\frac{\theta}{2}e^{-\imath\varphi/2} \\ \sin\frac{\theta}{2}e^{\imath\varphi/2} \end{bmatrix} \qquad 0 \leq \theta \leq \pi\,, \quad 0 \leq \varphi < 2\pi \tag{1.3}$$

A physical observable in the case of a finite-dimensional quantum system is given by a self-adjoint linear operator on the Hilbert space (in the finite dimensional case this is just a Hermitian matrix). The possible outcomes of the measurement correspond to the eigenvalues of the operator (hence the self-adjointness requirement).

Suppose we want to measure an operator A with eigenvalues a_i and with projectors P_i on the corresponding eigenspaces. The probability p_i of measuring outcome a_i under a pure state $|\psi\rangle$ is then given by

$$p_i = \langle\psi|P_i\psi\rangle\,.$$

The expectation value of the measurement of A is thus

$$\langle A\rangle_{|\psi\rangle} = \sum_i a_i p_i = \sum_i a_i \langle\psi|P_i\psi\rangle = \langle\psi|A\psi\rangle$$

Example 5. *The three spatial components of a spin* $\frac{1}{2}$ *particle are given by the following observables:*

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -\imath \\ \imath & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The state $|0\rangle$ *for example has its spin pointing in the z direction:*

$$\langle \sigma_x \rangle = 0\,, \langle \sigma_y \rangle = 0\,, \langle \sigma_z \rangle = 1$$

The meaning of the word 'pure' in 'pure quantum state' describes the case where we know for certain that the quantum system is prepared in a certain state. Often we cannot know this for certain and we can only say the system is in a statistical mixture of pure states. Such a state is called a mixed quantum state. If for example we have n possible states $|\psi_i\rangle$ which can appear with probability p_i, the expectation value of an observable A then becomes

$$\langle A \rangle = \sum_{i=1}^{n} p_i \langle \psi_i | A \psi_i \rangle = \sum_{i=1}^{n} p_i \operatorname{Tr}(|\psi_i\rangle\langle\psi_i| A) = \operatorname{Tr} \rho A\,.$$

The matrix ρ is called the density matrix of the system and it is given by

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

This matrix is positive:

$$\langle \psi | \rho \psi \rangle \geq 0 \quad \forall |\psi\rangle$$

and it is normalized, i.e. its trace equals 1:

$$\operatorname{Tr} \rho = \sum_i \lambda_i = 1\,,$$

where the λ_i are the eigenvalues of ρ_i. Conversely any positive and normalized linear operator on the vector space describes a quantum state.

Example 6. *A mixed two-dimensional state is described by a* 2×2 *density matrix. It is useful to write this matrix using the Pauli matrices* $\{\sigma_0 = \mathbb{1}, \sigma_x, \sigma_y, \sigma_z\}$ *as basis. As* σ_0 *is the only Pauli matrix with trace different from zero, the following characterization gives us all normalized Hermitian* 2×2 *matrices:*

$$\rho = \frac{1}{2}(\sigma_0 + r_x \sigma_x + r_y \sigma_y + r_z \sigma_z) \qquad r_x, r_y, r_z \in \mathbb{R}\,.$$

As we already have normalization it suffices to have $\det \rho \geq 0$ *for positivity. This can be easily calculated:*

$$\det \rho = 1 - r_x^2 - r_y^2 - r_z^2 \geq 0\,.$$

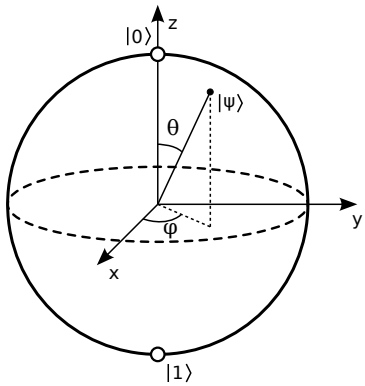

Figure 1.6: The Bloch sphere

We see that the mixed states correspond to a point in the 3-dimensional unit ball. This representation of the set of mixed 2×2 *states is called the Bloch ball, as shown in Figure 1.6. The pure states are the extreme points of the unit ball and lie on a sphere, as we have seen before. The pure state in Equation 1.3 corresponds to the parameters*

$$(r_x, r_y, r_z) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)\,,$$

which clarifies the choice of parameters in Equation 1.3.

The classical description can be recovered from the quantum case by assuming all matrices commute. More precisely, the algebra of observables $\mathcal{A}_c$ is the algebra of finite-dimensional real diagonal matrices:

$$\mathcal{A}_c = \left\{ \sum_{i=1}^{d} a_i |i\rangle\langle i| \,\middle|\, a_i \in \mathbb{R} \right\}.$$

Taking the trace with a density matrix ρ, we get

$$\langle A \rangle_\rho = \operatorname{Tr} A\rho = \sum_{i=1}^{d} a_i \rho_{ii}\,,$$

so the states on this observable algebra are essentially diagonal matrices. Since the state is positive and normalized, the diagonal elements ρ_{ii} form a probability distribution. The diagonal elements a_i correspond to a classical observable and by the above equation, the expectation value reduces to the classical expectation value of a_i under the distribution ρ_{ii}.

1.2.2 Bipartite states

When we are studying two quantum systems that interact with each other, we need a way of describing the two systems as one. In quantum mechanics this composition is done through a tensor product. If we have one system with a Hilbert space $\mathcal{H}_1$ and another with a Hilbert space $\mathcal{H}_2$, the tensor product of the two is denoted by $\mathcal{H}_1 \otimes \mathcal{H}_2$. The tensor product space $\mathcal{H}_1 \otimes \mathcal{H}_2$ is spanned by elements of the form $|\varphi\rangle \otimes |\psi\rangle$ with $|\varphi\rangle \in \mathcal{H}_1$, $|\psi\rangle \in \mathcal{H}_2$, which satisfy the following equations:

$$(|\varphi_1\rangle + |\varphi_2\rangle) \otimes |\psi\rangle = |\varphi_1\rangle \otimes |\psi\rangle + |\varphi_2\rangle \otimes |\psi\rangle \tag{1.4}$$

$$|\varphi\rangle \otimes (|\psi_1\rangle + |\psi_2\rangle) = |\varphi\rangle \otimes |\psi_1\rangle + |\varphi\rangle \otimes |\psi_2\rangle \tag{1.5}$$

$$\alpha|\varphi\rangle \otimes |\psi\rangle = |\varphi\rangle \otimes \alpha|\psi\rangle = \alpha(|\varphi\rangle \otimes |\psi\rangle) \tag{1.6}$$

If $|\varphi_i\rangle$ and $|\psi_j\rangle$ are vectors in $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, the inner product between their tensor products is defined as follows:

$$\langle \varphi_1 \otimes \psi_1 | \varphi_2 \otimes \psi_2 \rangle = \langle \varphi_1 | \varphi_2 \rangle \langle \psi_1 | \psi_2 \rangle$$

A bipartite pure state is then given by a normalized vector in the tensor product space:

$$|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2 \text{ such that } \langle \psi | \psi \rangle = 1\,.$$

Correspondingly, a bipartite mixed state is given by a positive normalized linear operator on the tensor product space:

$$\rho \in \mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2) \text{ such that } \rho \geq 0 \text{ and } \operatorname{Tr} \rho = 1$$

For a composite system, restricting to a sub-system amounts to taking partial traces over remaining parties

$$\rho_1 = \operatorname{Tr}_2 \rho_{12}.$$

The partial trace operation Tr_2 is defined as follows by its operation on product operators:

$$\operatorname{Tr}_2(A \otimes B) = (\operatorname{Tr} B)A$$

It can then extended by linearity to $\mathcal{B}(\mathcal{H}_1 \otimes \mathcal{H}_2)$

1.2.3 Entropy

The Shannon entropy as described for probability distributions needs to be adapted to density matrices. It will be called the von Neumann entropy:

$$\mathsf{S}(\rho) = -\operatorname{Tr} \rho \log \rho\,.$$

It is equal to the Shannon entropy of the eigenvalue distribution of ρ.

The von Neumann entropy retains some of the properties of the Shannon entropy. For example it is also concave:

$$\mathsf{S}(\lambda\rho_1 + (1-\lambda)\rho_2) \geq \lambda\mathsf{S}(\rho_1) + (1-\lambda)\mathsf{S}(\rho_2)$$

Also subaddivity and strong subaddivity[42] remain true:

$$\mathsf{S}(\rho_{12}) \leq \mathsf{S}(\rho_1) + \mathsf{S}(\rho_2) \tag{1.7}$$

$$\mathsf{S}(\rho_{123}) + \mathsf{S}(\rho_2) \leq \mathsf{S}(\rho_{12}) + \mathsf{S}(\rho_{23}) \tag{1.8}$$

However, the most basic and perhaps most intuitive property of the Shannon entropy, monotonicity, does not carry over. Consider for example the two-qubit state $|\Phi^+\rangle\langle\Phi^+|$ with $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$. Its entropy $\mathsf{S}(|\Phi^+\rangle\langle\Phi^+|)$ is zero as it is a pure state, while its restriction ρ_1 is the maximally mixed state, which has maximal entropy, so clearly $\mathsf{S}(\rho_1) \not\leq \mathsf{S}(\rho_{12})$.

1.2.4 Entanglement

The structure of quantum mechanical systems consisting of two parts has some important consequences. Let's look for example at a pure state $\mu_{1,2}$ on a classical bipartite system $\Omega_1 \times \Omega_2$. Classical pure states are Kronecker deltas:

$$\mu_{1,2}(i,j) = \delta_{(c,d)}(i,j) = \delta_c(i)\delta_d(j)\,.$$

It is clear that every classical pure state can be written as a product of pure states on the components.

For quantum systems this is no longer the case. Let $|\psi_{1,2}\rangle$ be a pure state vector in $\mathcal{H}_1 \otimes \mathcal{H}_2$. We say that the state is separable if it can be written as a product of pure states on the two components:

$$|\psi_{1,2}\rangle = |\psi_1 \otimes \psi_2\rangle\,,$$

where $\psi_1 \in \mathcal{H}_1$ and $\psi_2 \in \mathcal{H}_2$. States that are not separable are called entangled.

Some quantum states can however not be written as such a product state. Let's assume that our bipartite pure state $|\psi\rangle$ can in fact be written as a product state $|\psi_1 \otimes \psi_2\rangle$. This would imply that the corresponding bipartite density matrix is given by

$$\rho = |\psi\rangle\langle\psi| = |\psi_1\rangle\langle\psi_1| \otimes |\psi_2\rangle\langle\psi_2|\,.$$

The density matrix of the first part alone is given by the partial trace over the second part of the bipartite density matrix:

$$\rho_1 = \mathrm{Tr}_2\,\rho = |\psi_1\rangle\langle\psi_1|$$

If the bipartite pure state is separable, the reduced density matrix ρ_1 is also pure.

Example 7. *Let's look at the following bipartite pure state in a system consisting of two 2-dimensional subsystems:*

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle)$$

The density matrix of the first subsystem now becomes

$$\rho_1 = \frac{1}{2}\mathbb{1}\,.$$

This density matrix is not pure, so ψ is not separable. This state is an entangled state.

The mixedness of the reduced density matrix ρ_1 is thus a method of detecting entanglement. As we have seen, von Neumann entropy measures the mixedness of quantum states. Indeed, the entropy of the reduced density matrix is a so-called entanglement measure for pure states [52]. In Chapter 6 we will study this quantity for a specific class of states. For mixed states it is as yet unclear yet how to measure entanglement. This is one of the reasons we study a geometric way to look at quantum correlations in Chapter 5.

The definition of separable pure states can be extended to bipartite mixed states. A density matrix $\rho_{1,2}$ on $\mathcal{H}_1 \otimes \mathcal{H}_2$ is called separable if there exists a classical decomposition

$$\rho_{1,2} = \sum_{i=1}^{n} \lambda \rho_{i,1} \otimes \rho_{i,2}\,,$$

where $\rho_{i,j}$ is a density matrix on $\mathcal{H}_j$ and the λ_i form a probability distribution. Note that this definition operationally means that the state can be prepared by generating numbers according to the classical distribution λ_i and then preparing state $\rho_{i,1}$ in system 1 and $\rho_{i,2}$ in system 2. This means no quantum information has to be exchanged between the two parts to construct the state. If a mixed state cannot be constructed in this way, it is called entangled.

1.2.5 Quantum dynamics

Until now we have only looked at static quantum systems. To discuss communication we need more than this. First of all quantum states evolve in time. Second of all, the system under consideration often interacts with an environment that is outside of the observer's control. This is how noise is introduced in a quantum communication channel. An isolated system is often called closed, while a system that is in contact with an environment is called open.

The evolution of a closed system in a pure quantum state is described by the Schrödinger equation:

$$\imath\hbar\frac{\partial}{\partial t}|\phi\rangle = H|\phi\rangle\,,$$

where H is the Hamiltonian describing the energy of the system. Integrating this equation we get that after time t the state has evolved to

$$|\phi_t\rangle = e^{-\imath tH/\hbar}|\phi_0\rangle\,.$$

As we are not interested in the entire time evolution, but only the result at the output of the transmission channel, we can fix the time at t and write

$$U(t) = e^{-\imath tH/\hbar}\,,$$

which is a unitary operator.

From the evolution of closed systems in a pure state the evolution of a mixed state can be derived. The evolution of a density matrix of a closed system is given by

$$\rho_t = U(t)\rho_0 U(t)^*\,. \tag{1.9}$$

It is however impossible to shield our communication device completely from the environment. There will always be noise in realistic circumstances. Noise arises by the system of interest interacting in an uncontrolled way with the environment. This is the setting of open quantum dynamics.

Taking the system and environment together, we again have a closed system. The communication channel is now modelled as follows:

$$\Phi(\rho) = \mathrm{Tr}_E\, U(\rho\otimes\rho_E)U^*\,.$$

Here ρ is the density matrix of the system of interest, E denotes the environment the system interacts with and ρ_E is its initial state. U is now the evolution under the total system-plus-environment Hamiltonian. Linear maps of density operators of this form are called completely positive maps.

A way to determine whether or not a linear map is completely positive is given by the isomorphism due to Choi[11] and Jamiołkowski[32]. This isomorphism maps completely positive maps into positive matrices. The map is completely positive if and only if it is positive after applying the Choi-Jamiołkowski isomorphism. For a given map Φ, the isomorphism is given by applying Φ on a maximally entangled state:

$$I_{CJ}(\Phi) = \sum_i |i\rangle\langle i|\otimes\Phi(|i\rangle\langle i|)$$

Example 8. *The qubit depolarizing channel $\mathcal{E}$ is a map that 'flips' the 2×2 density matrix ρ with probability $1-x$:*

$$\mathcal{E}(\rho) = x\rho + (1-x)((\operatorname{Tr}\rho)\mathbb{1} - \rho)\,.$$

The Choi-Jamiołkowski isomorphism gives us:

$$I_{CJ}(\mathcal{E}) = \begin{bmatrix} x & 0 & 0 & 2x-1 \\ 0 & 1-x & 0 & 0 \\ 0 & 0 & 1-x & 0 \\ 2x-1 & 0 & 0 & x \end{bmatrix}.$$

This matrix has eigenvalues $\{3x-1, 1-x, 1-x, 1-x\}$, so $\mathcal{E}$ is completely positive if and only if $\frac{1}{3} \leq x \leq 1$. Note that the corresponding classical map does not have such a restriction.

1.3 Quantum communication

We can now wonder how the quantum communication channels described in the previous section can be used to transmit information. The first question we have to answer is what kind of information we want to send. We can think for example of a full quantum state ρ or of classical information in the form of a probability distribution. We will focus in this section on the rate at which classical information can be sent through a quantum channel. This rate is described by the quantum counterpart of the classical coding theorem we have encountered before. For a review of other possible capacities, see [63]. The layout of this section largely follows that in [30].

With the classical information we want to send encoded using an input alphabet $A = \{1, \ldots, a\}$, we choose for every element $i \in A$ an encoding quantum state ρ_i on a Hilbert space $\mathfrak{H}$. This input state is then transmitted using a quantum channel $\Lambda : \mathcal{B}(\mathfrak{H}) \to \mathcal{B}(\mathfrak{K})$. For the channel to be a valid quantum channel it must be a completely positive trace preserving map.

Transmitting the element $i \in A$ results in a quantum state $R_i = \Lambda(\rho_i)$ being received on the output side. On this side, the received quantum state is measured using a resolution of identity in $\mathfrak{K}$. This resolution of identity is a set of positive operators $X = \{X_i\}$ on $\mathfrak{K}$ such that $\sum_i X_i = \mathbb{1}$.

The conditional probability of the receiver measuring j, when the input i was sent, is given by $p(j|i) = \operatorname{Tr} R_i X_j$. If at the input side the element i is sent with a probability π_i, the amount of information that will be received is quantified by the classical Shannon information,

$$\mathsf{I}_{\Lambda,1}(\pi, \rho, X) = \sum_{i,j\in A} \pi_i p(j|i) \log\left(\frac{p(j|i)}{\sum_{k\in A}\pi_k p(j|k)}\right)\,. \tag{1.10}$$

If the sender is allowed to use the channel n times, the channel use can be described by the product channel $\Lambda_n = \otimes^n \Lambda$ on $\otimes^n \mathfrak{H} = \mathfrak{H} \otimes \ldots \otimes \mathfrak{H}$. The input alphabet is now A^n and the probability distribution of a word $u = (i_1, \ldots, i_n) \in A^n$ being sent is again denoted by π_u. The codeword corresponding to the input u is given by

$$\rho_u = \rho_{i_1} \otimes \ldots \otimes \rho_{i_n}$$

and results in $R_u = R_{i_1} \otimes \ldots \otimes R_{i_n}$ being received. The conditional probability and the Shannon information $I_{\Lambda,n}$ for the n-product of the channel can now be introduced completely analogously to Eq. (1.10), with the summations over A^n instead of A.

The maximum amount of information that can be sent with n channel uses is now given by

$$C_n(\Lambda) = \sup_{\pi,\rho,X} \mathsf{I}_{\Lambda,n}(\pi, \rho, X) \; .$$

Due to the fact that $C_n + C_m \leq C_{m+n}$, the limit

$$C_{\text{class}}(\Lambda) = \lim_{n\to\infty} \frac{C_n(\Lambda)}{n}$$

exists. Using Shannon's coding theorem, we see that C_{class} is the least upper bound of the rate of information that can be transmitted with asymptotically vanishing error.

The HSW theorem [30, 59] gives an expression for this classical product state capacity of noisy memoryless quantum channels,

$$C_{\text{class}}(\Lambda) = \chi^* = \sup_{\pi,\rho} \chi(\Lambda),$$

where χ is the Holevo χ quantity

$$\chi(\{(\pi_i, \Lambda(\rho_i))\})$$
$$= \mathsf{S}(\sum_i \pi_i \Lambda(\rho_i)) - \sum_i \pi_i \mathsf{S}(\Lambda(\rho_i)) \; .$$

Due to the convexity of the von Neumann entropy, the supremum can in fact be taken over pure states ρ_i.

Example 9. *Although the depolarizing channel is about the simplest quantum channel one can think of, its classical capacity has been determined only quite recently. See [36] for details.*

Outlook

The overview and examples in this chapter have introduced many of the aspects of information theory that will be studied in more detail in this thesis.

In Chapter 2 we will demonstrate a new method to efficiently calculate the classical capacity of a quantum depolarizing channel with memory. This is a channel similar to the one introduced in Example 8, with the difference that the probability of error is not constant across different uses of the channel. The amount of applied noise is described by a classical Markov process. As we will show, the calculation of this capacity boils down to the calculation of the entropy density of a so-called hidden Markov chain, much like in Example 2. This calculation relies on a method to calculate entropy densities first used by Blackwell[8]. The difference with Example 2 is that in the case of a hidden Markov process such a direct calculation is no longer tractable. Instead of the asymptotic entropy density $\lim_{n\to\infty} \mathsf{S}_n/n$, the asymptotic entropy rate $\lim_{n\to\infty} \mathsf{S}_{n+1} - \mathsf{S}_n$ can be calculated here. Using the strong subadditivity property (Eq. 1.8), one can demonstrate that these quantities are equal for stationary processes.

The method used to calculate the entropy density relies on the asymptotic equality of entropy density and rate. The properties that appear in the proof of this equality are the strong sub-additivity and stationarity. Strong sub-additivity is also valid for quantum processes, so as long as the process is stationary, the same proof can be used for both the classical and quantum cases. The obvious question then arises whether it is possible to do a calculation similar to the Blackwell calculation for quantum processes. The Blackwell calculation makes use of the hidden Markov structure of the process. To apply this method successfully in the quantum case, we need a quantum counterpart with similar properties. Such states are the finitely correlated states and are introduced in Chapter 3.

The entropy density calculation is worked out in this thesis for a specific class of states, namely the free Fermionic states. These states are fermionic versions of Gaussian states. These states form a nice class of quantum states that are true quantum states, but where computations are still tractable. These states and some of their properties are introduced in Chapter 4.

The construction of the finitely correlated states in Chapter 3 will show us the importance of so called conditional state spaces. These conditional state spaces show the influence of one subsystem on another. Otherwise stated, they show the correlations between two subsystems. It will be no surprise then that the 'finitely correlated' in 'finitely correlated states' actually means having a finite-dimensional conditional state space. Two different methods to characterize the possible conditional state spaces of free Fermions are described in Chapter 5.

Finally, we calculate the entropy density of the free Fermionic finitely correlated

states in Chapter 6. As we have demonstrated, the entropy density of a subsystem is an important measure of entanglement for pure states. As is demonstrated in Chapter 2 however, it also arises in the calculation of the capacity of memory channels. Furthermore the entropy density measures the mixedness of mixed states and therefore is an important quantity in statistical mechanics. For the case of free Fermionic finitely correlated states the density matrices correspond to Toeplitz matrices. By calculating the entropy density and rate we obtain a new Szegö theorem for Toeplitz matrices.

Chapter 2

Classical hidden Markov chains

2.1 Introduction

In Section 1.2.5, we have seen how the transfer of information using quantum systems can be modelled by a quantum channel. The capacity of such channels can be calculated using the Holevo-Schumacher-Westmoreland (HSW) theorem.

To derive this theorem the assumption has been made that the channel we are using returns to the state it was in before we had used it. In other words, the noise that affects our message during transmission is the same every time we use the channel. Such a channel is called a memoryless channel, as it has no memory of what has happened to it in the past.

In reality this will seldom be the case. The channel will retain some memory of what state it was in previously. If the channel returns to its initial state on a time scale that is much shorter than the interval between our uses of the channel, the memoryless approximation is justified. This is however an idealized situation. In real-world applications, we need to consider memory effects.

Two examples of quantum systems where memory effects play a role are spin chains [9] and micromasers [20]. In spin chains, particles that behave as miniature magnets are positioned on a one dimensional chain. The sender interacts with one end of the chain and manipulates it in such a way that the receiver on the other end of the chain can decode the message. After the message has been propagated from one end to the other, the particles will be in a different configuration than the one they started out in. By applying a strong magnetic field to particles, they are then reset to the initial configuration. In practice they will however never be perfectly aligned with the magnetic field and retain a memory of their previous

configuration.

In a micromaser a beam of atoms with a transition in the microwave regime is sent through an optical resonator. This resonator has a resonance mode at the same frequency as the transition in the atoms. The atoms are excited to one of the higher atomic level involved in the transition prior to entering the resonator. They emit photons inside the resonator and in this way a microscopic maser or micromaser is created. Since the cavity is a rather delicate instrument, no measurement can be performed on it directly. Instead, the particles coming out of the cavity are measured. If the lifetime of the photons inside the cavity is longer than the time it takes for a new atom to enter the cavity, the atoms will be correlated when they leave the cavity.

In this chapter we study a channel that has similar correlations between successive uses. We do not study memory channels in all their full glory though. We still make an assumption on the correlations in the noise. We look at channels that only remember the last particle that has passed through it. Such processes are called Markovian, as we have encountered in Example 2. The type of noise that is applied at each time step is also of a specific form. We employ a very simple model of the channel noise here, namely the depolarizing channel from Example 8. With a certain probability, the input is perfectly copied to the output. In the other cases, the output is a completely random signal, independent of the input. So there are two possibilities: either we have perfect transfer of information or the input information is completely erased. We use two different erasure probabilities or noise levels here. At each channel use, one of the two noise levels is chosen. The channel has a memory of which level of noise was applied the last time, i.e. the switching between different erasure probabilities is a classical Markov process.

As a new result, we derive an efficient method for numerically computing the classical capacity of the quantum depolarizing channel with memory. This work has been published in [70, 1].

Intuitively, we can expect that correlations in the noise make the noise more predictable. The sender could use this to his advantage as he has more information about what errors to expect and thus can better counter them. This has been demonstrated in some classical and quantum memory channels [15, 48]. This is one of the properties we will look at using our numerical method.

This chapter is structured as follows. In Section 2.2, we look at how the calculation of the capacity of memoryless quantum channels can be extended to certain channels with memory. The class of channels that will be considered is the class of forgetful channels. We introduce some recent results that allow us to rewrite the classical capacity of this quantum memory channel in terms of the capacity of a classical channel. In Section 2.3, we construct our model memory channel with Markovian noise correlations. In Section 2.4, algebraic measures are introduced, which are used in Section 2.5 to further simplify the calculation of the channel capacity.

Finally, in Section 2.6, we show how this reformulation allows us to easily calculate the capacity of the channel numerically.

2.2 Classical capacity of quantum memory channels

In Section 1.3, we have looked at the capacity of a memoryless quantum channel for communicating classical information. As this setting is not appropriate in many real-world applications, it has to be relaxed. In [39], the channel capacity is determined for *forgetful* channels. These channels do have memory, but the effects of the initializing memory essentially disappear after a number of time steps.

Recall that a memoryless channel in the Schrödinger picture is a completely positive trace preserving map that maps the state space $S(\mathcal{A})$ on an algebra $\mathcal{A}$ corresponding to the sender to $S(\mathcal{B})$, the state space on $\mathcal{B}$, at the receiving end. Successive applications are uncorrelated and are given by a tensor product of the channel:

$$\Gamma_n = \underbrace{\Gamma \otimes \Gamma \otimes \ldots \otimes \Gamma}_{n}$$

A general memory channel is then constructed as follows. When the channel has a memory, we add a third system that represents the state of the memory. The operator algebra of the memory is denoted by $S(\mathcal{M})$. The memory channel then transforms states from $S(\mathcal{A})$ to $S(\mathcal{B})$, while also transforming the memory. We have

$$\Lambda : S(\mathcal{M} \otimes \mathcal{A}) \to S(\mathcal{B} \otimes \mathcal{M})\,. \tag{2.1}$$

Successive applications of the memory channel result in the following channel Λ_n from $S(\mathcal{M} \otimes \mathcal{A}^{\otimes n})$ to $S(\mathcal{B}^{\otimes n} \otimes \mathcal{M})$:

$$\Lambda_n = (\mathrm{id}_{\mathcal{B}}^{n-1} \otimes \Lambda) \circ (\mathrm{id}_{\mathcal{B}}^{n-2} \otimes \Lambda \otimes \mathrm{id}_{\mathcal{A}}) \circ \ldots \circ (\Lambda \otimes \mathrm{id}_{\mathcal{A}}^{n-1})$$

This construction is represented in Figure 2.2. At each time step the channel takes in a part of the input (seen at the bottom of the figure), propagates it to the output (top of the figure) and transforms the memory (on the horizontal axis of the figure). Obviously the channel uses are correlated in this case.

The forgetful channels are a special class of memory channels. These are channels where for large n, the initial state of the memory has no effect. More concretely, for large n we can model the behaviour of the memory by a set of quantum channels that do not depend on the initial state of the memory.

The behaviour of the memory can be described by disregarding the output of the channel:

$$\hat{\Lambda}_n : S(\mathcal{M} \otimes \mathcal{A}^{\otimes n}) \to S(\mathcal{M}) : \rho \mapsto \mathrm{Tr}_{\mathcal{B}^{\otimes n}}(\Lambda_n(\rho))$$

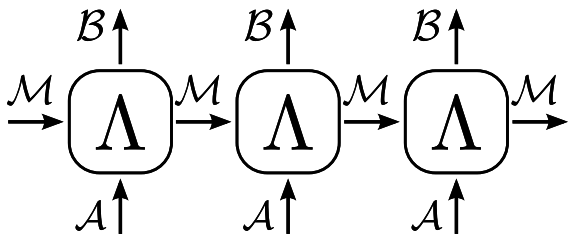

Figure 2.1: Memory channel construction

A forgetful channel is a channel where this behaviour of the memory is close to a model of the memory that does not depend on the initial memory configuration.

In order to talk about channels being close to one another, we need a distance measure. A norm that is convenient when talking about channels is the norm of complete boundedness, defined for a general linear operator $T : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$:

$$\|T\|_{cb} = \sup_n \|T \otimes \mathrm{id}_n\|$$

where $\|.\|$ is the operator norm. The norm of complete boundedness is used because of its nice properties, namely multiplicativity ($\|T_1 \otimes T_2\|_{cb} = \|T_1\|_{cb}\|T_1\|_{cb}$) and unitality for any channel T: $\|T\|_{cb} = 1$. It can be shown to be equivalent to several other norms used for channels [38].

Now we can define what it means for a channel to be forgetful.

Definition 2. *A memory channel Λ_n is forgetful iff there exists a set of quantum channels $\tilde{\Lambda}_n : S(\mathcal{A}^{\otimes n}) \to S(\mathcal{M})$ such that*

$$\lim_{n\to\infty} \|\hat{\Lambda}_n - \tilde{\Lambda}_n \circ \mathrm{Tr}_{\mathcal{M}}\|_{cb} = 0$$

For these forgetful channels the classical product state capacity has been shown [39] to be given by

$$C^* = \lim_{n\to\infty} \frac{C_{\mathrm{class}}(\Lambda_n)}{n}\ , \tag{2.2}$$

Example 10. *The simplest memory channel is the identity channel, where the input and output algebras $\mathcal{A}$ and $\mathcal{B}$ are equal:*

$$I : S(\mathcal{M} \otimes \mathcal{A}) \to S(\mathcal{A} \otimes \mathcal{M}) : \rho_M \otimes \rho \mapsto \rho \otimes \rho_M\ .$$

This channel is not forgetful, the memory stays in it initial state, so we cannot write the evolution of the memory as a map depending only on the inputs.

Note that although the map in itself it not the usual identity map, due to the construction of memory channel we have used, the map interchanges algebras by default (see Eq. 2.1). Therefore, this channel has no effect on the memory state and the output is equal to the input.

Example 11. *A second simple example is the switch channel, where the input, output and memory algebras $\mathcal{A}, \mathcal{B}$ and $\mathcal{M}$ are equal:*

$$\Lambda^s : S(\mathcal{A} \otimes \mathcal{A}) \to S(\mathcal{A} \otimes \mathcal{A}) : \rho_M \otimes \rho \mapsto \rho_M \otimes \rho \,.$$

Although the map in itself is the usual identity map, it interchanges memory and input states at each use. Hence, after one use the memory state will be replaced by the first input state and at the second use the evolution of the memory will no longer depend on the initial memory state. Therefore this is a forgetful channel.

2.3 The depolarizing memory channel

We will now look at one specific class of forgetful channels, namely depolarizing channels with noise correlations determined by a classical Markov process. If the Markov process determining the memory is irreducible and aperiodic, the state of the memory will converge to the stationary distribution, independent of the initial state and the channel is a forgetful one. We can then apply the tools from the previous section. Other cases can be recovered by using the results from [17].

2.3.1 Construction of the channel

The forgetful channel is constructed by combining two memoryless single qubit depolarizing channels ($\mathcal{E}_0$ and $\mathcal{E}_1$), each with its own error statistics, switching between them using a two-state Markov chain ($Q = (q_{ij}), \quad i, j \in \{0, 1\}$). Q is the 2×2 Markov transition matrix with q_{ij} being the probability of switching from channel i to channel j. Hence, $q_{ij} \geq 0$ and $q_{i0} + q_{i1} = 1$ for $i, j \in \{0, 1\}$. The channel is forgetful when the Markov chain is *aperiodic* and *irreducible*.

The depolarizing channels, which we encountered before in Example 8, can be written as

$$\mathcal{E}_i(\rho) = x_i^0 \rho + x_i^1 (\mathbb{1} - \rho) \,,$$

where $x_i^0 + x_i^1 = 1$. These single qubit channels can be thought of as probabilistically mixing the identity channel $\mathrm{id}(\rho) = \rho$ (with probability x_i^0) and the 'bit flip' $F(\rho) = \mathbb{1} - \rho$ (with probability $x_i^1 = 1 - x_i^0$) acting on a single qubit density operator ρ. This channel is only completely positive for $1/3 \leq x_i^0 \leq 1$. This can be verified by writing out the Choi matrix of the channel and checking its positivity.

The channel memory is represented by a classical two-level state space. The transformation of the memory is determined by the Markov transition matrix Q and at each time step the channel $\mathcal{E}_i$ is applied according to the probability distribution of the memory:

$$\Lambda(\boldsymbol{\nu} \otimes \rho) = \sum_i \mathcal{E}_i \otimes (Q\boldsymbol{\nu})_i \delta_i$$

The classical memory of this channel converges exponentially fast to the stationary state of the Markov process. As we are only interested in asymptotic quantities as the channel capacity, this evolution towards stationarity will have no influence, as it is finite in time. The initial memory configuration will hence be of no importance. Therefore, we can set ν equal to the stationary state γ from the beginning. Furthermore, we are only interested in the information input-output properties of the channel. When we trace out the memory output, the channel Λ_n built up from Λ, corresponding to n successive uses is

$$\Lambda_n = \rho_1 \otimes \ldots \otimes \rho_n \mapsto$$

$$\sum_{i_1,\ldots,i_n} \gamma_{i_1} q_{i_1 i_2} \cdots q_{i_{n-1} i_n} \mathcal{E}_{i_1}(\rho_1) \otimes \ldots \otimes \mathcal{E}_{i_n}(\rho_n) \; .$$

The sum is over all possible paths $(i_1, \ldots, i_n) \in \{0,1\}^n$ and each term is a tensor product of the selected sub-channels weighted by the probability of occurrence (γ_i is the initial probability of selection set to the stationary distribution of the Markov process: $Q^T\gamma = \gamma$).

2.3.2 Classical capacity

We calculate the capacity with this n-use form of the channel and regularize by taking the limit $n \to \infty$ as in Eq. (2.2). We will see that the capacity is equivalent to the capacity of the classical Gilbert-Elliott channel [48]. However, due to the quantum nature of the channel, we had to impose complete positivity and this restricts the choice in switching probability to the interval $[\frac{1}{3}, 1]$, in contrast to the classical case, where any probability is allowed.

We restrict ourselves to the product state capacity, meaning we do not allow entanglement across channel inputs at different times. We choose the input to the channel as

$$\rho_i = \rho_{\vec{l}}^{(n)} = \rho_{(l_1,\ldots,l_n)}^{(n)} := |l_1\rangle\langle l_1| \otimes \ldots \otimes |l_n\rangle\langle l_n| \; ,$$

where the l_i are arbitrary pure qubit states.

Applying the channel Λ_n, we get

$$\Lambda_n(\rho_{\bar{l}}^{(n)}) = \sum_{i_1,\ldots,i_n} \gamma_{i_1} q_{i_1 i_2} \ldots q_{i_{n-1} i_n}$$

$$(x_{i_1}^0 |l_1 \oplus 0\rangle\langle l_1 \oplus 0| + x_{i_1}^1 |l_1 \oplus 1\rangle\langle l_1 \oplus 1|) \otimes \ldots$$

$$\otimes (x_{i_n}^0 |l_n \oplus 0\rangle\langle l_n \oplus 0| + x_{i_n}^1 |l_n \oplus 1\rangle\langle l_n \oplus 1|) \ ,$$

where $(l_i \oplus 1)$ denotes the qubit state with a flipped Bloch vector with respect to $l_i = (l_i \oplus 0)$

$$|l_i \oplus 1\rangle\langle l_i \oplus 1| = \mathbb{1} - |l_i \oplus 0\rangle\langle l_i \oplus 0|$$

By expanding the product above we see that the eigenvalues of the output state are given by

$$\lambda_n(\bar{k}) = \sum_{i_1,\ldots,i_n} \gamma_{i_1} q_{i_1 i_2} \ldots q_{i_{n-1} i_n} x_{i_1}^{k_1} \ldots x_{i_n}^{k_n} \ . \tag{2.3}$$

Note that these eigenvalues are independent of the choice of the input state.

The channel output can now be written as

$$\Lambda_n \left(\rho_{\bar{l}}^{(n)}\right) = \sum_{\bar{k}} \lambda_n(\bar{k}) \rho_{\bar{l} \oplus \bar{k}}^{(n)} \ .$$

Hence, if we calculate the first term in the Holevo χ quantity taking π to be the uniform distribution ($\pi_{\bar{l}} = 1/2^n$) over the set of input states $\rho_{\bar{l}}^{(n)}$, we see that

$$\rho_{\text{out}} := \sum_{\bar{l}} \frac{1}{2^n} \Lambda_n \left(\rho_{\bar{l}}^{(n)}\right)$$

$$= \frac{1}{2^n} \sum_{\bar{k}} \lambda_n(\bar{k}) \sum_{\bar{l}} \rho_{\bar{l} \oplus \bar{k}}^{(n)} \ .$$

Since $\bar{l}$ goes over all possible combinations, so does $\bar{l} \oplus \bar{k}$, so we can relabel them

$$\Phi_{\text{out}} = \frac{1}{2^n} \sum_{\bar{k}} \lambda_n(\bar{k}) \sum_{\bar{l}'} \rho_{\bar{l}'}^{(n)} \ .$$

Since the eigenvalues in Eq. (2.3) sum to one, we see that ρ_{out} is the maximally mixed state

$$\rho_{\text{out}} = \frac{1}{2^n} \sum_{\bar{l}'} \rho_{\bar{l}'}^{(n)} \tag{2.4}$$

$$= \frac{1}{2^n} (\sum_{l_1} |l_1\rangle\langle l_1|) \otimes \ldots \otimes (\sum_{l_n} |l_n\rangle\langle l_n|) \tag{2.5}$$

$$= \frac{1}{2^n} \mathbb{1} \otimes \ldots \otimes \mathbb{1}; . \tag{2.6}$$

Thus, $\mathsf{S}(\rho_{\text{out}})$ is maximal and is equal to $\log_2(2^n) = n$.

The second term in the Holevo χ quantity is

$$-\sum_{\bar{l}} \pi_{\bar{l}} \mathsf{S}\left(\Lambda_n(\rho_{\bar{l}})\right) .$$

Since the eigenvalues $\lambda_n(\bar{k})$ of $\Lambda_n(\rho_{\bar{l}})$ do not depend on the choice of $\rho_{\bar{l}}$, this term does not influence the maximization. Hence our choice of π and ρ maximizes the Holevo χ quantity.

Thus, the final expression for the regularized capacity Eq. (2.2) is

$$C^* = \lim_{n\to\infty} \frac{1}{n} C_{\text{class}}(\Lambda_n)) = 1 - \lim_{n\to\infty} \frac{1}{n} \mathsf{S}(\Lambda_n(\rho)) . \tag{2.7}$$

If we were to calculate the output entropy using the eigenvalues in Eq. (2.3), the calculation would be exponentially long in n. Therefore, other techniques are needed. The way we approach the problem is by using a formula due to Blackwell [8] to calculate the entropy of a hidden Markov process. The eigenvalues of the output state correspond to the probabilities of such a process.

A hidden Markov process can be defined as follows. If we have a translation-invariant measure ν with the Markov property on $\Omega_H^{\mathbb{Z}}$, where Ω_H is a finite set, then a hidden Markov measure can be constructed on $\Omega^{\mathbb{Z}}$ through a function $\Phi : \Omega_H \to \Omega$, with the following local densities

$$\mu((\omega_m, \ldots, \omega_n)) = \sum_{\substack{\epsilon_m,\ldots,\epsilon_n \in \Omega_H \\ \Phi(\epsilon_m)=\omega_m \ldots \Phi(\epsilon_n)=\omega_n}} \nu((\epsilon_m, \ldots, \epsilon_n)) , \tag{2.8}$$

where $\omega_m, \ldots, \omega_n \in \Omega$. For obvious reasons, these processes are also called functions of Markov processes.

2.4 Algebraic measures

An algebraic measure μ is a translation-invariant measure on a set $\{0,\ldots,q-1\}^{\mathbb{Z}}$, with probabilities determined by matrices E_a with positive entries, one for each of the q states. The probability of a sequence is obtained by applying a positive linear functional σ to a matrix product of the corresponding matrices of the states of the sequence: $\mu(i_1,\ldots,i_n)=\sigma(E_{i_1}\ldots E_{i_n})$. This matrix algebraic construction is the reason for the name *algebraic measure*, studied in detail in Ref. [21]. As we shall see, the hidden Markov processes correspond to a set of algebraic measures with a specific positivity structure and the converse holds as well.

2.4.1 Manifestly positive measures

In [21] it was shown that hidden Markov processes correspond to manifestly positive algebraic measures. The local densities of such a manifestly positive algebraic measure on an infinite chain $\Omega^{\mathbb{Z}}$ of classical state spaces $\Omega=\{0,\ldots,d-1\}$ are of the form

$$\mu((\omega_1,\ldots,\omega_n))=\langle\tau|E_{\omega_1}\ldots E_{\omega_n}\sigma\rangle\ ,$$

where $\omega_i\in\Omega$, τ and σ are vectors in $\mathbb{R}^d$ with non-negative elements (denoted $(\mathbb{R}^d)^+$) and the E_i are $d\times d$ real matrices with non-negative elements (denoted M_d^+). In order to have a well-defined and compatible measure, we require that $E|\sigma\rangle=|\sigma\rangle$, $E^*|\tau\rangle=|\tau\rangle$ where $E=\sum_\omega E_\omega$ and that $\langle\tau|\sigma\rangle=1$. By compatible, we mean that by tracing out one site, we recover the measure on the remaining sites:

$$\sum_{\omega_n\in\Omega}\mu((\omega_1,\ldots,\omega_n))=\mu((\omega_1,\ldots,\omega_{n-1}))$$

Example 12. *A regular Markov chain $\mu((\omega_m,\ldots,\omega_n))$ on $\{0,\ldots,q-1\}^{\mathbb{Z}}$ can be written as an algebraic measure. If we choose τ, σ and the E_i as*

$$\sigma\in(\mathbb{R}^d)^+:\ \ \sigma_a=1\ \textit{for}\ a\in\Omega\ ,$$

$$\tau\in(\mathbb{R}^d)^+:\ \ \tau_a=\mu((a))\ \textit{for}\ a\in\Omega\ ,$$

$$E_a\in M_d^+:\ \ (E_a)_{b,c}=\delta_{a,b}\frac{\mu((b,c))}{\mu((b))}\ \textit{for}\ a,b,c\in\Omega\ ,$$

one can check that $\langle\tau|E_{\omega_m}\ldots E_{\omega_n}\sigma\rangle$ indeed gives the correct densities.

From this example it is easy to see that if we have a hidden Markov process on $\Omega^{\mathbb{Z}}$ defined by a map $\Phi:\Omega_H\to\Omega$ and a Markov measure μ on Ω_H with corresponding matrices E_a, the manifestly positive algebraic measure corresponding to the hidden

Markov measure is given by the same vectors σ and τ as before and the following matrices:

$$F_a \in M_d^+ : F_a = \sum_{\epsilon \in \Omega_H, \Phi(\epsilon)=a} E_\epsilon \quad \text{for } a \in \Omega \,. \tag{2.9}$$

For a proof of the converse, namely that every manifestly positive algebraic measure corresponds to a hidden Markov measure, we refer to [21].

2.4.2 Entropy density

In this section we present a method for calculating entropy densities of hidden Markov processes due to Blackwell [21, 8].

An important step in this method is to use strong subadditivity of the entropy to show that the entropy density of the process is equal to its asymptotic entropy production. Strong subadditivity for a process means that

$$\mathsf{S}_{\Lambda_1 \cap \Lambda_2}(\mu) + \mathsf{S}_{\Lambda_1 \cup \Lambda_2}(\mu) \leq \mathsf{S}_{\Lambda_1}(\mu) + \mathsf{S}_{\Lambda_2}(\mu)$$

where S_Λ is the entropy of the process restricted to the set of indices Λ:

$$\mathsf{S}_\Lambda(\mu) = -\sum_{\omega_\Lambda \in \Omega^\Lambda} \mu(\omega_\Lambda) \log \mu(\omega_\Lambda) \,.$$

We are looking at a translationally invariant measure, so we can always assume the first index of a set to be 0 and denote the entropy of n consecutive points by S_n.

Using the strong subadditivity of the entropy, one can show for translation-invariant measures that [2, 35, 66]

$$\mathsf{s} := \lim_{n\to\infty} \frac{\mathsf{S}_n}{n} = \lim_{n\to\infty} (\mathsf{S}_{n+1} - \mathsf{S}_n) \,,$$

which means that the entropy density equals the asymptotic entropy production.

Theorem 1. *Given a shift-invariant state ω on quantum spin chain $\mathcal{A}^{\mathbb{Z}}$, denote by S_n the von Neumann entropy of its restrictions to n consecutive spins. We then have*

$$\lim_{n\to\infty} \frac{\mathsf{S}_n}{n} = \lim_{n\to\infty} \mathsf{S}_{n+1} - \mathsf{S}_n \tag{2.10}$$

Proof. Let $\mathsf{S}_{[k,l]}$ denote the von Neumann entropy of the restriction of the spin chain to the interval $[k, l]$. By translation invariance we have

$$\mathsf{S}_n = \mathsf{S}_{[0,n-1]} = \mathsf{S}_{[x,n+x]} \,.$$

Together with strong sub-additivity on the partition $\mathcal{A}^{\otimes n+2} = \mathcal{A} \otimes \mathcal{A}^{\otimes n} \otimes \mathcal{A}$ this gives us that the sequence of entropy increments is decreasing:

$$\mathsf{S}_{n+2} - \mathsf{S}_{n+1} \leq \mathsf{S}_{n+1} - \mathsf{S}_n$$

By the triangle inequality for the von Neumann entropy, we also have that

$$\mathsf{S}_{n+1} \geq \mathsf{S}_n - \mathsf{S}_1 \,.$$

The sequence is decreasing and bounded from below, so the limit exists.

If we denote this limit by s, we now have that for a small $\epsilon > 0$ there exists an $n_0 \in \mathbb{N}$ such that $|\mathsf{S}_{n+1} - \mathsf{S}_n - \mathsf{s}| \leq \epsilon$ whenever $n \geq n_0$. We then have

$$\begin{aligned}
&|\mathsf{S}_n - \mathsf{S}_{n_0} - (n - n_0)\mathsf{s}| \\
&= |(\mathsf{S}_n - \mathsf{S}_{n-1} - \mathsf{s}) + (\mathsf{S}_{n-1} - \mathsf{S}_{n-2} - \mathsf{s}) + \ldots + (\mathsf{S}_{n_0+1} - S_{n_0} - \mathsf{s})| \\
&\leq |\mathsf{S}_n - \mathsf{S}_{n-1} - \mathsf{s}| + \ldots + |\mathsf{S}_{n_0+1} - \mathsf{S}_{n_0} - \mathsf{s}| \leq (n - n_0)\epsilon
\end{aligned}$$

By dividing by $n - n_0$, we get that the limit of S_n/n exists and is equal to s. □

Example 13. *First, let us use this relation to calculate the entropy density of a regular Markov process on $\Omega^{\mathbb{Z}}$ with transition matrix Q and stationary measure γ:*

$$\begin{aligned}
h &= \lim_{n\to\infty} (\mathsf{H}_{n+1} - \mathsf{H}_n) \\
&= \lim_{n\to\infty} \sum_{i_1,\ldots,i_{n+1}\in\Omega} [-\mu(i_1,\ldots,i_{n+1}) \log \mu(i_1,\ldots,i_{n+1}) \\
&\qquad\qquad +\mu(i_1,\ldots,i_{n+1}) \log \mu(i_1,\ldots,i_n)] \\
&= \lim_{n\to\infty} \sum_{i_1,\ldots,i_{n+1}\in\Omega} \mu(i_1,\ldots,i_n)(-\mu(i_{n+1}|i_1\ldots i_n) \log \mu(i_{n+1}|i_1\ldots i_n)) \\
&= \lim_{n\to\infty} \sum_{i_n,i_{n+1}\in\Omega} \mu(i_n)(-\mu(i_{n+1}|i_n) \log \mu(i_{n+1}|i_n)) \\
&= \sum_{a,b\in\Omega} \gamma(a)(-Q_{a,b} \log Q_{a,b}) \,.
\end{aligned}$$

By rewriting the entropy production in terms of conditional probabilities, the entropy density calculation splits into two parts. One is a probability distribution at time step n, which converges to the invariant measure γ. The other part is an entropy associated to the transition from one time step to the next.

A similar calculation as in the example above can be done for hidden Markov processes, although the more complicated conditional probabilities make it more involved in this case.

As we noted in the previous section, the probability distribution of a hidden Markov process can be expressed as an algebraic measure:

$$\mu(i_1, \ldots, i_n) = \langle \tau | F_{i_1} \ldots F_{i_n} \sigma \rangle$$

with the F_i, τ and σ manifestly positive matrices and vectors. The entropy production calculation now becomes:

$$s = \lim_{n\to\infty} \sum_{i_1,\ldots,i_{n+1}\in\Omega} \mu(i_1, \ldots, i_n)(-\mu(i_{n+1}|i_1 \ldots i_n) \log \mu(i_{n+1}|i_1 \ldots i_n)) \,.$$

Since

$$\mu(i_{n+1}|i_1 \ldots i_n) = \frac{\langle F_{i_1}^* \ldots F_{i_n}^* \tau | F_{i_{n+1}} \sigma \rangle}{\mu((i_1, \ldots, i_n))} \,,$$

this limit can be written as

$$s = \lim_{n\to\infty} \int_{\mathcal{B}_\sigma} \phi_n(d\nu) h_F(\nu) \,,$$

where

$$\mathcal{B}_\sigma = \{\nu \in (\mathbb{R}^d)^+ | \, \langle \nu | \sigma \rangle = 1\}$$

$$h_F(\nu) = -\sum_{i_{n+1}\in\Omega} \langle \nu | F_{i_{n+1}} \sigma \rangle \log \langle \nu | F_{i_{n+1}} \sigma \rangle$$

$$\phi_n(d\nu) = \sum_{i_1,\ldots,i_n\in\Omega} \mu((i_1, \ldots, i_n)) \delta_{\frac{F_{i_n}^* \ldots F_{i_1}^* \tau}{\mu((i_1,\ldots,i_n))}} (d\nu) \,.$$

The calculation again turns into a dynamical system of measures ϕ_n and an entropy function that is related to the transition from one point to the next on the underlying Markov process.

The transformation T_μ from one measure ϕ_n to the next is given by:

$$(T_\mu f)(\nu) = \sum_{i\in\Omega} \langle \nu | F_i \sigma \rangle f\left(\frac{F_i^* \nu}{\langle \nu | F_i \sigma \rangle}\right)$$

This transformation is expressed on functions $\mathcal{B}_\sigma$, but can also be seen as a transformation of measures on $\mathcal{B}_\sigma$ through $T_\mu(\phi)(f) = \phi(T_\mu(f))$. We then have $\phi_{n+1} = T_\mu(\phi_n)$.

This transformation can be shown [21] to be a contraction, so a fixed point argument can be used to show that ϕ_n converges to a unique measure ϕ that is invariant under T_μ

$$\phi(T_\mu f) = \phi(f) \ .$$

This measure allows us to calculate the entropy density

$$s = \int_{\mathcal{B}} \phi(d\nu) h_F(\nu) \ . \tag{2.11}$$

2.5 Algebraic measure of the channel

Our goal in the remaining part of the chapter is to translate the switching depolarizing channel into the setting of algebraic measures. We characterize the invariant measure that allows us to calculate the entropy density using Blackwells entropy density formula.

2.5.1 The manifestly positive matrices F_i

The relationship between a hidden Markov measure, say μ' on $\Omega^{\mathbb{Z}}$, and the underlying Markov measure ν with the Markov property on $\Omega_H^{\mathbb{Z}}$ is through a 'tracing' function $\Phi : \Omega_H \to \Omega$, as is shown in Eq. (2.8).

The underlying Markov process for the overall quantum channel has a four state configuration space corresponding to channel selection and error occurrence: $\Omega_H = \{(0,0),(0,1),(1,0),(1,1)\}$. The first index indicates which depolarizing channel has been chosen and the second indicates whether a bit flip occurred. The elements of the transition matrix, E, for this process are then given by

$$(E)_{\{(i,j)(i',j')\}} = q_{ii'} x_{i'}^{j'} \ , \tag{2.12}$$

so we have a transition matrix

$$E = \begin{pmatrix} q_{00}x_0^0 & q_{00}x_0^1 & q_{01}x_1^0 & q_{01}x_1^1 \\ q_{00}x_0^0 & q_{00}x_0^1 & q_{01}x_1^0 & q_{01}x_1^1 \\ q_{10}x_0^0 & q_{10}x_0^1 & q_{11}x_1^0 & q_{11}x_1^1 \\ q_{10}x_0^0 & q_{10}x_0^1 & q_{11}x_1^0 & q_{11}x_1^1 \end{pmatrix} \ .$$

The probability of going from (i,j) to (i',j') is given by the switching probability $q_{ii'}$ from channel i to i', multiplied by the probability $x_{i'}^{j'}$ that channel i' produces the error-occurrence j'.

The function that produces the correct hidden Markov process is then given by

$$\Phi((i,j)) = j \ .$$

This function reflects the fact that we are unaware of the choice of channel that has been made. The only effect that is visible from the outside is whether or not an input qubit has been flipped. Thus, Φ has to 'trace out' the choice of channel. Φ maps into the two-state error configuration space containing 'no flip' and 'flip': $\Omega = \{0,1\}$.

Using the fact that the matrices $E_{(i,j)}$ defining the algebraic measure of a Markov process ((Sec. 2.4.1, Pg. 33), $a = (i,j) \in \Omega_H$) have only one non-zero row and Eq. (2.9), we get the matrices F_0 and F_1 that define the algebraic measure corresponding to μ'. The matrix corresponding to 0, the first element of Ω, is given by

$$\begin{aligned} F_0 &= \sum_{(i,k),\Phi'((i,k))=0} E_{(i,k)} = \sum_i E_{(i,0)} \\ &= \begin{pmatrix} q_{00}x_0^0 & q_{00}x_0^1 & q_{01}x_1^0 & q_{01}x_1^1 \\ 0 & 0 & 0 & 0 \\ q_{10}x_0^0 & q_{10}x_0^1 & q_{11}x_1^0 & q_{11}x_1^1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

and similarly for 1, the second element of Ω.

The hidden Markov process then gives us the same probabilities as the eigenvalues in Eq. (2.3)

$$\begin{aligned} p((k_1,\ldots,k_n)) &= \langle \tau | F_{k_1} \ldots F_{k_n} \sigma \rangle \\ &= \sum_{i_1,\ldots,i_n} \tau_{i_1,k_1} q_{i_1 i_2} \ldots q_{i_{n-1} i_n} x_{i_2}^{k_2} \ldots x_{i_n}^{k_n} \ . \end{aligned}$$

Note that according to our discussion in Section 2.4, the vector τ is the stationary distribution of the full matrix E. Using Eq. (2.12), one can see that the invariant distribution τ is in fact $\tau_{(i,k)} = \gamma_i x_i^k$, so the probabilities of the hidden Markov process coincide with the eigenvalues in Eq. (2.3).

2.5.2 Support of ϕ

Having constructed the correct algebraic measure, we can determine T_μ explicitly and use it to simplify the corresponding invariant measure ϕ.

The expression for T_μ, as can be found in [21], is

$$(T_\mu f)(\hat{\nu}) = \sum_{a \in \Omega} \langle \hat{\nu} | F_a 1 \rangle f\left(\frac{F_a^* \hat{\nu}}{\langle \hat{\nu} | F_a 1 \rangle} \right) \ ,$$

where $\hat{\nu}$ is any 4-dimensional vector such that $\langle\hat{\nu}|1\rangle = 1$, $|1\rangle$ is a vector with all elements equal to 1 and f is a continuous real-valued function on the set of such vectors. For the case of our hidden Markov measure, the form of this transformation can be greatly simplified. Due to the stochasticity of the matrix E, we have the following:

$$F_0|1\rangle = \begin{pmatrix}1\\0\\1\\0\end{pmatrix} \text{ and } F_1|1\rangle = \begin{pmatrix}0\\1\\0\\1\end{pmatrix} .$$

If we furthermore denote the four row vectors of E by $\hat{\mu}_0$, $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$, we can write

$$F_0^*\hat{\nu} = \nu_0\hat{\mu}_0 + \nu_2\hat{\mu}_2 \text{ and } F_1^*\hat{\nu} = \nu_1\hat{\mu}_1 + \nu_3\hat{\mu}_3 .$$

On top of this, $\mu_0 = \mu_1$ and $\mu_2 = \mu_3$, so the total form of the transformation becomes

$$\begin{aligned}(T_\mu f)(\hat{\nu}) =&(\nu_0+\nu_2)f\left(\frac{\nu_0\hat{\mu}_0+\nu_2\hat{\mu}_2}{\nu_0+\nu_2}\right)\\ &+(\nu_1+\nu_3)f\left(\frac{\nu_1\hat{\mu}_0+\nu_3\hat{\mu}_2}{\nu_1+\nu_3}\right) .\end{aligned}$$

From this form of the transformation, we can already deduce a restriction on the support of ϕ. The support of ϕ is restricted to the set of convex combinations of $\hat{\mu}_1$ and $\hat{\mu}_3$

$$\text{supp}(\phi) \subset \{a\hat{\mu_0} + (1-a)\hat{\mu_2} \,|\, a \in [0,1]\} := S .$$

To show this, let's suppose that $\hat{\nu} \in \text{supp}(\phi)$ and $\hat{\nu} \notin S$. Take $\zeta_{\hat{\nu}}$ a function such that $\zeta_{\hat{\nu}}(\hat{s}) = 0$ for all $\hat{s} \in S$ and $\zeta_{\hat{\nu}}(\hat{\nu}) \neq 0$. As ϕ has non-zero weight in ν, we have $\phi(\zeta_{\hat{\nu}}) \neq 0$. However, if we apply the invariance of ϕ under T_μ, we get

$$\begin{aligned}0 \neq&\phi(\zeta_{\hat{\nu}}) = \phi(T_\mu\zeta_{\hat{\nu}}) = \int \phi(d\nu)T_\mu(\zeta_{\hat{\nu}}(\nu))\\ =&\int \phi(d\nu)\Big[(\nu_0+\nu_2)\zeta_{\hat{\nu}}\Big(\frac{\nu_0\hat{\mu}_0+\nu_2\hat{\mu}_2}{\nu_0+\nu_2}\Big)\\ &+(\nu_1+\nu_3)\zeta_{\hat{\nu}}\Big(\frac{\nu_1\hat{\mu}_0+\nu_3\hat{\mu}_2}{\nu_1+\nu_3}\Big)\Big] = 0 ,\end{aligned}$$

since the arguments to $\zeta_{\hat{\nu}}$ run over the set S, where the function is zero.

As the measure is non-zero only on S, the convex combinations of $\hat{\mu}_0$ and $\hat{\mu}_2$, we can write it as a measure on the interval $[0,1]$:

$$\phi(f) = \int_0^1 d\lambda(a) f(a\hat{\mu}_1 + (1-a)\hat{\mu}_3) \ . \tag{2.13}$$

The transformation T_μ now acting on the measure λ becomes:

$$\begin{aligned}
\phi(T_\mu f) &= \int \phi(d\nu) \ \Big[(\nu_1+\nu_3) f\Big(\frac{\nu_1\hat{\mu}_1 + \nu_3\hat{\mu}_3}{\nu_1+\nu_3}\Big) \\
&\qquad\qquad + (\nu_2+\nu_4) f\Big(\frac{\nu_2\hat{\mu}_1 + \nu_4\hat{\mu}_3}{\nu_2+\nu_4}\Big)\Big] \\
&= \int_0^1 d\lambda(a) \ \Big[(\hat{\mu}_{a,1}+\hat{\mu}_{a,3}) f\Big(\frac{\hat{\mu}_{a,1}\hat{\mu}_1 + \hat{\mu}_{a,3}\hat{\mu}_3}{\hat{\mu}_{a,1}+\hat{\mu}_{a,3}}\Big) \\
&\qquad\qquad + (\hat{\mu}_{a,2}+\hat{\mu}_{a,4}) f\Big(\frac{\hat{\mu}_{a,2}\hat{\mu}_1 + \hat{\mu}_{a,4}\hat{\mu}_3}{\hat{\mu}_{a,2}+\hat{\mu}_{a,4}}\Big)\Big] \qquad (2.14) \\
&= \int_0^1 d\lambda(a) \ \alpha_1(a) f_S(\beta_1(a)) + \alpha_2(a) f_S(\beta_2(a)) \ , \qquad (2.15)
\end{aligned}$$

where

$$\begin{aligned}
\hat{\mu}_a &= a\hat{\mu}_1 + (1-a)\hat{\mu}_3 \\
\alpha_1(a) &= \hat{\mu}_{a,1} + \hat{\mu}_{a,3} \\
\alpha_2(a) &= \hat{\mu}_{a,2} + \hat{\mu}_{a,4} \\
\beta_1(a) &= \frac{\hat{\mu}_{a,1}}{\hat{\mu}_{a,1}+\hat{\mu}_{a,3}} \\
\beta_2(a) &= \frac{\hat{\mu}_{a,2}}{\hat{\mu}_{a,2}+\hat{\mu}_{a,4}} \qquad \text{and} \\
f_S(a) &= f(\hat{\mu}_a)
\end{aligned}$$

The action of T_μ on f_S can be seen as taking two subintervals of f_S, scaling the function on these intervals to the unit interval and taking convex combinations of the resulting functions, where the weights depend on the variable a. Likewise, the dual action of T_μ on the measure λ is a contraction of the measure to two

subintervals. By repeated applications of this transformation, we get the invariant measure: $\phi = \lim_{n\to\infty}(T_\mu)^n\phi_0$. Repeated application means that at each iteration the support of ϕ is further restricted to finer subsets. Thus the support of ϕ will be on a Cantor-like set. These iterated function systems have been studied in more detail in [62].

2.6 Numerics

The invariant measure can numerically be approached by taking any initial probability measure and applying the transformation T_μ a sufficiently large number of times. Any measure will converge to the invariant measure, but some choices will give a lower computational complexity. A convenient choice is a sum of Dirac delta measures as these can easily be represented numerically by their weights and positions.

Figure 2.2 shows a plot of the capacity of a memory depolarizing channel, where the switching between the channels is taken to be symmetric and parametrized by a parameter s, the eigenvalue of the transition matrix different from 1:

$$Q = \begin{pmatrix} \frac{1+s}{2} & \frac{1-s}{2} \\ \frac{1-s}{2} & \frac{1+s}{2} \end{pmatrix} \tag{2.16}$$

Secondly, we parametrize the error probabilities by their average and difference: $x_0^0 = a + d$, $x_1^0 = a - d$. In this plot a is varied over the allowed interval $[\frac{1}{3}, 1]$ and d is taken as the largest possible value such that both error probabilities still lie in this interval.

The expected increase in channel capacity with increasing noise correlations is evident. The results were compared to a Monte Carlo calculation of the capacity and while the results were the same, the dynamical system calculation takes only a fraction of the time of a full Monte Carlo calculation.

2.7 Memory dependence

Explicitly determining the invariant measure of the dynamical system proves to be very difficult due to its fractal nature. Nevertheless it is possible to prove more general characteristics of the measure.

In the numerical simulation, we observe an increase of capacity with a stronger correlated memory. This behaviour has been analytically proven for the binary symmetric channel [48]. Here we present an alternative proof of this property that generalizes easily to the more general setting of algebraic measures.

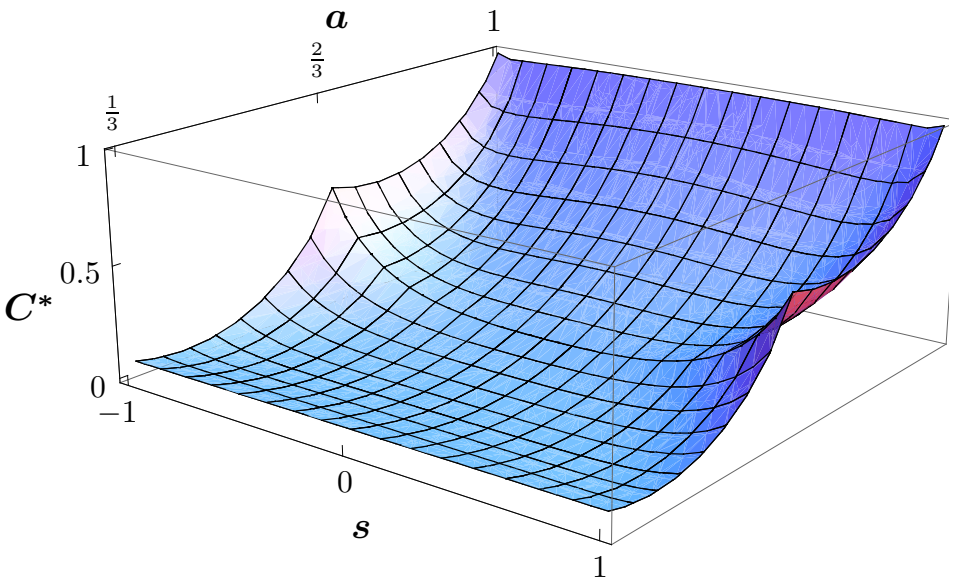

Figure 2.2: Capacity for maximally different sub-channels increases with memory

The rationale behind the proof is that by probabilistically interrupting the memory of the channel, we get a new hidden Markov measure with a memory that correlated less strongly. Since the process is a probabilistic mix, the entropy can be shown to increase and hence the capacity is lower. Hence less strongly correlated gives a lower channel capacity.

Proposition 1. *Given an algebraic measure μ generated by manifestly positive matrices E_i and manifestly positive vectors $|\sigma\rangle$ and $|\tau\rangle$. Let A the manifestly positive matrix*

$$A = \alpha \mathbb{1} + (1-\alpha)|\sigma\rangle\langle\tau| \,.$$

For the algebraic measure μ', generated by $E'_i = \sqrt{A}E_i\sqrt{A}$ and $|\sigma'\rangle = \sqrt{A}|\sigma\rangle = |\sigma\rangle$ and $|\tau'\rangle = (\sqrt{A})^*|\tau\rangle = |\tau\rangle$, we have

$$\mathsf{S}(\mu'_n) \geq \mathsf{S}(\mu_n) \qquad \forall n \in \mathbb{N}_0 \,.$$

Proof. The probability of a path $(\omega_1, \ldots, \omega_n)$ in the new process μ' is given by

$$\mu'_n((\omega_1, \ldots, \omega_n)) = \langle\sigma'|E'_{\omega_1} \ldots E'_{\omega_n}|\tau'\rangle \tag{2.17}$$

$$= \langle\sigma|AE_{\omega_1}A \ldots AE_{\omega_n}A|\tau\rangle \tag{2.18}$$

This can be written as a mix of product states, for example for the length two probabilities

$$\begin{aligned}\mu_2'((\omega_1,\omega_2)) &= \langle\sigma|E_{\omega_1}AE_{\omega_2}|\tau\rangle \\ &= \alpha\langle\sigma|E_{\omega_1}E_{\omega_2}|\tau\rangle + (1-\alpha)\langle\sigma|E_{\omega_1}|\tau\rangle\langle\sigma|E_{\omega_2}|\tau\rangle\,.\end{aligned}$$

In the general case the probabilities can be written as a sum over all partitions π_n of the set $\{1,\ldots,n\}$:

$$\mu_n' = \sum_{\pi_n} p(\pi_n)\mu_{|\pi_n(1)|}\times\ldots\times\mu_{|\pi_n(k)|}\,, \tag{2.19}$$

where $\pi_n(i)$ denotes the i-th subset in the partition and $|\pi_n(i)|$ denotes its cardinality. The convex coefficients $p(\pi_n)$ are determined by the number of times we switch between different processes $\mu_{|\pi_n(i)|}$. There are $k-1$ such switches in a partition with k subsets, so

$$p(\pi_n) = (1-\alpha)^{(k-1)}\alpha^{(n-k)}\,.$$

By using the concavity and the sub-additivity of the Shannon entropy, we get the desired result:

$$\begin{aligned}\mathsf{S}(\mu_n') &\geq \sum_{\pi_n} p(\pi_n)\left(\mathsf{S}(\mu_{|\pi_n(1)|}) + \ldots\mathsf{S}(\mu_{|\pi_n(k)|})\right) \\ &\geq \sum_{\pi_n} p(\pi_n)\mathsf{S}(\mu_n) \\ &= \mathsf{S}(\mu_n)\end{aligned}$$

□

Corrolary 1. *For the memory channel we have considered in this chapter, the classical capacity increases as the absolute value of the memory parameter s.*

Proof. What we will show is that the process μ' constructed in the proof of Proposition 1 is in fact the hidden Markov process corresponding to the memory channel with memory parameter αs. The smaller the mixing parameter α, the higher the entropy by Proposition 1 and hence the lower the capacity.

The probabilities of the original process μ are given by (cf. Eq. 2.3)

$$\mu((\omega_1,\ldots,\omega_n)) = \sum_{i_1,\ldots,i_n}\gamma_{i_1}q_{i_1i_2}\ldots q_{i_{n-1}i_n}x_{i_1}^{\omega_1}\ldots x_{i_n}^{\omega_n}$$

By Eq. 2.19, the new process has probabilities

$$\begin{aligned}&\mu'((\omega_1,\ldots,\omega_n))\\&=\sum_{i_1,\ldots,i_n}\left(\sum_{\pi_n}p(\pi_n)p_Q((i_1,\ldots,i_{|\pi(1)|}))\ldots p_Q((i_{n-|\pi(k)|+1},\ldots,i_n))\right)\\&\qquad x_{i_1}^{\omega_1}\ldots x_{i_n}^{\omega_n},\end{aligned}$$

where p_Q are the Markov channel switching probabilities

$$p_Q((i_1,\ldots,i_l))=\gamma_{i_1}q_{i_1i_2}\ldots q_{i_{l-1}i_l}$$

It is not difficult to see that the sum of switching probabilities between brackets in the previous equation is in fact the expansion of Markov switching probabilities with transition matrix

$$Q'=\alpha Q+(1-\alpha)\begin{bmatrix}\gamma_0 & \gamma_1\\ \gamma_0 & \gamma_1\end{bmatrix}.$$

Take for example the length 3 probabilities:

$$\begin{aligned}&\alpha^2\gamma_{i_1}q_{i_1i_2}q_{i_2i_3}+\alpha(1-\alpha)\gamma_{i_1}q_{i_1i_2}\gamma_{i_3}\\&\quad+(1-\alpha)\alpha\gamma_{i_1}\gamma_{i_2}q_{i_2i_3}+(1-\alpha)^2\gamma_{i_1}\gamma_{i_2}\gamma_{i_3}\\&=\gamma_{i_1}(\alpha q_{i_1i_2}+(1-\alpha)\gamma_{i_2})(\alpha q_{i_2i_3}+(1-\alpha)\gamma_{i_3})\end{aligned}$$

In the case of a transition matrix of the form of Eq. 2.16, this mixing simply means scaling the memory parameter s by α. □

2.8 Non-forgetful limit

To complete the discussion concerning correlations we need to look at the two extreme cases: $s=1$, corresponding to the case where a sub-channel is selected and used for every channel use afterwards, and $s=-1$, corresponding to the case where the choice of sub-channel is flipped with every channel use. Therefore, in constructing the overall channel and taking into account the initial random channel selection, we just have the mixing of two n-use channels. Specifically, in the $s=1$ case, we have the mixing of the two n-fold tensor products of the two sub-channels separately and in the $s=-1$ case we have the mixing of two n-use channels where

each deterministically alternates between the sub-channels but starting with a different sub-channel.

Both these extreme cases are non-forgetful since the initial sub-channel selection (the initial noise) is remembered and the forgetful Holevo capacity theorem no longer applies (the Markov selection matrix is periodic in the $s = -1$ case and reducible in the $s = 1$ case). For $s = -1$ the capacity can be calculated using [17] and agrees with the limit of the forgetful approach, the capacity is the average capacity of the two sub-channels separately. However, for $s = 1$ case there is a discontinuity and the capacity suddenly drops to the minimum capacity of the sub-channels [16].

The intuition is that in the $s = -1$ case the deterministic flip can be used to determine 'on-the-fly' which sub-channel is being used and then it is the same as using the two channels separately each half the time, so the capacity must be the average capacity. For the $s = 1$ case once you have the poorer channel you are stuck with it forever and so because of the mixture you can only guarantee the lower capacity.

2.9 Conclusion

We have constructed a simple forgetful quantum memory channel. Using results on the capacity of quantum memory channels, we can turn the capacity calculation into a calculation of the entropy density of a hidden Markov process. The entropy density of this hidden Markov process can be studied through a dynamical system defined by its conditional probabilities. The entropy of the process is calculated from the invariant measure of the dynamical system.

By studying the transformation involved in this dynamical system, it is possible to greatly reduce the complexity of the problem. This allows for easy numerical calculations of the channel capacity.

A possibility for further research on this topic is the study of memory channels where the memory state process is a quantum Markov process instead of a classical process. The entropy density calculation of some quantum Markov processes is worked out further on in this thesis.

Chapter 3

Quantum Markov chains

3.1 Introduction

In the previous chapter a classical hidden Markov process appeared in the context of quantum memory channels. The method used to calculate the entropy density of such hidden Markov processes relies only on the strong sub-additivity of the entropy and the translation-invariance on the one hand and the structure of the correlations in the state on the other. The first two properties ensure that the entropy density equals the asymptotic entropy rate. The last property then is used to simplify the calculation of this entropy rate.

The strong sub-additivity and translation-invariance are properties that translate directly to quantum systems. A good candidate for a quantum Markov process is less clear however. In this chapter we will try to explain some of the issues that arise in translating correlation properties to the quantum case.

There are several different ways to single out the Markov processes from among the set of classical processes. First of all, the Markov processes are those processes where the dependence on the past is limited to one time step. Another possibility is the well-known transfer matrix construction, where the transfer matrix holds the transition probabilities. Yet another way is looking at extensions of two-site probability distributions that agree on one site. Such distributions can be extended to processes, leading to a convex set of processes. The extension with maximal entropy corresponds to the Markov process.

These constructions can however not be put into a quantum context in a straightforward manner. The first two descriptions rely on the use of conditional states, for which there is no clear extension to quantum systems [64, 51]. As we will

show in the next section, the state extension construction is also not generalizable to quantum systems. To introduce quantum Markov processes, we will use another tool, called the conditional state space. This is a very general construction that allows to characterize correlations between two quantum parties.

The main application of the conditional state spaces lies in the construction of so called finitely correlated states on a spin chain [22]. These are translation-invariant states on the chain that have a finite dimensional conditional state space with respect to the division of the chain into two semi-infinite parts. When restricted to the classical case, these states correspond exactly to the hidden Markov processes.

It can be shown that all finitely correlated states can be constructed using a completely positive map that plays the role of the transition matrix [22]. In this chapter we will look at a specific example of such finitely correlated states for qubit chains, where we impose a lot of symmetry on the defining completely positive map. In the following chapters we will work out constructions similar to the ones presented in this chapter for fermions.

Finitely correlated states are interesting states to study for many different reasons. First of all the pure states among them are ground states of quantum spin systems that can be explicitly constructed [22]. Not a lot of states with these properties are known. Secondly, finitely correlated states can be used to quantify the randomness of its corresponding completely positive map. The randomness is given by the randomness of the produced state, determined by its entropy density. This is the analogue of measuring randomness of a transition matrix by the entropy density of the corresponding Markov measure. This method of measuring the randomness of a quantum channel can be seen as an alternative to minimal output entropy[37]. Finally, pure finitely correlated states also have been used to numerically simulate spin systems in the so called density matrix renormalization group method [58], but this will not be discussed here.

Finitely correlated states are well-known in quantum information theory and solid state physics. The pure finitely correlated states are also known as matrix product states. In recent years a lot of progress has been made on topics such as their entanglement [46], finite-dimensional and non-translation-invariant generalizations [50] and their applicability as trial states for numerical calculations of ground state properties. Here we present a new view emphasising the importance of conditional state spaces. The thorough investigation of conditional state spaces and possible applications to other concepts such as entanglement is a new contribution.

The ideas presented in this chapter have been published in [25, 24].

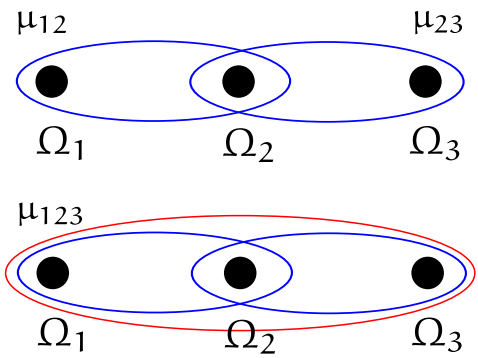

Figure 3.1: The classical state extension problem.

3.2 State extensions

In this section we first describe some properties of classical Markov processes. We look at their construction and the behaviour of the Shannon entropy. We then describe some related questions that arise in the context of quantum systems.

3.2.1 Extending classical states

Let us first look at ways to construct classical Markov processes. The configuration space of a classical register with d states is just a finite set $\Omega = \{1, 2, \ldots, d\}$. The states are length-$d$ probability vectors

$$\boldsymbol{\mu} = \{\mu(1), \mu(2), \ldots, \mu(d)\}, \;\; \mu(\epsilon) \geq 0, \;\; \sum_{\epsilon} \mu(\epsilon) = 1.$$

The state space is a simplex and the extreme points are the Dirac measures δ_ϵ assigning probability 1 to the configuration ϵ. The Shannon entropy

$$\mathsf{H}(\boldsymbol{\mu}) := -\sum_{\epsilon} \mu(\epsilon) \log \mu(\epsilon)$$

quantifies the randomness in the state.

We can now consider the following state extension problem. Suppose that we are given two probability vectors $\boldsymbol{\mu}_{12}$ and $\boldsymbol{\nu}_{23}$ that agree on the middle system: $\boldsymbol{\mu}_2 = \boldsymbol{\nu}_2$. Can we find a joint extension for $\boldsymbol{\mu}_{12}$ and $\boldsymbol{\nu}_{23}$? More explicitly: can we find a state $\boldsymbol{\xi}_{123}$ on Ω_{123} that restricts to $\boldsymbol{\mu}_{12}$ on Ω_{12} and to $\boldsymbol{\nu}_{23}$ on Ω_{23}?

This is indeed possible and clearly the set of joint extensions $\boldsymbol{\xi}_{123}$ is convex and compact. We can therefore refine the question and ask for a joint extension of

maximal entropy. A straightforward computation yields this extension:

$$\boldsymbol{\rho}_{123}(\epsilon_1,\epsilon_2,\epsilon_3) := \frac{\boldsymbol{\mu}_{12}(\epsilon_1,\epsilon_2)\,\boldsymbol{\nu}_{23}(\epsilon_2,\epsilon_3)}{\boldsymbol{\mu}_2(\epsilon_2)} = \frac{\boldsymbol{\mu}_{12}(\epsilon_1,\epsilon_2)\,\boldsymbol{\nu}_{23}(\epsilon_2,\epsilon_3)}{\boldsymbol{\nu}_2(\epsilon_2)}. \tag{3.1}$$

Actually, $\boldsymbol{\rho}_{123}$ saturates the strong sub-additivity inequality:

$$\mathsf{H}(\boldsymbol{\rho}_{123}) + \mathsf{H}(\boldsymbol{\rho}_2) = \mathsf{H}(\boldsymbol{\rho}_{12}) + \mathsf{H}(\boldsymbol{\rho}_{23}).$$

Unsurprisingly, there is a direct connection with thermal equilibrium states. If we introduce Hamiltonians

$$\boldsymbol{\mu}_{12} = \mathrm{e}^{-h_{12}},\ \ \boldsymbol{\nu}_{23} = \mathrm{e}^{-h_{23}},\ \ \text{and}\ \ \boldsymbol{\mu}_2 = \boldsymbol{\nu}_2 = \mathrm{e}^{-h_2}$$

then

$$\boldsymbol{\rho}_{123} = \mathrm{e}^{-(h_{12}+h_{23}-h_2)}.$$

Let us start with a two-party probability vector $\boldsymbol{\mu}$ that is shift-invariant in the sense that

$$\sum_{\epsilon_2} \boldsymbol{\mu}(\epsilon,\epsilon_2) = \sum_{\epsilon_1} \boldsymbol{\mu}(\epsilon_1,\epsilon)\ \text{ for all } \epsilon. \tag{3.2}$$

We can repeatedly apply the Markov extension procedure (3.1) to get a stationary process

$$\boldsymbol{\omega}(\epsilon_0,\epsilon_1,\ldots,\epsilon_n) = \frac{\boldsymbol{\mu}(\epsilon_0,\epsilon_1)\,\boldsymbol{\mu}(\epsilon_1,\epsilon_2)\,\cdots\,\boldsymbol{\mu}(\epsilon_{n-1},\epsilon_n)}{\boldsymbol{\mu}(\epsilon_1)\,\boldsymbol{\mu}(\epsilon_2)\,\cdots\,\boldsymbol{\mu}(\epsilon_{n-1})}. \tag{3.3}$$

Another procedure is to start with a $d\times d$ stochastic matrix T. The entry $T_{\epsilon_1\epsilon_2}$ is the probability for jumping from state ϵ_1 to ϵ_2, therefore

$$T_{\epsilon_1\epsilon_2} \geq 0\ \text{ and }\ \sum_{\epsilon_2} T_{\epsilon_1\epsilon_2} = 1. \tag{3.4}$$

The invariant state $\boldsymbol{\mu}$ is a row vector determined by $\boldsymbol{\mu}\,T = \boldsymbol{\mu}$. The Markov process is now obtained by putting

$$\boldsymbol{\omega}(\epsilon_0,\epsilon_1,\ldots,\epsilon_n) = \boldsymbol{\mu}(\epsilon_0)\,T_{\epsilon_0\epsilon_1}\,\cdots\,T_{\epsilon_{n-1}\epsilon_n}. \tag{3.5}$$

Both constructions (3.3) and (3.5) agree if we put

$$T_{\epsilon_1\epsilon_2} = \frac{\boldsymbol{\mu}(\epsilon_1,\epsilon_2)}{\boldsymbol{\mu}(\epsilon_1)}.$$

The randomness of the Markovian evolution T can be quantified by the minimal output entropy.

$$\mathsf{H}^{\min}(\Gamma) = \min\Big(\big\{\mathsf{H}(\Gamma(\rho)) \;\big|\; \rho \text{ state}\big\}\Big).$$

The amount of randomness in the Markov process ω up to time n is quantified by the entropy

$$\mathsf{H}_n = \mathsf{H}\big(\boldsymbol{\omega}_{\{0,1,\ldots,n\}}\big).$$

The rows of a stochastic matrix T are probability vectors. The minimal output entropy of T is simply

$$\mathsf{H}^{\min}(T) = \text{smallest entropy of rows of } T$$

while the entropy of the process is a smooth version of this quantity

$$\begin{aligned}\mathsf{h} &= \boldsymbol{\mu}\text{-average of entropies of rows of } T \\ &= \sum_{ij} \mu(i)(-T_{ij}\log T_{ij})\,,\end{aligned}$$

as we have seen in Section 2.4.2.

3.2.2 Extending quantum states

When turning to quantum state extension the situation gets more complicated. Quantum states allow for more freedom, as they exhibit correlations that are not present in classical systems, but this also makes it more difficult to have two states agree on a common part, as we did with classical states in the previous section.

States on a full matrix algebra $\mathcal{M}_d$ can be identified with density matrices: non-negative matrices with trace one. The convex set of density matrices is very unlike a simplex. A density matrix that is not an extreme point of the state space, i.e. that is not a one-dimensional projector, allows many decompositions in extreme states. In contrast with classical systems such a state can therefore not be seen as a well-defined ensemble of pure states. We need d^2-1 real parameters to describe the state space of $\mathcal{M}_d$ while $2d-2$ parameters suffice to label the pure states. This means that the boundary of the state space contains many flat subsets. Nevertheless the pure states form a very nice smooth manifold. The case of a single qubit is exceptional: its state space is affinely isomorphic to the Bloch ball by the standard parametrization

$$\rho = \tfrac{1}{2}\,(\mathbb{1} + \mathbf{x}\cdot\boldsymbol{\sigma}),\;\; \mathbf{x}\in\mathbb{R}^3,\;\; |\mathbf{x}|\leq 1. \tag{3.6}$$

In this case, every point of the boundary is also an extreme point. For higher d, a smooth parametrization of the pure states does not define a boundary of a convex set.

An important property that holds both for classical and quantum systems is that if the marginal ρ_1 of a bipartite state is pure then $\rho_{12} = \rho_1 \otimes \rho_2$. This is an important ingredient of the theory: namely it allows a system to be isolated from the rest of the universe. At the same time it is also a severe constraint on quantum systems because there are plenty of pure states of a composite system. In particular the restriction of an entangled pure state can never be pure and we can therefore not separate a party of an entangled system from the outside world.

Factorisation of extensions of pure states has also a bearing on joint extensions of states as considered in the previous section [68]. Indeed, suppose that ρ_{12} and ρ_{23} are pure and agree on the middle system,

$$\mathrm{Tr}_1 \rho_{12} = \mathrm{Tr}_3 \rho_{23} ,$$

which is easily feasible. A joint extension ρ_{123} can then only exist for ρ_{12} and ρ_{23} pure product states. Therefore a generic pure two-party state with inner shift-invariance as in (3.2) cannot be extended.

Suppose that density matrices ρ_{12} and σ_{23} agree on the middle system and can be jointly extended. The set of extensions is still convex and compact and so we can still look for the maximal entropy extension. Finding the maximal entropy extension is hard, however, because generally

$$\big[\rho_{12} \otimes \mathbb{1}_3 \,,\, \mathbb{1}_1 \otimes \sigma_{23}\big] \neq 0$$

or, equivalently, if ρ_{12} and σ_{23} are equilibrium states corresponding to Hamiltonians h_{12} and h_{23}

$$\mathrm{Tr}_3 \exp\big(h_{12} + h_{23}\big) \not\propto \exp\big(h_{12} + h_2\big).$$

Moreover, the maximal entropy extension will not saturate the strong sub-additivity.

Actually, a nice characterisation of equality in strong sub-additivity for a state ρ_{123} on a space $\mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ in terms of decompositions of the middle space has been obtained in [29]. The necessary and sufficient condition is that the middle Hilbert space $\mathcal{H}_2$ decomposes as

$$\mathcal{H}_2 = \bigoplus_\alpha \mathcal{H}^\alpha_{\mathrm{left}} \otimes \mathcal{H}^\alpha_{\mathrm{right}} \;\text{ and }\; \rho_{123} = \bigoplus_\alpha \lambda_\alpha \, \rho^\alpha_{12} \otimes \rho^\alpha_{23}$$

with $\{\lambda_\alpha\}$ convex weights.

A Qubit Example with SU(2)-symmetry

An example of the limitations imposed on quantum state extensions can be worked out for qubits with a SU(2)-symmetry. We look at the amount of entanglement

that can be achieved between two subsequent spins in a chain that is invariant under SU(2) transformations. We compare the amount of entanglement that can be theoretically be attained in the general case and the amount attainable in some cases where an explicit state is given. The general case gives a significantly higher value, showing the need for a more extensive class of explicit translation-invariant quantum chains.

In order to impose SU(2)-symmetry on single qubit observables we use the adjoint representation of SU(2)

$$\mathrm{Ad}(U) : A \mapsto U\,A\,U^*,\ \ U \in \mathrm{SU}(2),\ \ A \in \mathcal{M}_2.$$

This is a reducible representation that decomposes into a spin 0 and a spin 1 irrep:

$$\mathcal{M}_2 = \mathbb{C}\mathbb{1} \oplus \mathbb{C}\boldsymbol{\sigma}.$$

The action of the adjoint $\mathrm{Ad}(U)$ on the Bloch ball is a rotation about an arbitrary axis. Hence the only SU(2)-invariant state on $\mathcal{M}_2$ is the uniform state

$$\rho = \tfrac{1}{2}\,\mathbb{1},$$

at the center of the ball.

For 2 qubits $\mathrm{Ad}(U \otimes U)$ decomposes into 2 spin 0, 3 spin 1 and 1 spin 2 irrep. There exists now a one parameter family of SU(2)-invariant states

$$\rho = \tfrac{1}{3}\,(1-\lambda)(\mathbb{1} - \mathbf{p}) + \lambda\,\mathbf{p},\ \ 0 \le \lambda \le 1.$$

Here $\mathbf{p}$ is the projector on the singlet vector $\frac{1}{\sqrt{2}}\,(|10\rangle - |01\rangle)$ in $\mathbb{C}^2 \otimes \mathbb{C}^2$. This projector commutes with every unitary of the form $U \otimes U$ and every two-qubit observable that is SU(2)-invariant is a linear combination of $\mathbf{p}$ and $\mathbb{1}$. Clearly, SU(2)-invariant two-qubit states satisfy

$$0 \le \langle \mathbf{p} \rangle = \lambda \le 1.$$

The two-qubit state ρ is separable for $0 \le \lambda \le \frac{1}{2}$ and entangled for $\frac{1}{2} < \lambda \le 1$. Hence the expectation value of this projector for a certain process tells us how much bipartite entanglement between two neighbouring spins is attainable.

For 3 qubits the SU(2)-invariant states can still easily be determined but things become more complicated with increasing number of parties. Let

$$\mathbf{p}_1 = \mathbf{p} \otimes \mathbb{1} \ \text{ and } \ \mathbf{p}_2 = \mathbb{1} \otimes \mathbf{p} \ \text{ and put}$$

$$\mathbf{q} = \tfrac{4}{3}\,\big(\mathbf{p}_1 + \mathbf{p}_2 - \mathbf{p}_1\mathbf{p}_2 - \mathbf{p}_2\mathbf{p}_1\big).$$

The algebra of three-qubit observables that are SU(2)-invariant is not Abelian. It can be decomposed into a direct sum of $\mathbb{C}$ and $\mathcal{M}_2$ where $\mathbb{C}$ is identified with $\mathbb{C}\mathbf{q}$ and

$\mathcal{M}_2$ with the algebra generated by $\mathbf{p}_1$ and $\mathbf{p}_2$, not including $\mathbb{1}$. An SU(2)-invariant three-qubit state is of the form

$$\rho = \tfrac{1}{4}(1-\lambda)(\mathbb{1}-\mathbf{q}) + \lambda\big(a\,\mathbf{p}_1 + b\,\mathbf{p}_2 + c\,\mathbf{p}_1\mathbf{p}_2 + \overline{c}\,\mathbf{p}_2\mathbf{p}_1\big)$$

with

$$0 \le \lambda \le 1,\ \ a, b \in \mathbb{R},\ \ c \in \mathbb{C},\ \ 2a + 2b + \Re\mathfrak{e}(c) = 1,\ \text{ and }\ |c|^2 \le 4ab.$$

If we look for a SU(2)-invariant three-party state with partial shift-invariance, then we find the following constraint on the expectation of $\mathbf{p}$

$$0 \le \langle \mathbf{p}_1 \rangle = \langle \mathbf{p}_2 \rangle \le 3/4. \tag{3.7}$$

SU(2) and shift-invariant states on more parties will satisfy stronger upper bounds on the expectations of $\mathbf{p}$, see (3.7). Ultimately, if we look for a shift-invariant extension on the full half-chain then, using the Bethe Ansatz [7, 34, 33], one can show that

$$0 \le \langle \mathbf{p} \rangle \le \log 2 \approx 0.69. \tag{3.8}$$

We may look for the largest expectation value of $\mathbf{p}$ that can be obtained within classes of shift-invariant states that can easily be handled. Consider as a first example point-wise limits of shift-invariant product states. Such states are actually invariant under arbitrary finite permutations of sites on the half-chain and are usually called exchangeable. Using the Bloch parametrization (3.6) and

$$\mathbf{p} = \tfrac{1}{4}(\mathbb{1} - \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2) \ \text{ with } \ \boldsymbol{\sigma}_1 = \boldsymbol{\sigma} \otimes \mathbb{1} \text{ and } \boldsymbol{\sigma}_2 = \mathbb{1} \otimes \boldsymbol{\sigma}$$

we have to maximize

$$\mathbf{x} \in \mathbb{R}^3 \mapsto \tfrac{1}{16}\ \mathrm{Tr}\big[(\mathbb{1} + \mathbf{x}\cdot\boldsymbol{\sigma}_1)(\mathbb{1} + \mathbf{x}\cdot\boldsymbol{\sigma}_2)(\mathbb{1} - \boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2)\big]\,,$$

subject to the constraint $|\mathbf{x}| \le 1$. It is easily seen that the maximum is reached for $\mathbf{x} = 0$ for which value $\langle \mathbf{p} \rangle = \frac{1}{4}$. Hence

$$\langle \mathbf{p} \rangle \le \tfrac{1}{4} \ \text{ for all exchangeable states.} \tag{3.9}$$

The largest expectation for $\mathbf{p}$ that can be reached within the class of product states is

$$\max\Big(\tfrac{1}{16}\ \mathrm{Tr}\big[(\mathbb{1} + \mathbf{x}_1\cdot\boldsymbol{\sigma}_1)(\mathbb{1} + \mathbf{x}_2\cdot\boldsymbol{\sigma}_2)(\mathbb{1} - \boldsymbol{\sigma}_1\cdot\boldsymbol{\sigma}_2)\big]\Big)$$

subject to the constraint $|\mathbf{x}_1|, |\mathbf{x}_2| \le 1$. The maximum $\frac{1}{2}$ is attained for $\mathbf{x}_1 = -\mathbf{x}_2 = \mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^3$ is an arbitrary vector of length 1. Therefore

$$\langle \mathbf{p} \rangle \le \tfrac{1}{2} \ \text{ for any separable state.}$$

Moreover, this maximum is attained for shift-invariant separable states that are equal weight mixtures of period-2 product states

$$\tfrac{1}{2}\,|e_0\rangle\langle e_0|\otimes|e_1\rangle\langle e_1|\otimes|e_0\rangle\langle e_0|\otimes\cdots+\tfrac{1}{2}\,|e_1\rangle\langle e_1|\otimes|e_0\rangle\langle e_0|\otimes|e_1\rangle\langle e_1|\otimes\cdots\,, \quad (3.10)$$

where $\{e_0, e_1\}$ is any orthonormal basis in $\mathbb{C}^2$. Hence

$$\langle \mathbf{p}\rangle \leq \tfrac{1}{2} \text{ for all shift-invariant separable states}$$

is an optimal upper bound. States of the form (3.10) are extreme shift-invariant states which allow a convex decomposition in clustering period-2 states. This is called Néel order of period 2. The value of $\frac{1}{2}$ for shift-invariant separable states is still not close to the maximum value of $\log 2$. One can get closer by constructing more general quantum processes. A possible class of more general quantum processes is the class of finitely correlated states. They are states on the spin chain for which the correlations across any link can be modelled by a finite dimensional vector space. The way to study such correlations between two parts of a quantum state is introduced in the next section. In Section 3.4 the finitely correlated states are then introduced. We will then return to this example and demonstrate that a higher value can indeed be obtained using such states.

3.3 Conditional state spaces

In the previous sections we have seen how difficult it can be to explicitly construct translation-invariant processes that have non-trivial correlations. Surely, to describe translation-invariant states in general is hopeless. An approach that has proven to be successful is restricting the correlations between two semi-infinite halves of the process to be finite. This is quite a severe restriction, as both halves of the process are infinite, but they nevertheless have finite correlations.

First of all, we need a way of describing correlations between two parties. To study these correlations we consider the influence of perturbations of one party on the other. The states that arise in this way could be called conditional states even if the classical notion of conditioning cannot be extended to the quantum [64]. We first describe the general recipe for constructing these conditional states and work out some applications. We then derive a different description of the conditional state space that has some useful positivity properties.

3.3.1 Construction

We consider here mostly finite dimensional algebras of observables $\mathcal{A}$. These can always be taken to be unital sub-algebras of some complex matrix algebra $\mathcal{M}$,

closed under Hermitian conjugation. Such algebras are direct sums of full matrix algebras and therefore encompass both classical systems with finite state spaces and fully quantum systems with a finite number of accessible levels. The space of complex linear functionals on $\mathcal{A}$ is denoted by $\mathcal{A}^*$ and the pairing between a functional φ and an observable A by $\varphi \cdot A$. Because of the finite dimensions $(\mathcal{A}^*)^* = \mathcal{A}$. The state space of $\mathcal{A}$ is the convex subset $\mathcal{S}(\mathcal{A})$ of normalized, positive, linear functionals. The term state therefore means expectation functional rather than wave function as in standard quantum mechanics.

We now consider bipartite systems. The observables of both parties form algebras $\mathcal{A}_1$ and $\mathcal{A}_2$ and of the composite system $\mathcal{A}_{12} = \mathcal{A}_1 \otimes \mathcal{A}_2$. Generally, subsystems will be correlated and this is encoded in conditional state spaces

$$\mathcal{S}_1 := \big\{A_1 \mapsto \langle A_1 \otimes A_2\rangle_{12} \mid A_2 \in \mathcal{A}_2^+,\ \langle \mathbb{1}_1 \otimes A_2\rangle_{12} = 1\big\} \text{ and}$$
$$\mathcal{S}_2 := \big\{A_2 \mapsto \langle A_1 \otimes A_2\rangle_{12} \mid A_1 \in \mathcal{A}_1^+,\ \langle A_1 \otimes \mathbb{1}_2\rangle_{12} = 1\big\}. \tag{3.11}$$

$\mathcal{S}_i$ is a compact convex subset of $\mathcal{S}(\mathcal{A}_i)$.

We may also consider the linear spaces of functionals

$$\mathcal{V}_1 := \big\{A_1 \mapsto \langle A_1 \otimes A_2\rangle_{12} \mid A_2 \in \mathcal{A}_2\big\} \text{ and}$$
$$\mathcal{V}_2 := \big\{A_2 \mapsto \langle A_1 \otimes A_2\rangle_{12} \mid A_1 \in \mathcal{A}_1\big\}. \tag{3.12}$$

As any element in a C*-algebra is a linear combination of at most four positive elements, $\mathcal{V}_i$ is spanned by $\mathcal{S}_i$. Mostly, $\mathcal{S}_i$ is a proper subset of the space of positive normalized functionals in $\mathcal{V}_i$. Our main goal is to describe $\mathcal{S}_1$ preferably in terms of a manifestly positive model. This means that we want to generate the elements of $\mathcal{S}_1$ in terms of manifestly positive objects such as states and completely positive maps, see (3.22) for an example of a bipartite finite-dimensional quantum system and Chapter 5 for free Fermionic states.

A state $\langle\ \rangle_{12}$ of a composite system is a linear map from $\mathcal{A}_2$ to $\mathcal{A}_1^*$

$$S : A_2 \in \mathcal{A}_2 \mapsto \big(A_1 \in \mathcal{A}_1 \mapsto \langle A_1 \otimes A_2\rangle\big) \in \mathcal{A}_1^*. \tag{3.13}$$

This map is, moreover, positive. The transposed map S^T from $\mathcal{A}_1$ to $\mathcal{A}_2^*$

$$S^\mathsf{T}(A_1) \cdot A_2 = S(A_2) \cdot A_1,\ \ A_i \in \mathcal{A}_i \tag{3.14}$$

simply swaps the parties. As the rank of a map and its transpose are equal

$$\dim\big(\mathcal{V}_1\big) = \dim\big(\mathcal{V}_2\big) =: n. \tag{3.15}$$

The natural number n is the correlation dimension of $\langle\ \rangle_{12}$.

The conditional state $A_1 \mapsto \langle A_1 \otimes A_2 \rangle$ on $\mathcal{A}_1$ can now be written in the form

$$\langle A_1 \otimes A_2 \rangle_{12} = S^{\mathsf{T}}(A_1) \cdot A_2 = S^{\mathsf{T}}(A_1) \cdot B \tag{3.16}$$

for a suitably chosen B from $S^{\mathsf{T}}(\mathcal{A}_1)^*$, i.e., we have modelled the conditional states on $\mathcal{A}_1$ by a n-dimensional space. However, B does not have to be positive which makes (3.16) not very useful.

3.3.2 Applications

Classical states

The conditional state space defined previously corresponds to the usual conditional probabilities for classical systems. The observables A and B are continuous functions on the configuration space and the state ω is a measure p_{12}. The requirement on B that $\omega_B(\mathbb{1}) = 1$ becomes

$$\omega(\mathbb{1} \otimes B) = \sum_{ij} p_{12}(i,j)B(j) = \sum_j p_2(j)B(j) = 1\,.$$

So by defining $\lambda(j) := p_2(j)B(j)$, λ is a probability distribution since $B \geq 0$.

The conditional states are now given by

$$\omega_B(A) = \sum_{ij} p_{12}(i,j)A(i)B(j) = \sum_i \left(\sum_j \lambda(j) p_{12}(i|j) \right) A(i)\,.$$

Hence the conditional state space is a simplex with the usual conditional probability distributions at the vertices.

Product states

A product state is a state with no correlations. The expectation value factors into a product

$$\omega(A \otimes B) = \omega_1(A)\omega_2(B)\,.$$

For the conditional state ω_B to be a unital, we need that $\omega_2(B) = 1$. This gives us that for every such B, the conditional state equals ω_1

$$\omega_B(A) = \omega_1(A)\omega_2(B) = \omega_1(A)$$

Hence, the conditional state space of a state with trivial correlations is a trivial set with only one point.

Separable states

Separable states are states that can be constructed by local operations and classical communication. They don't exhibit quantum correlations and this property is mirrored in the geometry of the conditional state space.

By Caratheodory's theorem [55], a separable state ρ of a bipartite system $\mathcal{M}_1 \otimes \mathcal{M}_2$ can always be decomposed into a mixture of d pure product states where $d \leq d_1^2 d_2^2$:

$$\rho = \sum_{\alpha=1}^{d} \lambda_\alpha |\varphi_\alpha\rangle\langle\varphi_\alpha| \otimes |\psi_\alpha\rangle\langle\psi_\alpha|. \tag{3.17}$$

Here $\lambda = (\lambda_\alpha)$ is a probability vector and the φ_α and ψ_α are normalized but generally non-orthogonal vectors in $\mathbb{C}^{d_1}$ and $\mathbb{C}^{d_2}$. The conditional states are then of the form

$$A_1 \mapsto \sum_\alpha \mu_\alpha \langle \varphi_\alpha \,,\, A_1 \varphi_\alpha \rangle \tag{3.18}$$

where the probability vector $\mu = (\mu_\alpha)$ varies in a closed convex subset of the classical state space of probability vectors of length d. I.e., the conditional state space of a separable state admits a classical model.

Werner states

A 2×2 Werner state ρ_W is a state on $\mathcal{M}_2 \otimes \mathcal{M}_2$ which is a mixture between the completely mixed state $\mathbb{1}_4/4$ and the maximally entangled state $|\Psi^-\rangle\langle\Psi^-|$, where $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|01\rangle - |10\rangle)$:

$$\rho_W = (1-\lambda)\frac{\mathbb{1}}{4} + \lambda |\Psi^-\rangle\langle\Psi^-|$$

Depending on the mixing parameter λ the state ρ_W is entangled or separable [67]. For $0 \leq \lambda \leq \frac{1}{3}$ it is separable and for $\frac{1}{3} < \lambda \leq 1$. We can now look what influence this parameter has on the geometry of the conditional state space.

For the Werner state the restriction to one party $\mathrm{Tr}_1 \rho_W$ is the maximally mixed state $\mathbb{1}/2$. For the conditioning operator B, the requirement of unitality is then

$$\mathrm{Tr}\, \rho_W \mathbb{1} \otimes B = \mathrm{Tr}_2(\mathrm{Tr}_1(\rho_W) B) = \mathrm{Tr}\, \frac{B}{2} = 1\,,$$

so $B = 2\rho\prime$ for some qubit state $\rho\prime$. The state conditioned on B then becomes

$$\omega_B(A) = \mathrm{Tr}\, \rho_W A \otimes B = (1-\lambda)\,\mathrm{Tr}\, \frac{\mathbb{1}}{2} A + \lambda \langle \Psi^- | A \otimes 2\rho\prime | \Psi^- \rangle$$

If $\rho\prime$ has a Bloch vector $\mathbf{r}$ then the second term on the right hand side is a state with Bloch vector $-\mathbf{r}$. Hence the possible conditional states are mixes of all qubit states with $\mathbb{1}/2$ with a parameter λ. The conditional state space is then a Bloch ball with radius λ.

For qubits it is known that a separable state can always be decomposed as a mix of at most 4 separable pure states [57]. Note that this is lower than the 16 given by Caratheodory's theorem. This means that the conditional state space will lie inside a polygon with 4 vertices, i.e. a tetrahedron. Therefore, if the conditional state space cannot be put inside of a tetrahedron, the bipartite state must be entangled. This corresponds to what we see for the Werner state. The largest ball centered on the origin that can fit inside of a tetrahedron has radius 1/3. Hence, if the radius of the conditional state space, equal to the mixing parameter λ, becomes larger than 1/3, the Werner state must be entangled.

Bipartite quantum systems

It is possible to work out a description of the conditional state space for general bipartite quantum states. The description that is derived here consist of a quantum state space which is transformed by a completely positive map. This model fits in the general setting of generalized subsystems of [3]. Generalized subsystems describe quantum systems where only a subalgebra of the observables is accessible to the observer. In this case the observer only has access to one part of a bipartite state.

Consider a bipartite system with fully quantum parties, i.e., $\mathcal{A}_i = \mathcal{M}_i$ where $\mathcal{M}_i$ is a full matrix algebra of dimension d_i. The general finite dimensional situation can be handled by decomposition in a direct sum of full matrix algebras. A state of the composite system is given by a density matrix ρ_{12} of dimension $d_1 d_2$

$$\langle A_{12} \rangle_{12} = \mathrm{Tr}\big(\rho_{12} A_{12}\big), \;\; A_{12} \in \mathcal{M}_1 \otimes \mathcal{M}_2. \tag{3.19}$$

Let d_3 be the dimension of the range of ρ, then the GNS-construction [41] yields an essentially unique (up to unitary equivalence) normalized vector $\Omega \in \mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \otimes \mathbb{C}^{d_3}$ such that

$$\langle A_{12} \rangle_{12} = \langle \Omega \,,\, A_{12} \otimes \mathbb{1}_3 \, \Omega \rangle, \;\; A_{12} \in \mathcal{M}_1 \otimes \mathcal{M}_2. \tag{3.20}$$

We now perform the Schmidt decomposition of Ω with respect to $\mathcal{M}_2$ and $\mathcal{M}_1 \otimes \mathcal{M}_3$

$$\Omega = \sum_{j=1}^{m} c_j \, \Omega_{2\,j} \otimes \Omega_{13\,j}. \tag{3.21}$$

Here $c_j > 0$ and $\{\Omega_{2\,j}\}$ and $\{\Omega_{13\,j}\}$ are orthonormal families in $\mathbb{C}^{d_2}$ and $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_3}$. Thus, the conditional states on $\mathcal{M}_1$ are convex combinations of conditional states

defined by a rank one operator in $\mathcal{M}_2$. These are of the form

$$\begin{aligned} A_1 &\mapsto \langle \Omega \,,\, A_1 \otimes |\eta\rangle\langle\eta| \otimes \mathbb{1}_3\, \Omega\rangle \\ &= \sum_{k,\ell=1}^{m} \langle \Omega_{2\,k} \,,\, \eta\rangle \, \langle \eta \,,\, \Omega_{2\,\ell}\rangle \, \langle \Omega_{13\,k} \,,\, A_1 \otimes \mathbb{1}_3\, \Omega_{13\,\ell}\rangle \\ &= \langle \xi \,,\, A_1 \otimes \mathbb{1}_3\, \xi\rangle. \end{aligned}$$

Here ξ is a normalized vector in $\operatorname{span}\big(\{\Omega_{13\,j}\}\big)$. Moreover, any normalized ξ can be reached by an appropriate choice of η. Therefore, the conditional states are of the form

$$A_1 \mapsto \langle \xi \,,\, A_1 \otimes \mathbb{1}_3\, \xi\rangle, \;\; \xi \in \operatorname{span}\big(\{\Omega_{13\,j}\}\big).$$

Picking an isometry V from $\operatorname{span}\big(\{\Omega_{13\,j}\}\big)$ to $\mathbb{C}^m$ we obtain the following model for the conditional states

$$\mathcal{S}_1 = \big\{ A_1 \mapsto \operatorname{Tr}\big(\rho\, \Gamma(A_1)\big) \mid \rho \text{ density matrix on } \mathbb{C}^m \big\} \tag{3.22}$$

with

$$\Gamma(A_1) = V\, A_1 \otimes \mathbb{1}_3\, V^*. \tag{3.23}$$

The map Γ is completely positive and identity preserving. The model is manifestly positive, i.e. the complete positivity of the map and the positivity of the density matrix guarantee that the conditional state is positive. In Section 3.3.1 we described a specific conditional state by picking an operator in the inaccessible party, but this operator could not be chosen freely due to the positivity requirement. This difficulty is lifted in this model because of this manifest positivity.

Pure bipartite states

For a pure state $\langle\ \rangle_{12}$ on $\mathcal{M}_1 \otimes \mathcal{M}_2$ defined by a normalized vector Ω_{12} the forms (3.16) and (3.22) are actually very similar. We can identify the dual of $\mathcal{M}$ with $\mathcal{M}$ and use the pairing

$$\varphi \cdot A = \operatorname{Tr}(\varphi A). \tag{3.24}$$

Writing the Schmidt decomposition

$$\Omega_{12} = \sum_{i=1}^{p} r_i^{\frac{1}{2}}\, e_i \otimes f_i, \;\; r_i > 0, \tag{3.25}$$

we easily compute for $A_1 \in \mathcal{M}_1$

$$S^{\mathsf{T}}(A_1) = \sum_{k,\ell=1}^{p} r_k^{\frac{1}{2}}\, r_\ell^{\frac{1}{2}}\, \langle e_k\,,\, A_1 e_\ell\rangle\, |f_\ell\rangle\langle f_k|. \tag{3.26}$$

It is now not hard to verify that

$$\mathcal{S}_2 = \big\{ S^{\mathsf{T}}(A_1) \;\big|\; A_1 \geq 0 \;\text{ and }\; \langle A_1 \otimes \mathbb{1}_2\rangle_{12} = 1 \big\} \tag{3.27}$$

is affinely isomorphic to the state space of the p-dimensional complex matrices $\mathcal{M}_p$. We see therefore that the correlation dimension n is p^2. A general conditional state on $\mathcal{A}_1$

$$A_1 \mapsto \langle A_1 \otimes A_2\rangle_{12},\;\; A_2 \geq 0 \;\text{ and }\; \langle \mathbb{1}_1 \otimes A_2\rangle_{12} = 1 \tag{3.28}$$

can then be written as

$$A_1 \mapsto \mathrm{Tr}\big(S^{\mathsf{T}}(A_1)B\big) \tag{3.29}$$

for a suitable $B \in \big(S^{\mathsf{T}}(\mathcal{A}_1)\big)^*$. As $S^{\mathsf{T}}(\mathcal{A}_1)$ is the full state space of the n-dimensional matrices we must have that

$$B \geq 0 \;\text{ and }\; \mathrm{Tr}\big(S^{\mathsf{T}}(\mathbb{1}_1)B\big) = 1. \tag{3.30}$$

This means that (3.29) is manifestly positive. To obtain the equivalence with the form (3.22) we use the transposition with respect to the basis $\{f_j\}$ of $\mathbb{C}^p$:

$$\begin{aligned} \mathrm{Tr}\big(S^{\mathsf{T}}(A_1)B\big) \;&= \mathrm{Tr}\big(B^{\mathsf{T}}\big(S^{\mathsf{T}}(A_1)\big)^{\mathsf{T}}\big) = \mathrm{Tr}\big(B^{\mathsf{T}}\Lambda(A_1)\big) \\ &= \mathrm{Tr}\Big(\Lambda(\mathbb{1}_1)^{\frac{1}{2}} B^{\mathsf{T}} \Lambda(\mathbb{1}_1)^{\frac{1}{2}} \Gamma(A_1)\Big) \end{aligned} \tag{3.31}$$

with

$$\begin{aligned} &\Lambda(A_1) = \textstyle\sum_{k,\ell=1}^{p} r_k^{\frac{1}{2}} r_\ell^{\frac{1}{2}} \langle e_k\,,\, A_1 e_\ell\rangle\, |f_k\rangle\langle f_\ell| \;\text{ and} \\ &\Gamma(A_1) = \Lambda(\mathbb{1})^{-\frac{1}{2}} \Lambda(A_1) \Lambda(\mathbb{1}_1)^{-\frac{1}{2}}. \end{aligned} \tag{3.32}$$

It is clear that $\Gamma(\mathbb{1}) = \mathbb{1}$, i.e. Γ is unity preserving. Since $\Gamma(A_1) = B^* A_1 B$ with $B = \Lambda(\mathbb{1})^{-\frac{1}{2}} \sum_{k=1}^{p} r_k^{\frac{1}{2}} |e_k\rangle\langle f_k|$ it is also completely positive.

3.4 Constructing processes

We have introduced the conditional state space in the previous section in order to build quantum processes that have Markovian properties, i.e. finite correlations.

Here we construct such classical and quantum processes using as initial data a unity preserving CP map $\Gamma : \mathcal{M}_d \to \mathcal{M}_d$ with invariant state ρ. In the classical case this reduces to a stochastic matrix T with invariant measure $\boldsymbol{\mu}$. The construction is based on finitely correlated states [22], also called matrix product states. These states are more general than the ones we have considered up until now and are well-defined in the thermodynamic limit by construction, unlike the Bethe Ansatz states.

Finally we will show that the states that are constructed in this way are indeed the states we are looking for, namely states with finite dimensional conditional state spaces.

The starting point is a unity preserving CP map

$$\Lambda : \mathcal{M}_d \otimes \mathcal{M}_d \to \mathcal{M}_d$$

that is compatible with the given Γ in the following sense

$$\Lambda(A \otimes \mathbb{1}) = \Lambda(\mathbb{1} \otimes A) = \Gamma(A), \quad A \in \mathcal{M}_d. \tag{3.33}$$

A process $\boldsymbol{\omega}$ is now generated by repeatedly contracting the local observables on the half-chain. Consider a sequence of unity preserving CP maps

$$\begin{aligned}
\Lambda^{(0)} =& \Lambda : \mathcal{M}_d \otimes \mathcal{M}_d \to \mathcal{M}_d \\
\Lambda^{(1)} =& \Lambda \circ (\Lambda \otimes \mathrm{id}) : \big(\mathcal{M}_d \otimes \mathcal{M}_d\big) \otimes \mathcal{M}_d \to \mathcal{M}_d \\
&\vdots \\
\Lambda^{(n)} =& \Lambda \circ \big(\Lambda^{(n-1)} \otimes \mathrm{id}\big) : \underbrace{\big(\mathcal{M}_d \otimes \cdots \otimes \mathcal{M}_d\big)}_{(n+1) \text{ times}} \otimes \mathcal{M}_d \to \mathcal{M}_d.
\end{aligned} \tag{3.34}$$

The expectation of a local observable $A_n \in \otimes_0^n \mathcal{M}_d$ is then computed as

$$\boldsymbol{\omega}\big(A_n\big) = \mathrm{Tr}\Big\{\rho\, \Lambda^{(n)}(\mathbb{1} \otimes A_n)\Big\}. \tag{3.35}$$

Instead of looking at $\Lambda^{(n)}$ as a map contracting observables, we can also regard the conjugate maps as constructing ever larger local density matrices with the help of an ancillary algebra. The local density are then obtained by tracing out the ancilla:

$$\rho_n = \mathrm{Tr}_1\, \Lambda^{(n)^*}(\rho)\,.$$

To define a stationary process, (3.35) must satisfy a number of requirements. The definition should be consistent in the first place, namely $\boldsymbol{\omega}(A_n \otimes \mathbb{1}) = \boldsymbol{\omega}(A_n)$. This follows from the compatibility (3.33) and the invariance of ρ

$$\begin{aligned}
\boldsymbol{\omega}(A_n \otimes \mathbb{1}) &= \mathrm{Tr}\Big\{\rho\,\Lambda^{(n+1)}(\mathbb{1} \otimes A_n \otimes \mathbb{1})\Big\} \\
&= \mathrm{Tr}\Big\{\rho\,\Big(\Lambda \circ \big(\Lambda^{(n)} \otimes \mathrm{id}\big)\Big)(\mathbb{1} \otimes A_n \otimes \mathbb{1})\Big\} \\
&= \mathrm{Tr}\Big\{\rho\,\Lambda\Big(\Lambda^{(n)}(\mathbb{1} \otimes A_n) \otimes \mathbb{1}\Big)\Big\} \\
&= \mathrm{Tr}\Big\{\rho\,\Gamma\Big(\Lambda^{(n)}(\mathbb{1} \otimes A_n)\Big)\Big\} \\
&= \mathrm{Tr}\Big\{\rho\,\Lambda^{(n)}(\mathbb{1} \otimes A_n)\Big\} \\
&= \boldsymbol{\omega}(A_n).
\end{aligned}$$

Next, we need positivity. This follows immediately from the complete positivity of Λ. The compatibility condition implies that Λ maps the identity on $\mathbb{C}^d \otimes \mathbb{C}^d$ to the identity on $\mathbb{C}^d$. This implies the normalization and stationarity of $\boldsymbol{\omega}$.

It is important to observe that compatibility (3.33) imposes a severe restriction on Γ. Not every CP transformation Γ of $\mathcal{M}_d$ admits a compatible extension. Moreover, the compatible extensions of Γ, whenever such extensions exist, form a compact and convex set and one may wonder about particular extensions.

The following theorem characterizes the finitely correlated states in terms of their conditional state space. Classical Markov processes have the property that the future evolution is only conditioned on the present state and not the past. This theorem shows that finitely correlated states have a similar property, namely that their dependence on the past is finite. For a proof of the theorem, see [22].

Theorem 2. *Let ω be a translationally invariant state on the chain algebra $\mathcal{M}_d^{\mathbb{Z}}$. ω is a finitely correlated state as in (3.35) if and only if its correlation dimension is finite:*

$$\dim \mathcal{V} < \infty$$

with

$$\mathcal{V} := \big\{A \in \mathcal{M}_d^{\mathbb{N}} \mapsto \omega(X \otimes A) \;\mid\; X \in \mathcal{M}_d^{\mathbb{Z}\setminus\mathbb{N}}\big\}$$

We shall now turn to some classes of examples.

A qubit example with SU(2)-symmetry cont.

Here we return to the example with SU(2)-symmetry we have explored before. We will see that by considering the finitely correlated states, we can get more elaborate correlations while still being able to easily describe the translation-invariant states explicitly.

In order to have manifest SU(2)-invariance of the process we impose SU(2)-covariance both on the CP transformation of $\mathcal{M}_2$ and on its compatible extensions from $\mathcal{M}_2 \otimes \mathcal{M}_2$ to $\mathcal{M}_2$. Let $\mathcal{G} \mapsto U_g$ be a unitary representation of a group $\mathcal{G}$ on a Hilbert space $\mathcal{H}$. The adjoint representation lifts it to a representation of $\mathcal{G}$ on the bounded linear transformations $\mathcal{B}(\mathcal{H})$ of $\mathcal{H}$:

$$\mathrm{Ad}(U_g)(A) = U_g A U_g^*, \;\; g \in \mathcal{G}, \;\; A \in \mathcal{B}(\mathcal{H}).$$

Given two unitary representations $U^{(1)}$ and $U^{(2)}$ of $\mathcal{G}$ on $\mathcal{H}_1$ and $\mathcal{H}_2$ a map $\Gamma : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ is covariant if

$$\Gamma \circ \mathrm{Ad}(U^{(1)}) = \mathrm{Ad}(U^{(2)}) \circ \Gamma \tag{3.36}$$

The Choi-Jamiołkowski encoding of a linear map $\Gamma : \mathcal{M}_{d_1} \to \mathcal{M}_{d_2}$ is very convenient for handling complete positivity

$$\mathsf{C}(\Gamma) = \sum_{i,j} |i\rangle\langle j| \otimes \Gamma(|i\rangle\langle j|),$$

Γ is completely positive if and only if $\mathsf{C}(\Gamma)$ is positive semi-definite. The encoding depends on the chosen basis through the matrix units $|i\rangle\langle j|$ but only up to unitary equivalence as

$$\mathsf{C}(\Gamma \circ \mathrm{Ad}(U)) = \mathrm{Ad}(U^{\mathsf{T}} \otimes \mathbb{1}) \circ \mathsf{C}(\Gamma) \;\text{ and}$$

$$\mathsf{C}(\mathrm{Ad}(U) \circ \Gamma) = \mathrm{Ad}(\mathbb{1} \otimes U) \circ \mathsf{C}(\Gamma).$$

The covariance condition (3.36) for $\Gamma : \mathcal{M}_{d_1} \to \mathcal{M}_{d_2}$ translates for its Choi-Jamiołkowski encoding into

$$\left[\overline{U}_g^{(1)} \otimes U_g^{(2)}, \, \mathsf{C}(\Gamma)\right] = 0, \;\; g \in \mathcal{G}.$$

Here $\overline{A}$ is the complex conjugate of the matrix A. For SU(2) there is an additional simplification because the conjugate of SU(2) is unitarily equivalent to SU(2).

It turns out that there is a one-parameter family of SU(2)-covariant unit preserving CP transformations of $\mathcal{M}_2$

$$\Gamma(\boldsymbol{\sigma}) = \mu\, \boldsymbol{\sigma}, \;\; -\tfrac{1}{3} \leq \mu \leq 1.$$

The SU(2)-covariant CPUP maps $\Lambda : \mathcal{M}_2 \otimes \mathcal{M}_2 \to \mathcal{M}_2$ compatible with Γ are parametrized by three real parameters

$$\begin{aligned} &\Lambda(\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2) = \alpha\, \mathbb{1}, \\ &\Lambda(\boldsymbol{\sigma}_1) = \Lambda(\boldsymbol{\sigma}_2) = \Gamma(\boldsymbol{\sigma}) = \mu\, \boldsymbol{\sigma}, \ \text{and} \\ &\Lambda(\boldsymbol{\sigma}_1 \times \boldsymbol{\sigma}_2) = \eta\, \boldsymbol{\sigma}. \end{aligned} \tag{3.37}$$

Complete positivity imposes constraints on α, μ, and η

$$|6\mu - \alpha| \leq 3 \ \text{and} \ 3 - 2\alpha - \alpha^2 + 12\mu - 12\alpha\mu - 9\eta^2 \geq 0. \tag{3.38}$$

These conditions can be obtained by imposing positivity on the Choi matrix of Λ. The allowed region is a piece of a cone in $\mathbb{R}^3$. We then compute the expectation of $\mathbf{p}$

$$\begin{aligned} \langle \mathbf{p} \rangle &= \tfrac{1}{4} - \tfrac{1}{4} \langle \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2 \rangle = \tfrac{1}{4} - \tfrac{1}{8} \operatorname{Tr} \sum_{\gamma} \Lambda\big(\sigma^\gamma \otimes \Lambda(\sigma^\gamma \otimes \mathbb{1})\big) \\ &= \tfrac{1}{4} - \tfrac{1}{8}\, \mu \operatorname{Tr} \Lambda(\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2) = \tfrac{1}{4}\,(1 - \alpha\mu). \end{aligned}$$

The maximum in the allowed parameter region is attained for $\alpha = -\frac{3}{2}$ and $\mu = \frac{1}{4}$ and is independent of η (all η with $|\eta| \leq \frac{\sqrt{5}}{2}$ fulfil positivity for these α and μ). Therefore

$$\langle \mathbf{p} \rangle \leq \tfrac{11}{32} \tag{3.39}$$

for $\langle\ \rangle$ a stationary and SU(2)-invariant process as in (3.34). This should be compared with (3.9).

In passing from exchangeable to shift-invariant separable states we actually allowed product states of period 2. This can also be applied to processes of the type (3.37). Considering $\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2$ as the contribution to the energy of two neighbouring spins, a minimal value of $\langle \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2 \rangle$ corresponds to a maximal value of $\langle \mathbf{p} \rangle$ and this is expected to happen for spins as anti-parallel as possible. Therefore the second requirement in (3.37) is inappropriate and we should consider general SU(2)-covariant maps $\Lambda : \mathcal{M}_1 \otimes \mathcal{M}_2 \to \mathcal{M}_2$. These are determined by four real parameters

$$\begin{aligned} &\Lambda(\boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2) = \alpha\, \mathbb{1}, \\ &\Lambda(\boldsymbol{\sigma}_1) = \mu\, \boldsymbol{\sigma}, \ \ \Lambda(\boldsymbol{\sigma}_2) = \nu\, \boldsymbol{\sigma}, \ \text{and} \\ &\Lambda(\boldsymbol{\sigma}_1 \times \boldsymbol{\sigma}_2) = \eta\, \boldsymbol{\sigma}. \end{aligned} \tag{3.40}$$

Complete positivity imposes the constraints

$$\begin{aligned} &|3\mu + 3\nu - \alpha| \leq 3 \ \text{and} \\ &3 - 2\alpha - \alpha^2 + 6(1 - \alpha)(\mu + \nu) - 9(\mu - \nu)^2 - 9\eta^2 \geq 0. \end{aligned} \tag{3.41}$$

We now introduce two SU(2)-covariant maps $\Lambda_i : \mathcal{M}_2 \otimes \mathcal{M}_2 \to \mathcal{M}_2$ as in (3.40) with defining parameters $(\alpha_i, \mu_i, \nu_i, \eta_i)$. The expectation of $\mathbf{p}$ in the equal weight average of these period-2 processes is given by

$$\langle \mathbf{p} \rangle = \tfrac{1}{4} - \tfrac{1}{8}\,(\alpha_2\mu_1 + \alpha_1\nu_2).$$

Maximizing this in the allowed parameter region yields

$$\langle \mathbf{p} \rangle = \tfrac{5}{8} = 0.625$$

which is within 10% of the optimal bound and well within the entangled shift-invariant states.

Classical hidden Markov processes

There is good reason to consider the finitely correlated states as quantum Markov states. Indeed, when restricting the observable algebras to be Abelian, we recover the classical hidden Markov processes.

A classical observable, i.e. a $\mathbb{R}$-valued function f on configuration space $\Omega = \{1, 2, \ldots, d\}$ is naturally tabulated into a vector $\mathbf{f} = \big(f(1), f(2), \ldots, f(d)\big)^{\mathsf{T}} \in \mathbb{R}^d$ and identified with a diagonal matrix in $\mathcal{M}_d$ through the map

$$\mathrm{dia}(\mathbf{f}) = \sum_{\epsilon} f(\epsilon)\, |\epsilon\rangle\langle\epsilon|.$$

The relation between a (completely) positive transformation Γ of $\mathcal{M}_d$ and a stochastic $d \times d$ matrix is then

$$\Gamma\big(\mathrm{dia}(\mathbf{f})\big) = \mathrm{dia}(T\,\mathbf{f}).$$

This allows to rewrite the compatibility equation (3.33): a stochastic matrix $S : \mathbb{R}^d \otimes \mathbb{R}^d \to \mathbb{R}^d$ is compatible with a stochastic matrix $T : \mathbb{R}^d \to \mathbb{R}^d$ if

$$\sum_{\epsilon_2} S_{\varphi,(\epsilon,\epsilon_2)} = \sum_{\epsilon_1} S_{\varphi,(\epsilon_1,\epsilon)} = T_{\varphi,\epsilon}, \;\; \forall\, \varphi, \epsilon.$$

Let us introduce d matrices of size $d \times d$ with non-negative entries

$$E(\epsilon)_{\varphi,\eta} = S_{\varphi,(\eta,\epsilon)}.$$

The process generated by S is then seen to be

$$\boldsymbol{\omega}(\epsilon_0, \epsilon_1, \ldots, \epsilon_n) = \langle \boldsymbol{\mu}\,,\, E(\epsilon_0)E(\epsilon_1)\cdots E(\epsilon_n)\mathbf{1}\rangle,$$

where $\mathbf{1} \in \mathbb{R}^d$ has all its entries equal to one and $\boldsymbol{\mu}$ is the invariant probability vector for T.

A stochastic matrix T always allows the extension

$$S_{\varphi,(\eta,\epsilon)} = \delta_{\eta,\epsilon}\, T_{\varphi,\epsilon}.$$

The corresponding process is the usual Markov process (3.5). More general extensions $\boldsymbol{\omega}$ are hidden Markov processes: there exists a larger configuration space Ω_1, a function $F : \Omega_1 \to \Omega$ and a Markov process $\boldsymbol{\omega}_1$ on Ω_1 such that

$$\boldsymbol{\omega}(\epsilon_0, \epsilon_1, \ldots, \epsilon_n) = \sum_{F(\varphi_j)=\epsilon_j} \boldsymbol{\omega}_1(\varphi_0, \varphi_1, \ldots, \varphi_n).$$

The entropy of hidden Markov processes can be computed using the method due to Blackwell [21] that we have explained in the previous chapter. The starting point is the asymptotic entropy production formula (2.10). The construction of the process, adding one point at a time, see (3.34) and (3.35), defines a dynamical system on the length-d probability vectors. The entropy of the process is then obtained as an average over entropies of probability vectors with respect to the invariant measure of the dynamical system. Numerical evidence suggests that the Markov extension has the smallest entropy amongst all.

3.5 Conclusion

In this chapter we have looked for a generalization of classical Markov processes. As a first step, we have looked at some of the ways to construct a classical Markov process, for example through a transfer matrix or an extension of overlapping states. We have also seen that due to the intricacies of quantum correlations, these constructions cannot be straightforwardly transported to quantum systems.

A construction that can be carried out for quantum systems is that of conditional state spaces. For classical systems these conditional state spaces correspond to regular conditional states. Through a series of applications, we have also looked deeper into what can be said about conditional state spaces of quantum systems. The geometry and dimension of the conditional state space tells us about the nature and strength of the correlations in the system.

The final goal is then to construct quantum Markov processes. As was demonstrated in [22], an appealing way to construct such processes is the set-up of finitely correlated states. These are states where the conditional state space is finite-dimensional, hence the name finitely correlated. An interesting property of this construction is that it exactly coincides with the set of hidden Markov processes when restricted to Abelian algebras, as we have demonstrated.

The general concept of conditional state spaces has been studied in some detail for certain classes of states. These hint at possibilities to study quantum correlations

using this tool. It may also prove interesting to study these conditional state spaces in the setting of quantum phase transitions and DLR-like equations [23].

Conditional state spaces and finitely correlated states will be studied for a particular class of physical systems, the fermionic quasi-free states, in the following chapters.

Chapter 4

Free Fermions

4.1 Introduction

Quantum states are mostly indirectly given, typically as ground or equilibrium states for a given interaction and are hence difficult to work with as there is for example no explicit density matrix. Also, in general one has to deal with an enormous amount of parameters when the number of particles grows. As the number of components grows, typically the number of parameters grows exponentially. Free Fermionic states [60, 53, 5] are exceptional in both respects. These states describe systems of non-interacting fermions. They are given by an explicit recipe, reducing the calculation of higher order correlation to a simple combinatorial combination of second order correlations. Hence not only can they be calculated explicitly, they are also fully described by their second order correlation, resulting in a significant reduction in parameters.

For these reasons free states form an interesting class of quantum states. Constructions that are intractable in general quantum systems due to the high dimensionality can often be carried out for free fermions.

Although they allow for a simple description, the free Fermionic states nevertheless contain enough complexity to describe interesting quantum behaviour. They arise for example in the description of the free electron gas in quantum statistical mechanics [10] and in Hartree-Fock approximations in solid state physics [45].

This chapter gives an overview of some of the properties of free states, most of which are known in the literature. A new contribution is a lemma for rewriting certain expectation values.

4.2 Fermions

Assume we have two indistinguishable particles described by pure state vectors $|\varphi_1\rangle$ and $|\varphi_2\rangle$ in a Hilbert space $\mathcal{H}$. The state $|\varphi_1\rangle \otimes |\varphi_2\rangle$ is a possible description of the combined particles. However, due to the particles being indistinguishable, $|\varphi_2\rangle \otimes |\varphi_1\rangle$ is also a perfectly reasonable description. The possibilities for the state vector of a system of two indistinguishable particles are thus

$$|\phi_\pm\rangle = |\varphi_1\rangle \otimes |\varphi_2\rangle \pm |\varphi_2\rangle \otimes |\varphi_1\rangle \,.$$

These states remain the same or pick up a minus sign upon the interchange of particles. Let's denote by $U_{(12)}$ the operation that interchanges the two particles:

$$U_{(12)}|\psi_1\rangle \otimes |\psi_2\rangle = |\psi_2\rangle \otimes |\psi_1\rangle \,.$$

Then

$$U_{(12)}|\phi_\pm\rangle = \pm|\phi_\pm\rangle \,.$$

The particles that are symmetric under particle interchange are called bosons. The ones that are anti-symmetric are called fermions. We will only be concerned about the latter here.

If there are more than two particles, we need to consider more permutations than just the interchange of two particles. Denote by π a permutation of n elements. This permutation works on a number of particles as

$$U_\pi \left(|\psi_1\rangle \otimes \ldots \otimes |\psi_n\rangle\right) = |\psi_{\pi(1)}\rangle \otimes \ldots \otimes |\psi_{\pi(n)}\rangle \,.$$

Any permutation π can be written as a number of interchanges and when this number is even, its sign $\epsilon(\pi)$ is 1, otherwise it is -1. As every interchange will make the fermion state vector pick up a factor (-1), the sign change under the permutation will be $(-1)^{\epsilon(\pi)}$:

$$U_\pi|\psi\rangle = (-1)^{\epsilon(\pi)}|\psi\rangle \,.$$

One can see that all possible combinations of n indistinguishable particles $\psi_1, \ldots, \psi_n$ that obey these rules are given by

$$\psi_1 \wedge \psi_2 \wedge \ldots \wedge \psi_n = \frac{1}{n!} \sum_\sigma \epsilon(\sigma) \varphi_{\sigma(1)} \otimes \psi_{\sigma(2)} \otimes \ldots \otimes \psi_{\sigma(n)} \,,$$

where σ runs over all permutations of n elements. The space spanned by these anti-symmetric vectors is denoted by $\mathcal{H}^{\wedge n} = \mathcal{H} \wedge \ldots \wedge \mathcal{H}$.

When the number of particles is not fixed as we have assumed until now, we get what is called the Fermionic Fock space:

$$\mathcal{F}(\mathcal{H}) = \mathbb{C} \oplus \mathcal{H} \oplus (\mathcal{H} \wedge \mathcal{H}) \oplus \ldots = \bigoplus_{n=0}^{\infty} \mathcal{H}^{\wedge n} \,.$$

We can go from a system of n fermions to one with $n+1$ fermions by the creation operator

$$a^*(\varphi)(\psi_1 \wedge \ldots \wedge \psi_n) = \varphi \wedge \psi_1 \wedge \ldots \wedge \psi_n \,,$$

which inserts a particle in the state φ. The conjugate operator removes a particle and is called the annihilation operator

$$a(\varphi)(\psi_1 \wedge \ldots \wedge \psi_n) = \sum_i (-1)^i \langle \varphi | \psi_i \rangle (\psi_1 \wedge \ldots \wedge \psi_{i-1} \wedge \psi_{i+1} \wedge \ldots \wedge \psi_n)$$

These operators satisfy the following anti-commutation relations:

$$\{a(\varphi)\,,\, a(\psi)\} = 0 \;\text{ and }\; \{a(\varphi)\,,\, a^*(\psi)\} = \langle \varphi\,,\, \psi \rangle \, \mathbb{1}. \tag{4.1}$$

In fact, every linear transformation of the Fermionic Fock space $\mathcal{F}(\mathcal{H})$ can be written as a linear combination of products of a and a^*. This is why we use the algebra generated by the creation and annihilation operators to describe observables of Fermionic systems.

Instead of looking at a specific physical realization of the algebra of creation and annihilation operators (in terms of the Fermionic Fock space), we can look at the abstract algebra describing the behaviour of these operators. This algebra is called the CAR-algebra $\mathcal{A}(\mathcal{H})$ — CAR stands for canonical anti-commutation relations — with single-mode Hilbert space $\mathcal{H}$. It is the C*-algebra generated by an identity $\mathbb{1}$ and by creation and annihilation operators a^* and a that satisfy

$$\begin{gathered} \varphi \in \mathcal{H} \mapsto a^*(\varphi) \text{ is complex linear} \\ \{a(\varphi)\,,\, a(\psi)\} = 0 \;\text{ and }\; \{a(\varphi)\,,\, a^*(\psi)\} = \langle \varphi\,,\, \psi \rangle \, \mathbb{1}. \end{gathered} \tag{4.2}$$

The physical predictions that one makes for Fermionic systems described by the CAR-algebra should remain the same when we introduce an unimportant phase factor. This leads us to the gauge-invariant CAR algebra, which is a subalgebra of the CAR algebra. There is a representation from U(1) in the group $\{\alpha_z \mid z \in \mathrm{U}(1)\}$ of gauge automorphisms of $\mathcal{A}(\mathcal{H})$

$$z \in \mathrm{U}(1) \mapsto \alpha_z \;\text{ with }\; \alpha_z(a^*(\varphi)) = z a^*(\varphi). \tag{4.3}$$

It's fixed point algebra is the GICAR-algebra — gauge-invariant CAR —, it is generated as a linear space by monomials in creation and annihilation operators of the form $a^*(\varphi_1) \cdots a^*(\varphi_n) a(\psi_n) \cdots a(\psi_1)$. The states we will look at are defined on this GICAR algebra.

4.3 Free fermions

A gauge-invariant free state ω_Q on $\mathcal{A}(\mathcal{H})$ is determined by a symbol which is a linear operator Q on $\mathcal{H}$ satisfying $0 \leq Q \leq \mathbb{1}$. The ω_Q-expectations of all monomials vanish except for the elements of the GICAR-algebra:

$$\omega_Q\big(a^*(\varphi_1)a^*(\varphi_2)\cdots a^*(\varphi_n)a(\psi_n)\cdots a(\psi_2)a(\psi_1)\big) = \det\Big(\big[\langle\psi_k\,,\,Q\,\varphi_\ell\rangle\big]\Big). \quad (4.4)$$

Positivity holds if and only if $0 \leq Q \leq \mathbb{1}$. The set of symbols

$$\mathcal{Q}(\mathcal{H}) = \{Q \mid Q \text{ bounded linear operator on } \mathcal{H} \text{ such that } 0 \leq Q \leq \mathbb{1}\} \quad (4.5)$$

is convex and weakly compact. Convexity at the level of symbols is very different from convexity at the level of the free states. Nevertheless it can be shown that a free state is pure, i.e. extreme in the full state space of $\mathcal{A}(\mathcal{H})$, if and only if its symbol is an orthogonal projector, i.e. an extreme point of $\mathcal{Q}(\mathcal{H})$ [2].

Important quantities like the entropy of free states are expressible in terms of symbols, e.g.

$$\mathsf{S}(Q) = -\operatorname{Tr} Q \log Q - \operatorname{Tr}(\mathbb{1}-Q)\log(\mathbb{1}-Q). \quad (4.6)$$

This formula is an example of the reduction of complexity in free fermionic states. Although the ω_Q is a state on the CAR algebra $\mathcal{A}(\mathcal{H})$, its properties are defined by an operator on the single particle Hilbert space $\mathcal{H}$.

A state ω on $\mathcal{A}(\mathcal{H})$ is even if it vanishes on all monomials in creation and annihilation operators with an odd number of factors. Gauge-invariant states are automatically even. If ω_i is an even state on $\mathcal{A}(\mathcal{H}_i)$ for $i = 1, 2$, then there exists a unique state $\omega_1 \wedge \omega_2$ on $\mathcal{A}(\mathcal{H}_1 \oplus \mathcal{H}_2)$ such that

$$(\omega_1 \wedge \omega_2)(X_1X_2) = \omega_1(X_1)\,\omega_2(X_2), \;\; X_i \in \mathcal{A}(\mathcal{H}_i). \quad (4.7)$$

For a proof, see [2].

A symbol Q induces an orthogonal decomposition

$$\mathcal{H} = \mathcal{H}_0 \oplus \tilde{\mathcal{H}} \oplus \mathcal{H}_{\mathbb{1}} \quad (4.8)$$

where

$$\mathcal{H}_0 = \ker(Q) \;\text{ and }\; \mathcal{H}_{\mathbb{1}} = \ker(\mathbb{1}-Q) \quad (4.9)$$

and ω_Q factorizes into

$$\omega_Q = \omega_0 \wedge \omega_{\tilde{Q}} \wedge \omega_{\mathbb{1}} \;\text{ with }\; \tilde{Q} = Q\big|_{\tilde{\mathcal{H}}}. \quad (4.10)$$

The states ω_0 on $\mathcal{A}(\mathcal{H}_0)$ and $\omega_{\mathbb{1}}$ on $\mathcal{A}(\mathcal{H}_{\mathbb{1}})$ are pure, they are Fock and anti-Fock states. The Fock state ω_0 is the pure state $|\Omega\rangle$ which describes the vacuum:

$$a(\varphi)|\Omega\rangle = 0\,.$$

The anti-Fock state or Dirac sea is the state with all modes occupied:

$$\omega(a(\varphi)a^*(\varphi)) = 0\,.$$

Equilibrium states

To prove our first proposition we rely on the equilibrium properties of free states. The free states can be described as equilibrium states for a suitably chosen Hamiltonian.

For finite dimensional systems, equilibrium states are Gibbs states:

$$\omega_\beta(x) = \mathrm{Tr}(x\rho_\beta)\,,$$

where

$$\rho_\beta = \frac{e^{-\beta H}}{\mathcal{Z}}$$

and $\mathcal{Z}$ is a normalizing factor:

$$\mathcal{Z} = \mathrm{Tr}\, e^{-\beta H}\,.$$

By looking at the time evolution of the density matrix (see Eq. 1.9), it is clear that this state is invariant. Denoting the time evolution of observables by α_t, we have

$$\alpha_t(x) = e^{\imath t H} x e^{-\imath t H}\,,$$

and

$$\omega_\beta(\alpha_t(x)) = \omega_\beta(x)\,.$$

Using the structure of the Gibbs state ρ_β we can derive a property of the expectation values called the KMS-condition. For Gibbs states, this condition amounts to

$$\begin{aligned}
\omega_\beta(\alpha_t(x)y) &= \frac{1}{\mathcal{Z}}\,\mathrm{Tr}\, e^{-\beta H} e^{\imath t H} x e^{-\imath t H} y \\
&= \frac{1}{\mathcal{Z}}\,\mathrm{Tr}\, e^{-\beta H} y e^{-\beta H} e^{\imath t H} x e^{-\imath t H} e^{\beta H} \\
&= \omega_\beta(y e^{\imath(t+\imath\beta)H} x e^{-\imath(t+\imath\beta)H}) \\
&= \omega_\beta(y\alpha_{t+\imath\beta}(x))
\end{aligned}$$

Similarly, one can characterize an equilibrium state ω_β on a C*-algebra $\mathcal{A}$ by the KMS-condition corresponding to a dynamics in Heisenberg picture. Let $\{\alpha_t \mid t \in \mathbb{R}\}$ be a continuous group of automorphisms of $\mathcal{A}$, then ω_β is an α-KMS-state at inverse temperature $\beta > 0$ if there exists for any pair of observables $x, y \in \mathcal{A}$ a function

$$z \in \mathbb{C} \mapsto F_{x,y}(z) \in \mathbb{C} \tag{4.11}$$

that is analytic inside the strip $0 < \Im\mathfrak{m} z < \beta$, that extends continuously to the closure of the strip, and such that

$$F_{x,y}(t) = \omega_\beta(\alpha_t(x)y) \text{ and } F_{x,y}(t - i\beta) = \omega_\beta(y\alpha_t(x)), \ t \in \mathbb{R}. \tag{4.12}$$

It is straightforward to check that the KMS-states on a finite dimensional quantum system precisely coincide with the canonical Gibbs states.

Let $0 < Q < \mathbb{1}$ whereby we mean that for $0 \neq \varphi$

$$0 < \langle \varphi \, , \, Q\varphi \rangle \qquad \text{and} \qquad 0 < \langle \varphi \, , \, (\mathbb{1} - Q)\varphi \rangle. \tag{4.13}$$

The free state ω_Q is then the unique α KMS state on $\mathcal{A}(\mathcal{H})$ at inverse temperature $\beta = 1$ where α is the strongly continuous one-parameter group of automorphisms [54]

$$\alpha_t\big(a^*(\varphi)\big) = a^*\big(\mathrm{e}^{ith}\varphi\big), \ t \in \mathbb{R} \tag{4.14}$$

with

$$h = \ln(\mathbb{1} - Q) - \ln Q. \tag{4.15}$$

Exponential elements

An interesting approach when working with free states is the use of so-called exponential elements. These exponential elements provide us with a way to write down the density matrix of a free Fermionic state ω_Q in terms of its defining symbol Q. Studying the spectrum of such operators then for example allows to determine the entropy of a free Fermionic state; see [18, 19] for more details. Due to this form of the density matrix, a free Fermionic state can also easily be characterized by its action on these exponential elements, a property we will make use of later on.

The second quantization map

$$\Gamma : \mathcal{T}_1(\mathcal{H}) \to \mathcal{A}(\mathcal{H}) : \Gamma(A) := \sum_{k,\ell} \langle e_k \, , \, A\, e_\ell \rangle \, a^*(e_k)a(e_\ell) \tag{4.16}$$

takes a trace class single-particle operator $A \in \mathcal{T}_1(\mathcal{H})$ to an element $\Gamma(A)$ in the CAR-algebra $\mathcal{A}(\mathcal{H})$. This operator is independent of the chosen orthonormal basis

$\{e_i\}$ of $\mathcal{H}$. It acts on vectors in the Fermionic Fock space as

$$\Gamma(A)\varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_n = \sum_{j=1}^{n} \varphi_1 \wedge \varphi_2 \wedge \ldots \wedge A\varphi_j \wedge \ldots \wedge \varphi_n \,.$$

This map is complex-linear, continuous, and satisfies

$$\frac{1}{2}\,\|A\|_1 \leq \|\Gamma(A)\| \leq \|A\|_1 \,. \tag{4.17}$$

Here $\|X\|$ denotes the operator norm of X and $\|X\|_1 = \operatorname{Tr}\sqrt{X^*X}$ is the trace norm. This map is, moreover, completely positive and for a positive operator $A \in \mathcal{T}_1(\mathcal{H})$

$$\|\Gamma(A)\| = \operatorname{Tr} A. \tag{4.18}$$

In [19] a map E from the Fredholm operators $\mathbb{1} + \mathcal{T}_1(\mathcal{H})$ to $\mathcal{A}(\mathcal{H})$ is considered that satisfies

$$\begin{aligned} \mathrm{E}(X)\mathrm{E}(Y) &= \mathrm{E}(XY), \\ \mathrm{E}(X)^* &= \mathrm{E}(X^*), \\ \operatorname{Tr}\mathrm{E}(X) &= \det(\mathbb{1} + X), \ \text{ and} \\ \mathrm{E}(\exp A) &= \exp\bigl(\Gamma(A)\bigr), \ \ A \in \mathcal{T}_1(\mathcal{H}). \end{aligned} \tag{4.19}$$

On the fermionic Fock space, it is given by

$$\mathrm{E}(A)\varphi_1 \wedge \varphi_2 \wedge \ldots \wedge \varphi_n = A\varphi_1 \wedge A\varphi_2 \wedge \ldots \wedge A\varphi_n$$

The operators of the form $\mathrm{E}(X)$ are called exponential elements. The span of the exponential elements is in fact the entire GICAR algebra, so one can characterize gauge-invariant states by their action on the exponential elements. In the case of the free Fermionic states, this greatly simplifies calculations since the density matrices are given by exponential elements as well.

The map E obeys for positive trace-class A the bounds

$$\begin{aligned} 1 + \|A\|_1 \leq \|\mathrm{E}(\mathbb{1} + A)\| &\leq \exp\bigl(\|A\|_1\bigr) \ \text{ and} \\ \|\mathrm{E}(\mathbb{1} + A) - \mathbb{1}\| &\leq \exp\bigl(\|A\|_1\bigr) - 1. \end{aligned} \tag{4.20}$$

One can check that the density matrix ρ_Q of a free Fermionic state ω_Q with symbol Q is given by

$$\rho_Q = \det(\mathbb{1} - Q)E(\frac{Q}{\mathbb{1} - Q}) \,.$$

A gauge-invariant free state ω_Q can then be characterized by

$$\omega_Q(\mathrm{E}(X)) = \mathrm{Tr}\,\rho_Q E(X) = \det(\mathbb{1} - Q + QX), \;\; X \in \mathbb{1} + \mathcal{T}_1(\mathcal{H}). \tag{4.21}$$

Rearranging expectation values

Free states are the fermionic version of classical Gaussians. For Gaussians, expectations of a given function f multiplied by a Gaussian variable can be expressed as expectations of the derivative of the function with respect to the random variable. If $\langle\rangle_\sigma$ are the expectation values under a Gaussian measure then

$$\langle x f(x)\rangle_\sigma = \langle x^2\rangle_\sigma \langle \frac{\partial f}{\partial x}\rangle_\sigma$$

and

$$\langle x^2 f(x)\rangle_\sigma = \langle x^2\rangle_\sigma^2 \langle \frac{\partial^2 f}{\partial x^2}\rangle_\sigma + \langle x^2\rangle_\sigma \langle f\rangle_\sigma \,.$$

The following lemma provides such a formula in the fermionic context. Note that the classical second derivative becomes a combined commutation anti-commutation.

Lemma 1. *For any $Y \in \mathcal{A}(\mathcal{H})$ and $\varphi \in \mathcal{H}$, we have*

$$\omega_Q(a^*(\varphi) Y a(\varphi)) = \omega_Q(a^*(\varphi)a(\varphi))\omega_Q(Y) + \omega_Q\big(\big\{a(Q\varphi)\,,\,\big[a^*(\varphi)\,,\,Y\big]\big\}\big) \tag{4.22}$$

Proof. We may limit ourselves to gauge-invariant Y due to the gauge-invariance of the free state. Since we can approximate Y by linear combinations of gauge-invariant monomials in $\mathcal{A}(\mathcal{H})$, it suffices to show the lemma for $Y = a^*(\psi_1)\cdots a^*(\psi_n)a(\eta_n)\cdots a(\eta_1)$. For such Y, using the fact that ω_Q is free, the expression $\omega_Q(a^*(\varphi)Ya(\varphi))$ expands to

$$\begin{aligned}
&\omega_Q\big(a^*(\varphi)a(\varphi)\big)\,\omega_Q\big(a^*(\psi_1)\cdots a^*(\psi_n)a(\eta_n)\cdots a(\eta_1)\big) \\
&\quad + \sum_{k,\ell} \varepsilon_{k,\ell}\,\omega_Q\big(a^*(\varphi)a(\eta_\ell)\big)\,\omega_Q\big(a^*(\psi_k)a(\varphi)\big) \\
&\qquad \times \omega_Q\big(a^*(\psi_1)\cdots \widehat{a^*(\psi_k)}\cdots a^*(\psi_n)a(\eta_n)\cdots \widehat{a(\eta_\ell)}\cdots a(\eta_1)\big).
\end{aligned} \tag{4.23}$$

Here $\varepsilon_{k,\ell} = \pm 1$, depending on the parity of the permutation needed to put the modes in the original order and $\widehat{a^*(\psi_k)}$ means that the factor $a^*(\psi_k)$ is removed from the product $a^*(\psi_1)\cdots a^*(\psi_n)$.

We now compute by repeated application of (4.2)

$$\begin{aligned} & a^*(Q\varphi)a(\eta_n)\cdots a(\eta_1) \\ & \quad = \langle \eta_n\,,\, Q\varphi\rangle a(\eta_{n-1})\cdots a(\eta_1) - a(\eta_n)a^*(Q\varphi)a(\eta_{n-1})\cdots a(\eta_1) \\ & \quad = \sum_\ell \varepsilon_\ell\, \omega_Q\big(a^*(\varphi)a(\eta_\ell)\big)\, a(\eta_n)\cdots \widehat{a(\eta_\ell)}\cdots a(\eta_1) \pm a(\eta_n)\cdots a(\eta_1)a^*(Q\varphi), \end{aligned} \tag{4.24}$$

with the upper sign for n even and the lower sign for n odd, therefore

$$\sum_\ell \varepsilon_\ell\, \omega_Q\big(a^*(\varphi)a(\eta_\ell)\big)\, a(\eta_n)\cdots \widehat{a(\eta_\ell)}\cdots a(\eta_1) = \big[a^*(Q\varphi)\,,\, a(\eta_n)\cdots a(\eta_1)\big]_\mp. \tag{4.25}$$

Using this relation, its conjugate and the anti-commutation relations (4.2), we get the desired result for gauge-invariant monomials and hence for all gauge-invariant elements

$$\begin{aligned} & \sum_{k,\ell} \varepsilon_{k,\ell}\, \omega_Q\big(a^*(\varphi)a(\eta_\ell)\big)\, \omega_Q\big(a^*(\psi_k)a(\varphi)\big) \\ & \qquad\qquad \times a^*(\psi_1)\cdots \widehat{a^*(\psi_k)}\cdots a^*(\psi_n)a(\eta_n)\cdots \widehat{a(\eta_\ell)}\cdots a(\eta_1) \\ & = -\sum_\ell \varepsilon_\ell\, \omega_Q\big(a^*(\varphi)a(\eta_\ell)\big) \\ & \qquad\qquad \times \mp\big[a(Q\varphi)\,,\, a^*(\psi_1)\cdots a^*(\psi_n)\big]_\mp a(\eta_n)\cdots \widehat{a(\eta_\ell)}\cdots a(\eta_1) \\ & = \pm\sum_\ell \varepsilon_\ell\, \omega_Q\big(a^*(\varphi)a(\eta_\ell)\big) \\ & \qquad\qquad \times \big\{a(Q\varphi)\,,\, a^*(\psi_1)\cdots a^*(\psi_n)a(\eta_n)\cdots \widehat{a(\eta_\ell)}\cdots a(\eta_1)\big\} \\ & = \pm\big\{a(Q\varphi)\,, a^*(\psi_1)\cdots a^*(\psi_n)\big[a^*(Q\varphi), a(\eta_n)\cdots a(\eta_1)\big]_\mp\big\} \\ & = \big\{a(Q\varphi)\,, \big[a^*(Q\varphi), a^*(\psi_1)\cdots a^*(\psi_n)a(\eta_n)\cdots a(\eta_1)\big]\big\}. \end{aligned} \tag{4.26}$$

□

4.3.1 Bipartite free states

We have seen that in the case of distinguishable particles the composition of two subsystems amounts to taking the tensor product of the corresponding Hilbert spaces (see Section 1.2.2). For indistinguishable particles composition is done at the level of the single particle Hilbert space. The direct sum is the natural way of composing fermionic subsystems. Two free states on different CAR algebras $\mathcal{A}(\mathcal{H}_1)$ and $\mathcal{A}(\mathcal{H}_1)$ define a unique product state on $\mathcal{A}(\mathcal{H}_1 \oplus \mathcal{H}_2)$ (see Eq. 4.7). Here we look at more general states on composed CAR algebras.

Let P be an orthogonal projection of $\mathcal{H}$, then the restriction of the free state ω_Q on $\mathcal{A}(\mathcal{H})$ is a free state on the sub-CAR algebra $\mathcal{A}(P\mathcal{H})$ with symbol PQP. Conversely, an orthogonal decomposition $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ turns $\mathcal{A}(\mathcal{H})$ into a composite system with parties $\mathcal{A}(\mathcal{H}_i)$ up to a minor modification: $\mathcal{A}(\mathcal{H}_i)$ sits as a graded tensor factor in $\mathcal{A}(\mathcal{H})$ through the natural identification $a^*(\varphi_i) \mapsto a^*(\varphi_i \oplus 0)$. This is due to the fact that odd elements in $\mathcal{A}(\mathcal{H}_1 \oplus 0)$ anti-commute with odd elements in $\mathcal{A}(0 \oplus \mathcal{H}_2)$. To simplify notation we shall often write $a^*(\varphi)$ instead of $a^*(\varphi \oplus 0)$.

We now consider a free state on a bipartite fermionic system $\mathcal{A}(\mathcal{H}_1 \oplus \mathcal{H}_2)$ defined by a symbol Q with block matrix structure

$$Q = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix}. \tag{4.27}$$

The positivity of the bipartite state imposes restrictions on its constituent matrices A, B and C. For example it is obvious that

$$0 \leq A \leq \mathbb{1} \text{ and } 0 \leq C \leq \mathbb{1}\,.$$

Furthermore, from the positivity conditions $Q \geq 0$ and $Q \leq \mathbb{1}$ of the symbol (4.27) on $\mathcal{H}_1 \oplus \mathcal{H}_2$ it follows that

$$B\ \ker(C) = B\ \ker(\mathbb{1} - C) = 0. \tag{4.28}$$

By restricting $\mathcal{H}_2$ to $\ker(C)$, we can assume $C = 0$ and then the requirement $Q \geq 0$ becomes

$$Q = \begin{bmatrix} A & B \\ B^* & 0 \end{bmatrix} \geq 0\,.$$

This means that $\det(Q) = -\det(BB^*) \geq 0$, which is only possible if $B = 0$ on $\ker(C)$.

In the following chapter we will be interested in computing expectation values of the form $\omega_Q(XY)$, where X and Y are gauge-invariant elements of $\mathcal{A}(\mathcal{H}_1)$ and $\mathcal{A}(\mathcal{H}_2)$ respectively. In this case the subalgebras $\mathcal{A}(\ker(C))$ and $\mathcal{A}(\ker(\mathbb{1} - C))$ of

$\mathcal{A}(\mathcal{H}_2)$ are irrelevant for computing the expectation values. There is therefore no loss in generality to assume that the kernels of C and $\mathbb{1} - C$ are trivial. In this case the positivity conditions can be restated as

$$0 < C < \mathbb{1}, \;\; BC^{-1}B^* \leq A \;\text{ and }\; B(\mathbb{1} - C)^{-1}B^* \leq \mathbb{1} - A. \tag{4.29}$$

In these inequalities, even if C^{-1} or $(\mathbb{1} - C)^{-1}$ are unbounded, $BC^{-1}B^*$ and $B(\mathbb{1}-C)^{-1}B^*$ extend to bounded operators on $\mathcal{H}_1$. The positivity conditions (4.29) can be recast into the requirement that

$$0 \leq A \leq \mathbb{1} \;\text{ and }\; 0 \leq C \leq \mathbb{1} \tag{4.30}$$

and that there exist operators

$$\begin{aligned} &D_i : \mathcal{H}_2 \to \mathcal{H}_1, \;\; |D_i| \leq 1, \;\; i = 1, 2 \;\text{ such that} \\ &B = A^{\frac{1}{2}} D_1 C^{\frac{1}{2}} = (\mathbb{1} - A)^{\frac{1}{2}} D_2 (\mathbb{1} - C)^{\frac{1}{2}}. \end{aligned} \tag{4.31}$$

4.3.2 Free maps

There is a notion of gauge-invariant, free, completely positive, identity preserving maps between CAR algebras that naturally extends that of free states. These maps preserve the free nature of a state, i.e. a free state is mapped into a free state. Such a map $\Gamma_{R,S} : \mathcal{A}(\mathcal{H}) \to \mathcal{A}(\mathcal{K})$ is determined by operators

$$R : \mathcal{H} \to \mathcal{K} \;\text{ and }\; S : \mathcal{H} \to \mathcal{H} \tag{4.32}$$

that satisfy

$$0 \leq S \leq \mathbb{1} - R^* R. \tag{4.33}$$

The free map acts on monomials by

$$\begin{aligned} &\Gamma_{R,S}(a^*(\varphi_1) \dots a^*(\varphi_m) a(\psi_n) \dots a(\psi_1)) \\ &\qquad := \sum_{p_1, p_2} \epsilon_1 \epsilon_2 \omega_S(a^*(\varphi_{k_1}) \dots a^*(\varphi_{k_r}) a(\psi_{l_r}) \dots a(\psi_{l_1})) \\ &\qquad\qquad \times a^*(R\varphi_{i_1}) \dots a^*(R\varphi_{i_{m-r}}) a(R\psi_{j_{n-r}}) \dots a(R\psi_{j_1}) \end{aligned} \tag{4.34}$$

Here $p_1 = \{\{k_1, \dots, k_r\}, \{i_1, \dots, i_{m-r}\}\}$ is a partition of $\{1, \dots, m\}$ and $p_2 = \{\{l_1, \dots, l_r\}, \{j_1, \dots, j_{m-r}\}\}$ is a partition of $\{1, \dots, n\}$. ϵ_1 is the sign of the permutation corresponding to p_1 and ϵ_2 the sign corresponding to p_2.

Example 14. *An almost trivial example of a free map is the identity map written as* $id = \Gamma_{\mathbb{1},0}$.

Example 15. *The action of the map* $\Gamma_{R,S}$ *on* $a^*(\varphi)$ *is simply*

$$\Gamma_{R,S}(a^*(\varphi)) = a^*(R\varphi)\,.$$

The action on a gauge-invariant monomial of order two is given by

$$\Gamma_{R,S}(a^*(\varphi)a(\psi)) = a^*(R\varphi)a(R\psi) + \langle\psi\,,\,S\,\varphi\rangle,\ \ \varphi,\psi\in\mathcal{H}. \tag{4.35}$$

For a third order monomial like $a^*(\varphi_1)a^*(\varphi_2)a(\psi)$ *for example we get*

$$\begin{aligned}\Gamma_{R,S}(a^*(\varphi_1)a^*(\varphi_2)a(\psi)) =& a^*(R\varphi_1)a^*(R\varphi_2)a(R\psi)\\ &+\langle\varphi_1,S\psi\rangle a^*(R\varphi_2)+\langle\varphi_2,S\psi\rangle a^*(R\varphi_1)\,,\end{aligned}$$

and so on.

The quasi-free map $\Gamma_{R,S}$ can be shown [19] to be a concatenation of three simple completely positive maps. First there is the injection $a^*(\varphi)\mapsto a^*(\varphi\oplus 0)$ from $\mathcal{A}(\mathcal{H})$ to $\mathcal{A}(\mathcal{H}\oplus\mathcal{H})$. Then $\mathcal{A}(\mathcal{H}\oplus\mathcal{H})$ is mapped into $\mathcal{A}(\mathcal{K}\oplus\mathcal{K})$ by $a^*(\varphi\oplus\psi)\mapsto a^*(U(\varphi\oplus\psi))$ with U the unitary operator

$$U=\begin{bmatrix} R & \sqrt{\mathbb{1}-R^*R}\\ -\sqrt{\mathbb{1}-R^*R} & R\end{bmatrix}.$$

Finally $\mathcal{A}(\mathcal{K}\oplus\mathcal{K})$ is mapped back into $\mathcal{A}(\mathcal{K})$ by $a^*(\varphi\oplus\psi)\mapsto a^*(\varphi).\omega_S(a^*(\psi))$.

Furthermore it was shown there that the Jamiołkowski state corresponding to the free map $\Gamma_{R,S}$ is unitarily equivalent to the free state with symbol

$$Q_J=\frac{1}{2}\begin{bmatrix}\mathbb{1} & R\\ R^* & R^*R+2S\end{bmatrix}.$$

Note that this again gives the complete positivity of the map since positivity of Q_J corresponds to $0\le S\le\mathbb{1}-R^*R$.

As for free states, we introduce the set

$$\begin{aligned}\mathcal{CP}(\mathcal{H},\mathcal{K}) = \big\{(R,S)\ \big|\ & R:\mathcal{H}\to\mathcal{K}\text{ and }S:\mathcal{H}\to\mathcal{H}\text{ bounded linear}\\ &\text{operators such that }0\le S\le\mathbb{1}-R^*R\big\}.\end{aligned} \tag{4.36}$$

We use $\mathcal{CP}(\mathcal{H})$ for $\mathcal{CP}(\mathcal{H},\mathcal{H})$. The set of free CP maps extends that of free states by putting

$$Q\in\mathcal{Q}(\mathcal{H})\mapsto(0,Q)\in\mathcal{CP}(\mathcal{H},\mathcal{K}). \tag{4.37}$$

Another special distinguished class of maps are the free homomorphisms from $\mathcal{A}(\mathcal{H})$ to $\mathcal{A}(\mathcal{K})$

$$\{(V,0) \in \mathcal{CP}(\mathcal{H},\mathcal{K}) \mid V : \mathcal{H} \to \mathcal{K} \text{ isometric}\}. \tag{4.38}$$

The set $\mathcal{CP}(\mathcal{H},\mathcal{K})$ is also convex and weakly compact.

Free CP maps transform free states into free states and one checks from (4.4) and (4.35) that

$$\omega_Q \circ \Gamma_{R,S} = \omega_{R^*QR+S}. \tag{4.39}$$

For more details, we refer to [19].

Chapter 5

Fermionic conditional state spaces

5.1 Introduction

In Chapter 3 we have seen how conditional state spaces can be used to study the correlations between two subsystems. In this chapter we characterize the possible conditional state spaces that arise in bipartite free fermionic states.

Two different characterizations are given. One is in terms of bounds on the operator describing the two-point correlations of the free fermionic state. A second characterization is in terms of a completely positive map, similar to what was described in Section 3.3.2.

All results presented in this chapter are new contributions and have been published in [24].

5.2 Correlation bounds

In Chapter 3 we have introduced the conditional state spaces for composed systems of distinguishable particles. These conditional state spaces describe the states that can be obtained on one subsystem by perturbing another. Here we will do the same for composed Fermionic systems. The aim is to characterize all free states on $\mathcal{A}(\mathcal{H}_1)$ that arise as conditional states of bipartite free states. More precisely, to

characterize

$$\mathcal{S}_1^{\text{free}} = \Big\{ \omega_{\tilde{A}} \;\Big|\; \omega_{\tilde{A}} \text{ is a free state on } \mathcal{A}(\mathcal{H}_1) \text{ and } \exists \text{ a gauge-invariant } Y \in \mathcal{A}(\mathcal{H}_2) \text{ such that } Y \geq 0 \text{ and } \omega_{\tilde{A}}(X) = \omega_Q(XY),\ X \in \mathcal{A}(\mathcal{H}_1) \Big\}. \tag{5.1}$$

In this section we derive a description of the conditional state space S_1^{free} in terms of bounds on the two point correlations $\tilde{A}$. A first proposition bounds the two-point correlation matrix of a conditional state. A second proposition shows the converse for free Fermionic states, i.e. if the symbol satisfies the given bounds then it can be obtained as a conditional state.

First we show that if a state is obtained by conditioning with a certain gauge-invariant operator Y, then an operator $\tilde{A}$ that contains its second order moments fulfils certain bounds, depending on the operators that make up the composed state ω_Q. Note that here we do not yet restrict the conditional state to be a free one.

Proposition 2. *With the assumptions and notations of above there exists for any positive, gauge-invariant $Y \in \mathcal{A}(\mathcal{H}_2)$ with $\omega_Q(Y) = 1$ a bounded operator $\tilde{A}$ on $\mathcal{H}_1$ such that*

$$\omega_Q\big(a^*(\varphi)a(\psi)Y\big) = \langle \psi\,,\, \tilde{A}\varphi \rangle,\ \ \varphi, \psi \in \mathcal{H}_1 \tag{5.2}$$

and

$$A - BC^{-1}B^* \leq \tilde{A} \leq A + B(\mathbb{1} - C)^{-1}B^*. \tag{5.3}$$

Proof. $\tilde{A}$ is the single particle reduced density operator of the state $\omega_{Q,Y}$ conditioned on Y: $\omega_{Q,Y}(X) = \omega_Q(XY)$. It can be defined through its elements:

$$\langle e_i|\tilde{A}|e_j\rangle = \omega_Q(a^*(e_j)a(e_i)Y)\,,$$

where the e_i form a basis of $\mathcal{H}_1$. To prove the matrix inequalities, we determine bounds for $\langle \varphi|\tilde{A}\varphi\rangle$.

Since Y commutes with $a(\varphi)$ and $\omega_Q(Y) = 1$, we can use Lemma 1 to get

$$\begin{aligned} \langle \varphi\,,\, \tilde{A}\varphi\rangle &= \omega_Q(a^*(\varphi)Ya(\varphi)) \\ &= \langle \varphi\,,\, A\varphi\rangle + \omega_C\big(\big\{a(B^*\varphi)\,,\, \big[a^*(B^*\varphi)\,,\, Y\big]\big\}\big). \end{aligned} \tag{5.4}$$

The next step consists in rewriting this identity in such a way that we can use the information $\omega_C(Y) = 1$. This can be achieved through the KMS-condition. Using an approximation argument we may assume that $a^*(B^*\varphi)$ and Y are analytic elements for the automorphism group α of $\mathcal{A}(\mathcal{H}_2)$ defined by C. For an analytic element x we have

$$\begin{aligned}\big(\alpha_z(x)\big)^* &= \alpha_{\overline{z}}(x^*), \quad z \in \mathbb{C} \text{ and}\\ \omega_C(xY) &= \omega_C\big(xY^{\frac12}Y^{\frac12}\big) = \omega_C\big(\alpha_{-i}(Y^{\frac12})xY^{\frac12}\big)\\ &= \omega_C\big(\alpha_{-\frac{i}{2}}(Y^{\frac12})\alpha_{\frac{i}{2}}(x)\alpha_{\frac{i}{2}}(Y^{\frac12})\big),\end{aligned} \tag{5.5}$$

and similarly

$$\begin{aligned}\omega_C(Yx) &= \omega_C\big(\alpha_{-\frac{i}{2}}(Y^{\frac12})\alpha_{-\frac{i}{2}}(x)\alpha_{\frac{i}{2}}(Y^{\frac12})\big),\\ \omega_C(xYy) &= \omega_C\big(\alpha_{-\frac{i}{2}}(Y^{\frac12})\alpha_{-\frac{i}{2}}(y)\alpha_{\frac{i}{2}}(x)\alpha_{\frac{i}{2}}(Y^{\frac12})\big).\end{aligned} \tag{5.6}$$

This allows us to rewrite (5.4) as

$$\langle\varphi\,,\,\tilde{A}\varphi\rangle = \langle\varphi\,,\,A\varphi\rangle + \omega_C\big(\alpha_{-\frac{i}{2}}(Y^{\frac12})u\alpha_{\frac{i}{2}}(Y^{\frac12})\big) \tag{5.7}$$

with

$$u = \big\langle B^*\varphi\,,\,\big(\mathbb{1}+\mathrm{e}^{-h}\big)B^*\varphi\big\rangle - a^*\big(\mathrm{e}^{\frac{h}{2}}B^*\varphi+\mathrm{e}^{-\frac{h}{2}}B^*\varphi\big)a\big(\mathrm{e}^{\frac{h}{2}}B^*\varphi+\mathrm{e}^{-\frac{h}{2}}B^*\varphi\big). \tag{5.8}$$

Here h is the single mode Hamiltonian as in (4.15) replacing Q by C. Using

$$0 \le a^*(\zeta)a(\zeta) \le |\zeta|^2\mathbb{1} \tag{5.9}$$

we obtain the statement of the proposition

$$A - BC^{-1}B^* \le \tilde{A} \le A + B(\mathbb{1}-C)^{-1}B^*. \tag{5.10}$$

□

Obviously, since the states in S_1^{free} are also obtained as conditional states, their two-point correlations satisfy the bounds in this proposition. Since the two-point correlations of a free state ω_Q are encoded in its symbol Q, the operator Q will satisfy the bounds given for $\tilde{A}$ in Proposition 2. In Proposition 3, we show that the converse is also true, i.e. that every free state whose two-point correlations satisfy the given bounds is contained in the weak*-closure of S_1^{free}.

To prove this statement, we first look at conditional states generated by conditioning an exponential element Y in $\mathcal{A}(\mathcal{H}_2)$.

Lemma 2. *If ω_Q is a free state on $\mathcal{A}(\mathcal{H})$ with symbol Q as in (4.27) and $Y = E(L)/\omega_Q(E(L))$ is an exponential element in $\mathcal{A}(\mathcal{H}_2)$ with $L \geq 0$, then the conditional state $\tilde{\omega} : X \mapsto \omega_Q(XY)$ is a free state on $\mathcal{A}(\mathcal{H}_1)$ with symbol*

$$\tilde{A} = A - B(L - \mathbb{1})(\mathbb{1} - C + CL)^{-1}B^*. \tag{5.11}$$

Proof. We calculate the expectation value of elements $X = E(K)$ with K an operator on $\mathcal{H}_1$ in the state $\tilde{\omega}$. Since these elements $E(K)$ span the gauge invariant CAR algebra [19], these values determine the state $\tilde{\omega}$.

First we determine the normalization factor $\omega_Q(E(L))$ by using (4.21)

$$\begin{aligned} \omega_Q(E(L)) \quad &= \det(\mathbb{1} - Q + Q(\mathbb{1} \oplus L)) \\ &= \det \begin{bmatrix} \mathbb{1} & -B + BL \\ 0 & \mathbb{1} - C + CL \end{bmatrix} \\ &= \det(\mathbb{1} - C + CL). \end{aligned} \tag{5.12}$$

Likewise, we have that

$$\begin{aligned} \omega_Q(E(K)E(L)) \qquad &= \det(\mathbb{1} - Q + Q(K \oplus L)) \\ &= \det \begin{bmatrix} \mathbb{1} - A + AK & -B + BL \\ -B^* + B^*K & \mathbb{1} - C + CL \end{bmatrix} \\ &= \det(\mathbb{1} - C + CL) \det(\mathbb{1} - \tilde{A} + \tilde{A}K) \end{aligned} \tag{5.13}$$

with

$$\tilde{A} = A - B(L - \mathbb{1})(\mathbb{1} - C + CL)^{-1}B^*. \tag{5.14}$$

Hence, $\tilde{\omega}$ is a free state with symbol $\tilde{A}$

$$\tilde{\omega}(E(K)) = \frac{\omega_Q(E(K)E(L))}{\omega_Q(E(L))} = \det(\mathbb{1} - \tilde{A} + \tilde{A}K) = \omega_{\tilde{A}}(E(K)). \tag{5.15}$$

□

Now that we have an expression for the conditional states formed from exponential elements, we can use it to show that almost all symbols fulfilling the correlation bounds can be recovered by choosing the exponential element.

Lemma 3. *Let $0 < \varepsilon < 1$ and let $\tilde{A}$ be an operator on $\mathcal{H}_1$ such that $A - \tilde{A}$ is of finite rank and such that*

$$A - (1-\varepsilon)BC^{-1}B^* \leq \tilde{A} \leq A + (1-\varepsilon)B(\mathbb{1} - C)^{-1}B^*, \tag{5.16}$$

then there exists a positive $Y \in \mathcal{A}(\mathcal{H}_2)$ such that

$$\omega_{\tilde{A}}(X) = \omega_Q(XY), \;\; X \in \mathcal{A}(\mathcal{H}_1). \tag{5.17}$$

Proof. We consider the set of operators

$$\tilde{A} = A + BKB^* \tag{5.18}$$

with K a finite rank operator on $\mathcal{H}_2$ such that $\tilde{A}$ satisfies the bounds (5.16). This is the case if

$$-(1-\varepsilon)C^{-1} \leq K \leq (1-\varepsilon)(\mathbb{1} - C)^{-1}. \tag{5.19}$$

Using Lemma 2, we obtain the free state with symbol $\tilde{A}$ as the conditional state $X \mapsto \omega_Q(XY)$ with

$$Y = \frac{E(L)}{\omega_Q(E(L))} \in \mathcal{A}(\mathcal{H}_2) \tag{5.20}$$

if we are able to find a positive operator L on $\mathcal{H}_2$ such that

$$K = (\mathbb{1} - L)(\mathbb{1} - C + CL)^{-1} \text{ and } \mathbb{1} - L \text{ finite rank.} \tag{5.21}$$

Rewriting this in terms of a finite rank operator N, such that $L = \mathbb{1} + N$, we have

$$K = -N(\mathbb{1} + CN)^{-1}. \tag{5.22}$$

If $\mathbb{1} + CK$ is invertible, this equation is solved by

$$N = -K(\mathbb{1} + CK)^{-1}. \tag{5.23}$$

To show that $\mathbb{1} + CK$ is invertible, assume that $\varphi \in \ker\{\mathbb{1} + KC\}$. This implies that

$$\langle C\varphi, \varphi\rangle + \langle C\varphi, KC\varphi\rangle = 0 \tag{5.24}$$

and

$$0 \geq \langle C\varphi, \varphi\rangle - \langle C\varphi, (1-\varepsilon)C^{-1}C\varphi\rangle = \varepsilon\langle\varphi, C\varphi\rangle. \tag{5.25}$$

Hence $\ker\{\mathbb{1} + KC\} = \{0\}$. Therefore, as CK is of finite rank, $\mathrm{ran}(\mathbb{1} + CK) = \mathcal{H}_2$. Furthermore $\ker\{\mathbb{1} + CK\} = \{0\}$ as well and so $\mathbb{1} + CK$ is invertible. □

Proposition 2 gives us bounds that the symbol of states in $\mathcal{S}_1^{\text{free}}$ has to fulfil. Lemma 3 tells us that almost all the symbols that fulfil these bounds are in $\mathcal{S}_1^{\text{free}}$. Taken together, this gives us a characterization of $\mathcal{S}_1^{\text{free}}$ in terms of bounds on the symbol.

Proposition 3. *The weak*-closure of the set $\mathcal{S}_1^{\text{free}}$ of conditioned free states on $\mathcal{A}(\mathcal{H}_1)$ coincides with the set of free states on $\mathcal{A}(\mathcal{H}_1)$ whose symbols $\tilde{A}$ satisfy*

$$A - BC^{-1}B^* \leq \tilde{A} \leq A + B(\mathbb{1} - C)^{-1}B^*. \tag{5.26}$$

Proof. For free states, weak*-convergence is equivalent to weak convergence of their symbols. The proof then immediately follows from Proposition 2 and Lemma 3. □

Example 16. *As a trivial example consider a product state, i.e. a state with $B = 0$. From Proposition 3 it then follows that the conditional state space is zero-dimensional as we have seen in the general case in the Section 3.3.2.*

Example 17. *Another application of this characterization of the conditional state space can be made for pure states. A quasi-free state is pure if and only if its symbol Q is a projector [19], i.e. $Q^2 = Q$. For a bipartite state with symbol*

$$Q = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} \tag{5.27}$$

this means that

$$A^2 + BB^* = A$$

$$B^*B + C^2 = C$$

$$AB + BC = B$$

$$B^*A + CB^* = B^*$$

Through the polar decomposition of B, we get from the first equation that $B = \sqrt{A(\mathbb{1} - A)}U$ for some unitary U. Using this in the third equation, we get $C = U^*(\mathbb{1} - A)U$. From Proposition 3 it now follows that $0 \leq \tilde{A} \leq \mathbb{1}$. The conditional state space becomes a full quasi-free state space, as in the general case of Section 3.3.2. The role of the Schmidt number is played here by the dimension of the image of A.

5.3 Manifestly positive model

As in (3.22), we can write the free conditional states as generalized subsystems, using a free completely positive map to a suitable operator algebra and free states on the target algebra.

Proposition 4. *There exists a unique, free, minimal, identity preserving, completely positive map* Γ *such that the weak*-closure of* $\mathcal{S}_1^{\text{free}}$ *is the pull-back of the free states by* Γ.

Proof. Let $\mathcal{K} = \overline{\text{ran}(B)} \subset \mathcal{H}_1$. We construct operators

$$R : \mathcal{H}_1 \to \mathcal{K} \text{ and } S : \mathcal{H}_1 \to \mathcal{H}_1 \tag{5.28}$$

such that

$$0 \leq S \leq \mathbb{1} - R^* R. \tag{5.29}$$

These operators define a completely positive, free, identity preserving map Γ from $\mathcal{A}(\mathcal{H}_1)$ to $\mathcal{A}(\mathcal{K})$ as in (4.35). The pull-back of the free states on $\mathcal{A}(\mathcal{K})$ consists of the free states on $\mathcal{A}(\mathcal{H}_1)$ with symbols

$$\{\tilde{A} = R^* T\, R + S \mid 0 \leq T \leq \mathbb{1}\}. \tag{5.30}$$

We need to show that the set (5.30) coincides with (5.26). This is the case if and only if

$$R = U \sqrt{B\, C^{-1} B^* + B\, (\mathbb{1} - C)^{-1} B^*} \text{ and } S = A - B\, C^{-1} B^*. \tag{5.31}$$

In this expression U is an arbitrary unitary on $\mathcal{K}$. □

This manifestly positive model of the conditional state space corresponds to the general model of Section 3.3.2. As was the case there, the model consists of a full quasi-free state space and a completely positive map.

Example 18. *Let's consider the simple case where* $\mathcal{H}_1 = \mathcal{H}_2 = \mathbb{C}$. *The symbol of the bipartite free state then takes the form*

$$Q = \begin{bmatrix} a & b \\ b^* & c \end{bmatrix},$$

with $a, b, c \in \mathbb{C}$. *Positivity of the state* ω_Q *means* $0 \leq Q \leq \mathbb{1}$, *from which we see that* $0 \leq a \leq 1$, $0 \leq a \leq 1$, $|b|^2 \leq ac$ *and* $|b|^2 \leq (1-a)(1-c)$.

Proposition 3 tells us that the conditional states are all $\omega_{\tilde{a}}$ *with*

$$a - \frac{|b|^2}{c} \leq \tilde{a} \leq a + \frac{|b|^2}{1-c}.$$

Note that the positivity requirements on a, b and c ensure that $0 \leq \tilde{a} \leq 1$. It is evident that if the symbols a and c are fixed, the correlations are determined by b and that the conditional state space increases accordingly with larger values of $|b|$.

It is clear that by taking $R = e^{\imath\theta}\sqrt{\frac{|b|^2}{c} + \frac{|b|^2}{1-c}}$ and $S = a - \frac{|b|^2}{c}$ that this interval coincides with that prescribed by Proposition 4.

5.4 Conclusion

In this chapter we described the correlations in bipartite quantum states in terms of a space of conditional states. These states are restrictions of the global state to one party after perturbing the other. For finite dimensional systems one can always find a manifestly positive model of the conditional states in terms of an auxiliary system and a completely positive map. Separable states correspond to classical models.

The main result of this chapter is to work out the details of the manifestly positive model in the setting of Gaussian Fermionic systems. This greatly simplifies the analysis as Gaussian structures are characterized by one-particle objects or, stated differently, tensor constructions are replaced by direct sums. This allowed us to fully carry out the description of conditional state spaces in this chapter. The conditional state space has been characterized by two inequalities on the quasi-free symbol describing the conditional state and also by a manifestly positive description in terms of a quasi-free state space and a completely positive map.

Chapter 6

Fermionic Markov chains

6.1 Introduction

In Chapter 2, the entropy density of a hidden Markov process was calculated by determining the entropy rate instead of the entropy density, as these quantities are equal for translation-invariant processes. In Chapter 3 we then put forward a quantum counterpart for classical Markov processes. Finally in Chapter 4 we introduced a class of quantum states for which explicit calculations can be made, the free Fermions.

In this chapter we will put these ingredients together; we will do an entropy density calculation for free Fermionic Markov processes, based on the entropy rate approach that has proven successful for classical hidden Markov processes.

In the first part of the chapter we repeat the construction of quantum Markov states for the free Fermionic case. The operators describing the two-point correlations, the symbol of the state, will be in this case a Toeplitz matrix, a matrix where the elements along diagonals are equal.

In the second part we calculate the entropy rate of the Fermionic Markov process. For the states we propose in this chapter, a direct calculation of the entropy density is possible. This calculation is based on Szegö's theorem for densities of trace functions of Toeplitz matrices. We first give a new generalization of this theorem to block Toeplitz matrices before moving on to the calculation of rate functions for matrices. We then present a new theorem on the asymptotic behaviour of the eigenvalues of Toeplitz matrices, much like Szegö's theorem, but then focused on asymptotic rates rather than densities.

6.2 Constructing processes

In Section 4.3.1, we have encountered bipartite free Fermionic states. The symbol operators Q that define such states are 2×2 block matrices, where the submatrices describe the correlations in the different parties and the correlations between them (see Eq. 4.27). There are natural embeddings

$$a(\varphi) \mapsto a(\varphi \oplus 0) \text{ and } a(\psi) \mapsto a(0 \oplus \psi) \tag{6.1}$$

of $\mathfrak{A}(\mathfrak{H})$ and $\mathfrak{A}(\mathfrak{K})$ into $\mathfrak{A}(\mathfrak{H} \oplus \mathfrak{K})$. Both factors together generate the large algebra and they satisfy graded commutation relations as creation operators in different factors anticommute.

Now we construct a semi-infinite chain on an infinite number of Fermionic parties. We can transport the construction of the quantum Markov chain (3.35) to the free Fermionic setting. The spin chain algebra $\otimes^{\mathbb{N}}\mathcal{M}_d$ is replaced by a semi-infinite Fermionic chain $\mathfrak{A}(\bigoplus^{\mathbb{N}} \mathfrak{H})$ where $\mathfrak{A}(\mathfrak{H})$ is now the one site algebra. Hence the symbol will be a semi-infinite block matrix. As the construction presented here is analogous to that in Section 3.4 the state is also translation-invariant. For free Fermions this will mean that the symbol is a block Toeplitz matrix.

6.2.1 Extending free maps

The basic ingredient of the construction for distinguishable parties was a completely positive map $\Lambda : \mathcal{M}_d \otimes \mathcal{M}_d \to \mathcal{M}_d$ that allowed us to contract multi-partite observables to the single-party observable algebra $\mathcal{M}_d$. This map was constructed to be compatible with a single-party CP map $\Gamma : \mathcal{M}_d \to \mathcal{M}_d$ (see Eq. 3.33). In the free Fermionic setting, this map turns into a free completely positive transformation $\Lambda_{C,D}$ from $\mathfrak{A}(\mathfrak{H} \oplus \mathfrak{H})$ to $\mathfrak{A}(\mathfrak{H})$ such that $\Lambda_{C,D}$ extends the single-party free CP-map $\Lambda_{A,B}$ of $\mathfrak{A}(\mathfrak{H})$:

$$\Lambda_{C,D} \circ \jmath_1 = \Lambda_{C,D} \circ \jmath_2 = \Lambda_{A,B}. \tag{6.2}$$

Here, $\jmath_1$ and $\jmath_2$ are the natural embeddings of $\mathfrak{A}(\mathfrak{H})$ into the first and second factor of $\mathfrak{A}(\mathfrak{H} \oplus \mathfrak{H})$

$$\jmath_1(a(\varphi)) = a(\varphi \oplus 0) \text{ and } \jmath_2(a(\varphi)) = a(0 \oplus \varphi). \tag{6.3}$$

This compatibility condition on the map corresponds to the one given for distinguishable particles in Eq. 3.33.

Applying the compatibility condition (6.2) to $a(\varphi)$ we see that

$$a(C(\varphi \oplus 0)) = a(C(0 \oplus \varphi)) = a(A\varphi)$$

and so

$$C = \begin{bmatrix} A & A \end{bmatrix} . \tag{6.4}$$

Doing the same for $a^*(\varphi)a(\psi)$ we get

$$\begin{aligned} &a^*(C(\varphi \oplus 0))a(C(\psi \oplus 0)) + \langle \psi \oplus 0, D(\varphi \oplus 0)\rangle \\ &= a^*(C(0 \oplus \varphi))a(C(0 \oplus \psi)) + \langle 0 \oplus \psi, D(0 \oplus \varphi)\rangle \\ &= a^*(A\varphi)a(A\psi) + \langle \psi, B\varphi\rangle \end{aligned}$$

from which we get

$$D = \begin{bmatrix} B & X \\ X^* & B \end{bmatrix} ,$$

where X is as of yet undetermined and allows for some freedom in the choice of D.

Because of the structure of free CP maps, the compatibility conditions (6.4) are not only necessary but also sufficient and we can rephrase the whole construction on the level of symbols. Doing so, the graded tensor products in the case of distinguishable particles become direct sums in the case of free Fermions.

6.2.2 Invariant states

The construction of the quantum Markov process consists of using a completely positive map to contract the observable and then applying a single-party state that is invariant under $\Lambda_{A,B}$. We have the following lemma concerning the existence of such invariant states.

Lemma 4. *Let $\Lambda_{A,B}$ be a completely positive free transformation of $\mathcal{A}(\mathcal{H})$ as in (4.35) and assume that $\dim(\mathcal{H}) < \infty$, then $\Lambda_{A,B}$ has a unique invariant state if and only if $|A| < 1$. Moreover, the unique invariant state is free with symbol Q determined by*

$$Q = A^* Q\, A + B. \tag{6.5}$$

Proof. The condition $|A| < 1$ is equivalent to the non-existence of non-trivial solutions to the homogeneous equation $Q = A^* Q\, A$. It has to be satisfied to have uniqueness of the solution of the invariance condition (6.5) for symbols. Conversely, suppose that $|A| < 1$, then there exists by the fixed point theorem for contractions a unique Q such that

$$Q = A^* Q\, A + B. \tag{6.6}$$

This Q satisfies $0 \leq Q \leq \mathbb{1}$ as we may obtain Q by iterating the map $X \mapsto A^* X A + B$ with initial value 0. It is then easily checked that

$$\lim_{n\to\infty} \Lambda_{A,B}^n = \omega_Q \tag{6.7}$$

which guarantees both the uniqueness of the invariant state and its free character. □

6.2.3 Markov construction

We now have all necessary components to construct free Fermionic Markov processes.

Starting from a single-party CP map $\Lambda_{A,B} : \mathcal{A}(\mathcal{H}) \to \mathcal{A}(\mathcal{H})$ defined by operators A and B, and with an invariant free state ω_Q, we first have to extend it to an CP map $\Lambda_{C,D} : \mathcal{A}(\mathcal{H} \oplus \mathcal{H}) \to \mathcal{A}(\mathcal{H})$ compatible with $\Lambda_{A,B}$. This involves fixing a choice for X such that $\Lambda_{C,D}$ is still completely positive. We will shortly see what the requirements are on A and B for this to be possible.

Using the map $\Lambda_{C,D}$ we can then pull multi-partite operators down to the level of the state ω_Q. On the other hand, we can also look at this as the dual map of $\Lambda_{C,D}$ pulling the state ω_Q up to the level of the multi-partite observables. This will give us ever larger free states, with block symbols Q_n that increase dimensions at each step.

Construction. *Let $(A, B) \in \mathcal{CP}(\mathcal{H})$ and let $Q \in \mathcal{Q}(\mathcal{H})$ be such that ω_Q is invariant under $\Lambda_{A,B}$:*

$$Q = A^* Q A + B. \tag{6.8}$$

Let $X : \mathcal{H} \to \mathcal{H}$ satisfy the compatibility condition

$$(C, D) \in \mathcal{CP}(\mathcal{H} \oplus \mathcal{H}, \mathcal{H}) \text{ with } C \text{ and } D \text{ as in (6.4).} \tag{6.9}$$

The free Markov chain defined by X and Q is the symbol

$$Q_\infty = \underset{n\to\infty}{w\text{-}lim}\, P_n R_n P_n^* \text{ on } \oplus^{\mathbb{N}} \mathcal{H} \tag{6.10}$$

where

$$P_n : \mathcal{H} \oplus \left(\oplus_{k=0}^{n-1} \mathcal{H}\right) \to \left(\oplus_{k=0}^{n-1} \mathcal{H}\right) : \varphi \oplus \psi_n \mapsto \psi_n \tag{6.11}$$

$$R_0 = Q \text{ and } R_{n+1} = \Big(C^* \oplus (\oplus^n \mathbb{1})\Big) R_n \Big(C \oplus (\oplus^n \mathbb{1})\Big) + \Big(D \oplus (\oplus^n 0)\Big). \tag{6.12}$$

We can also explicitly calculate the block elements of the symbol Q_∞.

Proposition 5. *The symbol Q_∞ in (6.10) is an Hermitian block Toeplitz matrix with entries*

$$\big(Q_\infty\big)_{i\,i} = Q \;\; \textit{and} \;\; \big(Q_\infty\big)_{i\,i+n} = (A^*)^n(Q - B + X). \tag{6.13}$$

Here $i = 0, 1, 2, \ldots$ and $n = 1, 2, 3, \ldots$

Proof. The proof consists in a straightforward computation of the consecutive R_n in (6.12) combined with the invariance equation (6.8).

For example, we have that $R_0 = Q$, so

$$R_1 = C^*QC + D = \begin{bmatrix} A^*QA + B & A^*QA + X \\ A^*QA + X^* & A^*QA + B \end{bmatrix}.$$

Using the invariance $A^*QA + B = Q$ this becomes

$$R_1 = \begin{bmatrix} Q & Q - B + X \\ Q - B + X^* & Q \end{bmatrix}.$$

The next step gives

$$\begin{aligned} R_2 &= (C^* \oplus \mathbb{1})R_1(C \oplus \mathbb{1}) + (D \oplus 0) \\ &= \begin{bmatrix} Q & Q - B + X & A^*(Q - B + X) \\ Q - B + X^* & Q & A^*(Q - B + X) \\ (Q - B + X^*)A & (Q - B + X^*)A & Q \end{bmatrix}, \end{aligned}$$

and so on.

The projections P_n cut off the first row and column, giving in the limit the given result for Q_∞.

$$Q_\infty = \begin{bmatrix} Q & A^*(Q - B + X) & (A^*)^2(Q - B + X) & \cdots \\ (Q - B + X^*)A & Q & A^*(Q - B + X) & \cdots \\ (Q - B + X^*)A^2 & (Q - B + X^*)A & Q & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

□

6.2.4 Extendability

As we have mentioned earlier, there is a freedom in choosing the channel $\Lambda_{C,D}$. The operator X has to chosen such that the $\Lambda_{C,D}$ is completely positive, i.e.

$$0 \leq D \leq \mathbb{1} - C^*C\,.$$

One may wonder if and when this is possible. Let's look at the simplest case, where the single-particle space is $\mathcal{A}(\mathbb{C})$.

Example 19. *In the case where $\mathcal{H} = \mathbb{C}$, we have*

$$D = \begin{bmatrix} b & x \\ \overline{x} & b \end{bmatrix} \text{ and } C = \begin{bmatrix} a & a \end{bmatrix},$$

with $a, x \in \mathbb{C}$ and $b \in \mathbb{R}$. From the complete positivity of $\Lambda_{a,b}$, we know that $0 \leq b \leq 1 - |a|^2$. The complete positivity conditions for C and D limit the choice for x. From $0 \leq D$ we see that

$$|x| \leq b\,.$$

From $D \leq \mathbb{1} - C^*C$ on the other hand, we get that

$$|x + |a|^2| \leq 1 - |a|^2 - b\,.$$

These two inequalities means that x has to lie in the intersection of two circles in the complex plane, one centred around 0 with radius $r_1 = b$ and another one centred around $-|a|^2$ with radius $r_2 = 1 - |a|^2 - b$. These two circles have an intersection when the distance between the centres is smaller than the sum of the radii. Hence, the channel is extendible if

$$|a|^2 \leq \frac{1}{2}$$

For the general case we get the following lemma.

Lemma 5. *The compatibility condition (6.9) is satisfiable if and only if*

$$A^*A \leq \min\big(\{\tfrac{1}{2}\,\mathbb{1}, \mathbb{1} - B\}\big). \tag{6.14}$$

Proof. We look for the necessary and sufficient conditions for the existence of a $X : \mathcal{H} \to \mathcal{H}$ such that

$$\begin{bmatrix} B & X \\ X^* & B \end{bmatrix} \geq 0 \text{ and } \begin{bmatrix} \mathbb{1} - A^*A - B & -A^*A - X \\ -A^*A - X^* & \mathbb{1} - A^*A - B \end{bmatrix} \geq 0. \tag{6.15}$$

Clearly $0 \leq B \leq \mathbb{1} - A^*A$ as $(A, B) \in \mathcal{CP}(\mathcal{H})$. The remaining positivity conditions are then the existence of S and T with

$$\begin{aligned} &|S| \leq 1,\ \ |T| \leq 1,\ \ X = B^{\frac{1}{2}} S B^{\frac{1}{2}},\ \text{ and} \\ &A^*A + X = (\mathbb{1} - A^*A - B)^{\frac{1}{2}} T (\mathbb{1} - A^*A - B)^{\frac{1}{2}}. \end{aligned} \tag{6.16}$$

Replacing S and T by their Hermitian parts, we may restrict to Hermitian X and so we need

$$[-B\,,\, B] \cap [-\mathbb{1} + B\,,\, \mathbb{1} - 2A^*A - B] \neq \emptyset \tag{6.17}$$

or, equivalently, that

$$[\mathbb{1}\,,\,\mathbb{1}+2B]\cap[2B\,,\,2\mathbb{1}-2A^*A]\neq\emptyset. \tag{6.18}$$

But this is the case if and only if

$$\max\big(\{\mathbb{1},2B\}\big)\leq 2\mathbb{1}-2A^*A \;\text{ or }\; A^*A\leq\min\big(\{\tfrac{1}{2}\,\mathbb{1},\mathbb{1}-B\}\big). \tag{6.19}$$

□

6.3 Entropy density

In this section we compute the entropy h for our chain, see (6.21). We can associate an entropy to a Fermionic Markov chain using (4.6)

$$\mathsf{h}(X,Q):=\lim_{n\to\infty}\frac{1}{n}\,\mathsf{S}(P_nR_nP_n^*) \tag{6.20}$$

$$:=\lim_{n\to\infty}\frac{1}{n}\Big(-\operatorname{Tr}P_nR_nP_n^*\log(P_nR_nP_n^*) - \operatorname{Tr}(\mathbb{1}-P_nR_nP_n^*)\log(\mathbb{1}-P_nR_nP_n^*)\Big). \tag{6.21}$$

A first method to compute this relies directly on the expression (4.6) for the entropy of a free state in terms of its symbol and on the structure of the symbols Q_∞ in Proposition 5. A second way is to rewrite the entropy as the asymptotic rate of disorder, as in the classical case, see Section 2.4.2. This last approach was used in [8, 21] to compute the entropy of a hidden Markov process. The first method uses the full local restrictions of the state while the second relies on the incremental structure of the local states given by a transfer matrix like construction, see (3.35) and (6.12).

6.3.1 Direct approach

The first approach to calculating the entropy density uses an extension of Szegö's theorem to block Toeplitz matrices $\hat{T}$. This theorem allows to calculate asymptotic densities of trace functions of Toeplitz matrices. A block Toeplitz matrix is a block matrix $\hat{T}$ where the blocks along diagonals are equal

$$\hat{T}_{i,j}=\hat{T}_{i+k,j+k}\,,$$

where $\hat{T}_{i,j}$ denotes a block elements. Using Szegö's theorem, we can write densities

$$\lim_{n\to\infty}\operatorname{Tr}\frac{f(\hat{T}_n)}{n}$$

of a matrix function f and the finite projections $\hat{T}_n = P_n \hat{T} P_n$ in terms of a generating function $T(\theta)$. The Fourier coefficients of this $T(\theta)$ are the elements on the diagonals of $\hat{T}$. We will now formulate this more precisely.

Let $T : [-\pi, \pi[\to \mathcal{M}_d$ be an essentially bounded measurable matrix-valued function on the circle and denote its Fourier coefficients by

$$\hat{T}(k) := \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \, T(\theta) \, \mathrm{e}^{-ik\theta} \in \mathcal{M}_d.$$

A function T is essentially bounded if there exist a constant M such that $|T(\theta)| \leq M$ almost everywhere. The operator

$$\hat{T} = \begin{pmatrix} \hat{T}(0) & \hat{T}(1) & \hat{T}(2) & \dots \\ \hat{T}(-1) & \hat{T}(0) & \hat{T}(1) & \dots \\ \hat{T}(-2) & \hat{T}(-1) & \hat{T}(0) & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

defined on $\ell^0(\mathbb{N}) \otimes \mathbb{C}^d$ extends to a bounded linear transformation of $\ell^2(\mathbb{N}) \otimes \mathbb{C}^d$. Operators of this type are block Toeplitz matrices and one has

$$|\hat{T}| = |T|_\infty = \operatorname*{ess\,sup}_\theta |T(\theta)| \,,$$

where the essential supremum of T is the infimum of all constants M that bound $|T(\theta)|$ almost everywhere.

The Toeplitz matrices we are interested in are symbols and hence self-conjugate. For such Toeplitz matrices, we have that $\hat{T}^* = \hat{T}$ if and only if the function T takes values in the Hermitian matrices.

Szegö's theorem

An extension of Szegö's theorem to block Toeplitz matrices characterizes the limiting spectrum of principal submatrices $P_n \hat{T} P_n$ in terms of the generating function T, see [47]. Here P_n projects on the first n blocks in $\ell^2(\mathbb{N}) \otimes \mathbb{C}^d$. We obtain here a more general characterization of such limiting submatrices.

Let us denote for a simply connected compact subset $\mathcal{K}$ of $\mathbb{C}$ by $\mathcal{H}(\mathcal{K})$ the set of continuous functions $f : \mathcal{K} \to \mathbb{C}$ that are holomorphic in the interior $\overset{\circ}{\mathcal{K}}$ of $\mathcal{K}$. Mergelyan's theorem [56] asserts that the complex polynomials in the indeterminate z are dense in $\mathcal{H}(\mathcal{K})$: for any $f \in \mathcal{H}(\mathcal{K})$ and $\epsilon > 0$ there exists a polynomial p^ϵ such that

$$\max_{z \in \mathcal{K}} \big| f(z) - p^\epsilon(z) \big| \leq \epsilon.$$

Finally, let us denote by $\mathbb{E}_n$ the conditional expectation from $\mathcal{B}\big(\ell^2(\mathbb{N})\big)\otimes\mathcal{M}_d \to \mathcal{M}_d$ which traces out the first n blocks

$$\mathbb{E}_n(X) := \frac{1}{n}\sum_{j=0}^{n-1} X_{jj} \in \mathcal{M}_d.$$

We get the following generalization of Szegö's theorem [27].

Theorem 3. *Let $\{T_1, T_2, \ldots, T_k\} \subset \mathcal{L}^\infty\big([-\pi,\pi[,\mathcal{M}_d\big)$ be such that every $T_j(\theta)$ is θ-a.e. diagonalizable, let $f_j \in \mathcal{H}\big(\{z\in\mathbb{C} \mid |z| \le |T_j|_\infty\}\big)$ for $j = 1, 2, \ldots, k$ and let $A_j \in \mathcal{M}_d$, $j = 1, 2, \ldots, k+1$, then*

$$\begin{aligned}\lim_{n\to\infty} \mathbb{E}_n\Big((\mathbb{1}\otimes A_1)\, f_1\big(P_n\hat{T}_1P_n\big)\,(\mathbb{1}\otimes A_2)\,\cdots f_k\big(P_n\hat{T}_kP_n\big)\,(\mathbb{1}\otimes A_{k+1})\Big)\\ = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta\, A_1\, f_1(T_1(\theta))\, A_2 \cdots f_k(T_k(\theta))\, A_{k+1}.\end{aligned} \tag{6.22}$$

Proof. The proof relies on a continuity argument combined with a standard counting argument. First remark that given $\epsilon > 0$ every f_j can be approximated by a suitable complex polynomial p_j^ϵ

$$\max_{|z|\le|T_j|_\infty} \big|f_j(z) - p_j^\epsilon(z)\big| \le \epsilon.$$

Next, as $|T_j(\theta)| \le |T_j|_\infty$ a.e., we can use von Neumann's inequality [28] to get

$$|f_j(T_j(\theta))| \le \max_{|z|\le|T_j|_\infty} |f_j(z)| \text{ and} \tag{6.23}$$

$$\begin{aligned}\big\|f_j(T_j(\theta)) - p_j^\epsilon(T_j(\theta))\big\| &= \big\|\big(f_j - p_j^\epsilon\big)(T_j(\theta))\big\|\\ &\le \max_{|z|\le|T_j(\theta)|}\big(f_j - p_j^\epsilon\big)(z) \le \max_{|z|\le|T_j|_\infty}\big(f_j - p_j^\epsilon\big)(z) \le \epsilon.\end{aligned} \tag{6.24}$$

These estimates allow to replace the f_j in (6.22) by polynomials. It then remains to verify the statement for monomials, but this reduces to a standard counting argument.

In the case where there is only one function $f(X) = X^k$ and $A_j = \mathbb{1}$, the density limit can be worked out as follows:

$$\lim_{n\to\infty}\frac{1}{n}\operatorname{Tr}(P_n\hat{T}P_n)^k = \lim_{n\to\infty}\frac{1}{n}\sum_{i_1,\ldots,i_k=0}^{n} \hat{T}_{i_1,i_2}\hat{T}_{i_2,i_3}\ldots\hat{T}_{i_k,i_1} \tag{6.25}$$

$$= \lim_{n\to\infty}\frac{1}{n}\sum_{i_1,\ldots,i_k=0}^{n} \hat{T}(i_2-i_1)\hat{T}(i_3-i_2)\ldots\hat{T}(i_k-i_1) \tag{6.26}$$

By substituting $v_1 = i_2 - i_1, \ldots, v_{k-1} = i_k - i_{k-1}$, this sum becomes:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{v_1,\ldots,v_{k-1}=-n}^{n} \sum_{i_1 \in \mathcal{S}_n(v_1,\ldots,v_{k-1})} \hat{T}(v_1)\ldots\hat{T}(v_{k-1})\hat{T}(-v_1-\ldots-v_{k-1}),$$

where $\mathcal{S}_n(v_1,\ldots,v_{k-1})$ is the set of indices i such that $v_1+i, v_1+v_2+i, \ldots, v_1+\ldots v_{k-1}+i \in [0,n]$. The number of elements in this set increases by exactly one when n goes to $n+1$, so in the limit we get

$$\lim_{n\to\infty} \sum_{v_1,\ldots,v_{k-1}=-\infty}^{\infty} \hat{T}(v_1)\ldots\hat{T}(v_{k-1})\hat{T}(-v_1-\ldots-v_{k-1}).$$

This is exactly the zeroth Fourier coefficient of $T(\theta)^k$, so we get that the density equals

$$\frac{1}{2\pi}\int_{-pi}^{\pi} d\theta f(T(\theta)).$$

The general case of the theorem can be worked out in a similar manner. □

To deal with entropy we don't need the full amalgamated extension of Theorem 3 of Szegö's theorem but we may restrict ourselves for an Hermitian T to the asymptotic eigenvalue distribution of the principal blocks $P_n\hat{T}P_n$. Taking the trace of (6.22) with a single f and all $A_j = \mathbb{1}$ we recover the result [47]. We denote by $\inf(T)$ and $\sup(T)$ the largest and smallest real numbers such that

$$\inf(T) \leq T \leq \sup(T) \text{ a.e.} \tag{6.27}$$

The increasingly ordered eigenvalues $\big(\tau_1(\theta), \tau_2(\theta), \ldots, \tau_d(\theta)\big)$ of $T(\theta)$ are measurable functions of θ that satisfy

$$\inf(T) \leq \tau_1(\theta) \leq \cdots \leq \tau_d(\theta) \leq \sup(T). \tag{6.28}$$

The eigenvalue distribution of $P_n\hat{T}P_n$ is the atomic probability measure

$$\delta_n = \frac{1}{nd} \sum_{\lambda\in\sigma(P_n\hat{T}P_n)} \delta_\lambda. \tag{6.29}$$

Theorem 4. *With the assumptions of above*

$$\mathrm{w}^*\text{-}\lim_{n\to\infty} \delta_n = \delta_\infty, \tag{6.30}$$

where

$$\delta_\infty\big(]-\infty,t]\big) = \frac{1}{d}\sum_{k=1}^{d} \frac{1}{2\pi} \int_{\tau_k(\theta)\leq t} d\theta. \tag{6.31}$$

An equivalent way to express this result is saying that for any continuous complex function f on $[\inf(T), \sup(T)]$

$$\lim_{n\to\infty} \frac{1}{nd} \operatorname{Tr} f(P_n \hat{T} P_n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \, \frac{1}{d} \operatorname{Tr} f(T(\theta)). \tag{6.32}$$

This version is in some sense more natural as it doesn't involve the reordering of the eigenvalue functions τ_k used in the definition of the distribution function of the limiting eigenvalue distribution (6.31).

We can apply Theorem 4 to the computation of the entropy, replacing the Toeplitz operator T by Q_∞ in Proposition 5 and choosing in (6.32)

$$f(\lambda) = -\lambda \log(\lambda) - (1-\lambda)\log(1-\lambda) \text{ on } (0,1). \tag{6.33}$$

The generating function T becomes

$$\begin{aligned} \theta \mapsto Q &+ (Q - B + X) A \mathrm{e}^{\imath\theta} (\mathbb{1} - A\mathrm{e}^{\imath\theta})^{-1} \\ &+ A^* \mathrm{e}^{-\imath\theta} (\mathbb{1} - A^* \mathrm{e}^{-\imath\theta})^{-1} (Q - B + X^*). \end{aligned}$$

6.3.2 Entropy rate approach

The second approach expresses the entropy as an asymptotic rate. Let ω be a translation invariant state on a quantum spin chain $\otimes^{\mathbb{Z}} \mathcal{M}_d$ and denote by $\rho_{(0,n-1)}$ its reduced density matrices, i.e.

$$\omega(X) = \operatorname{Tr}\big(\rho_{(0,n-1)} X\big) \text{ for } X \in \otimes_{k=0}^{n-1} \mathcal{M}_d. \tag{6.34}$$

As we have seen before, subadditivity combined with translation invariance guarantee the existence of the mean entropy of ω for intervals

$$\mathsf{s}(\omega) = \lim_{n\to\infty} \frac{1}{n} \mathsf{S}(\rho_{(0,n-1)}). \tag{6.35}$$

Moreover, strong subadditivity in conjunction with translation invariance also guarantees that

$$n \mapsto \mathsf{S}(\rho_{(0,n-1)}) \text{ is monotonically increasing and} \tag{6.36}$$

$$\mathsf{s}(\omega) = \lim_{n\to\infty} \frac{1}{n} \mathsf{S}(\rho_{(0,n-1)}) = \lim_{n\to\infty} \Big(\mathsf{S}(\rho_{(0,n)}) - \mathsf{S}(\rho_{(0,n-1)}) \Big). \tag{6.37}$$

Both properties (6.36) and (6.37) fail for general quantum states or for general finite local regions [35]. These results for quantum spin chains extend to Fermionic lattices using the natural embeddings (6.1) and restricting to even states [4]. The

equality of both limits in (6.37) can be seen as a discrete version of de l'Hôpital's rule. Obviously, the existence of the limit of the differences is a much stronger requirement than that of the averages.

For free Fermionic states we can work at the level of symbols. E.g., strong subadditivity of entropy amounts to

$$\mathsf{S}(Q_{123}) + \mathsf{S}(Q_2) \leq \mathsf{S}(Q_{12}) + \mathsf{S}(Q_{23}) \tag{6.38}$$

where S is defined in (4.6) and where the symbols in the inequality are as follows

$$Q_{123} = \begin{bmatrix} Q_1 & T & S \\ T^* & Q_2 & R \\ S^* & R^* & Q_3 \end{bmatrix}, \; Q_{12} = \begin{bmatrix} Q_1 & T \\ T^* & Q_2 \end{bmatrix}, \text{ and } Q_{23} = \begin{bmatrix} Q_2 & R \\ R^* & Q_3 \end{bmatrix}. \tag{6.39}$$

Below, we extend the equality of the limit of differences with that of averages, as in (6.37), to a much wider class of functions than the entropy of a symbol (4.6). The argument relies on regularity of the functions and not on subadditivity or convexity which rarely hold. Szegö's theorem follows as a consequence.

We first show that the theorem holds for polynomials.

Lemma 6. *With the notation and assumptions on an Hermitian Toeplitz operator at the beginning of this section, for any polynomial p*

$$\lim_{n\to\infty} \Big(\operatorname{Tr} p(P_n \hat{T} P_n) - \operatorname{Tr} p(P_{n-1} \hat{T} P_{n-1}) \Big) = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \, \operatorname{Tr} p(T(\theta)). \tag{6.40}$$

Proof. It suffices to consider $p(\lambda) = \lambda^k$ for $k \in \mathbb{N}$. We have

$$\begin{aligned}
&\lim_{n\to\infty} \operatorname{Tr}(P_n \hat{T} P_n)^k - \operatorname{Tr}(P_{n-1} \hat{T} P_{n-1})^k \\
&\quad = \lim_{n\to\infty} \sum_{i=1}^{n} \operatorname{Tr}\big((P_n \hat{T} P_n)^k\big)_{ii} - \sum_{i=1}^{n-1} \operatorname{Tr}\big((P_{n-1} \hat{T} P_{n-1})^k\big)_{ii} \\
&\quad = \lim_{n\to\infty} \Big(\sum_{i_1,\dots,i_k=1}^{n} - \sum_{i_1,\dots,i_k=1}^{n-1} \Big) \\
&\qquad\qquad \operatorname{Tr}\Big\{ (P_n \hat{T} P_n)_{i_1 i_2} \cdots (P_n \hat{T} P_n)_{i_{k-1} i_k} (P_n \hat{T} P_n)_{i_k i_1} \Big\} \\
&\quad = \lim_{n\to\infty} \Big(\sum_{i_1,\dots,i_k=1}^{n} - \sum_{i_1,\dots,i_k=1}^{n-1} \Big) \hat{T}(i_2 - i_1) \dots \hat{T}(i_1 - i_k) \,,
\end{aligned}$$

where $(Q)_{ij}$ denotes the block at position (i, j) inside of a block matrix Q.

By substituting $v_1 = i_2 - i_1, \ldots, v_{k-1} = i_k - i_{k-1}$, this sum becomes:

$$\lim_{n\to\infty} \sum_{v_1,\ldots,v_{k-1}=-n}^{n} \Big(\sum_{i_1\in\mathcal{S}_n(v_1,\ldots,v_{k-1})} - \sum_{i_1\in\mathcal{S}_{n-1}(v_1,\ldots,v_{k-1})} \Big) \hat{T}(v_1)\ldots\hat{T}(v_{k-1})\hat{T}(-v_1-\ldots-v_{k-1}),$$

where $\mathcal{S}_n(v_1,\ldots,v_{k-1})$ is the set of indices i such that $v_1+i, v_1+v_2+i, \ldots, v_1+\ldots v_{k-1}+i \in [0,n]$. For fixed $v_1,\ldots,v_{k-1}$, the number of elements in these sets increases by exactly one when n goes to $n+1$. Hence, the difference of sums between brackets equals one and we arrive at the expression prescribed by the lemma. □

We can now use this lemma and an approximation argument to prove the general case.

Theorem 5. *With the notation and assumptions on an Hermitian Toeplitz operator at the beginning of this section, for any function f that is absolutely continuous on the interval* $[\inf(T), \sup(T)]$

$$\lim_{n\to\infty} \Big(\operatorname{Tr} f(P_{n+1}\hat{T}P_{n+1}) - \operatorname{Tr} f(P_n\hat{T}P_n) \Big) = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta \ \operatorname{Tr} f(T(\theta)). \tag{6.41}$$

Proof. By the continuity of the eigenvalues of a matrix and by the minimax principle [12] we can label the eigenvalues of $P_n\hat{T}P_n$ as

$$\{\tau^n_{k\,j} \mid k = 1,2,\ldots,d,\ j = 1,2,\ldots,n\} \text{ with}$$

$$\inf(T) \le \tau^n_{1\,j} \le \tau^n_{2\,j} \le \cdots \le \tau^n_{d\,j} \le \sup(T) \text{ and } \tau^{n+1}_{k\,j} \le \tau^n_{k\,j} \le \tau^{n+1}_{k\,j+1}. \tag{6.42}$$

See [31] for a proof of this interlacing property.

Let $f : [\inf(T), \sup(T)] \to \mathbb{C}$ be absolutely continuous with integrable derivative g, then for any $\lambda, \tau \in [\inf(T), \sup(T)]$

$$f(\lambda) = f(\tau) + \int_{\tau}^{\lambda} dx\, g(x). \tag{6.43}$$

Therefore

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta\, f(\tau(\theta)) = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta \Big\{ f(\tau) + \int_{\tau}^{\tau(\theta)} dx\, g(x) \Big\} \tag{6.44}$$

$$= f(\tau) + \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta \int_{\tau}^{\tau(\theta)} dx\, g(x) \tag{6.45}$$

$$= f(\tau) + \int_{\inf(T)}^{\sup(T)} dx\, g(x)\, \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta\, \eta(\tau, x, \theta). \tag{6.46}$$

Here, η is defined as

$$\eta(\tau, x, \theta) = \begin{cases} 1 & \tau < x < \tau(\theta) \\ -1 & \tau(\theta) < x < \tau \\ 0 & \text{otherwise} \end{cases}. \tag{6.47}$$

By (6.46) we rewrite the increment of traces of $f(P_n \hat{T} P_n)$ as

$$\operatorname{Tr} f(P_{n+1}\hat{T}P_{n+1}) - \operatorname{Tr} f(P_n \hat{T} P_n) \tag{6.48}$$

$$= \sum_{k=1}^{d} \Big\{ \sum_{j=1}^{n+1} f(\tau_{k\,j}^{n+1}) - \sum_{j=1}^{n} f(\tau_{k\,j}^{n}) \Big\} \tag{6.49}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta \sum_{k=1}^{d} f(\tau_k(\theta)) - \sum_{k=1}^{d} \int_{\inf(T)}^{\sup(T)} dx\, g(x)\, \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta$$

$$\Big\{ \sum_{j=1}^{n+1} \eta(\tau_{k\,j}^{n+1}, x, \tau_k(\theta)) - \sum_{j=1}^{n} \eta(\tau_{k\,j}^{n}, x, \tau_k(\theta)) \Big\} \tag{6.50}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta\, \operatorname{Tr} f(\hat{T}(\theta)) - \sum_{k=1}^{d} \int_{\inf(T)}^{\sup(T)} dx\, g(x)\, \frac{1}{2\pi} \int_{-\pi}^{\pi} d\theta\, h_k^n(x, \theta), \tag{6.51}$$

with

$$h_k^n(x, \theta) = \sum_{j=1}^{n+1} \eta(\tau_{k\,j}^{n+1}, x, \tau_k(\theta)) - \sum_{j=1}^{n} \eta(\tau_{k\,j}^{n}, x, \tau_k(\theta)). \tag{6.52}$$

The functions h_k^n are piecewise constant with values -1, 0 or 1 due to the interlacement (6.42) of the $\tau_{k\,j}^n$. As any integrable g on $[\inf(T), \sup(T)]$ can be arbitrarily well approximated in $\mathcal{L}^1$-norm by polynomials, the theorem follows from Lemma 6. □

6.4 Conclusion

We have studied a free Fermionic version of the quantum Markov processes from Chapter 3. The nature of the free Fermionic states allows us to fully characterize all possible Markov processes that one can construct. The density matrices of these states can be fully described by a Toeplitz matrix. By studying the behaviour of the eigenvalues of subsequent Toeplitz matrices, we have proved a new Szegö theorem that allows us to calculate the asymptotic entropy rate. This is what corresponds in the free Fermionic case to the method proposed by Blackwell [8].

It would be interesting to look for other quantum Markov processes for which an explicit calculation of the entropy rate is possible. Processes with a high symmetry are obvious first choices. The process with SU(2) symmetry presented in Chapter 3 is a good candidate. Hopefully, such calculation can lead to a quantum version of the Blackwell dynamical system.

Chapter 7

Conclusion

The main goal of this thesis is to generalize an existing approach of calculating entropy densities as rates to quantum systems. Along the way we have from time to time strayed off into related topics, such as determining dependence of the entropy density on a parameter in Chapter 2 or the use of conditional state spaces to study correlations in quantum systems. We will conclude by reviewing the results discussed in the text and proposing some possible future research.

First of all, one may wonder why it is interesting to calculate the entropy density of quantum states. This question we tried to tackle in the introduction. We have explored two major topics in quantum information science where entropies play a crucial role, namely the determination of entanglement in pure states and the information capacity of a quantum channel. Besides these topics one can think of applications in statistical physics, but they have not been addressed here.

We then introduced in Chapter 2 the method of calculating entropy densities as rates for classical hidden Markov processes in the context of channel capacities. The cornerstone of the entropy rate method is the following relation for translation-invariant processes:

$$\lim_{n\to\infty} \frac{\mathsf{S}_n}{n} = \lim_{n\to\infty} (\mathsf{S}_n - \mathsf{S}_{n-1})\,. \tag{7.1}$$

One can think of this relation as a discrete de l'Hôpital's rule, where the derivative has been replaced by a difference. Using the structure of probabilities of hidden Markov processes, the entropy rate calculation simplifies into an entropy of an invariant measure of a dynamical system.

This dynamical system was worked in detail for a so-called depolarizing channel with memory. It turned out that the invariant measure is too complex to fully characterize due to its fractal nature. By employing some numerical tricks to

approximate the invariant measure, it is however possible to calculate the channel capacity efficiently. Of course, it is not the invariant measure itself that is really of interest here, but averaged quantities of this measure and their dependence on the parameters in the problem. We have shown that it is possible to say something about this general behaviour by proving the increase of channel capacity as the noise correlations between subsequent uses becomes stronger.

As the goal of this work is to generalize an entropy calculation method that uses the specific structure of classical hidden Markov processes, it is necessary to specify what we consider to be hidden Markov processes in the quantum case. This was reviewed in Chapter 3. The crucial property proves to be the fact that the correlations between two semi-infinite halves of the chain remains in a sense finite. This is a property that can be compared to the standard definition of a Markov process as depending on the past through only one past configuration.

This is where we side-tracked to have a closer look at the way correlations are described in the definition of these finitely correlated states. This leads to the definition of conditional state spaces. We have then tried to determine the usefulness of this concept in characterizing correlations in general quantum states. From the applications we have worked out here, one can conclude that the dimension and the geometry of the conditional state space is an indicator of the strength and nature of the correlations in the state under study.

We then introduced our main object of study for the remainder of the text, namely the free Fermionic states. These are states with a simple structure that fulfil the Fermionic anti-commutation relations. Due to this structure, calculations on these states remain manageable. As we want to look at correlations between subsystems and construct processes consisting of many subsystems, we have explored briefly what we mean by a subsystem in a Fermionic setting.

The first application worked out for these free Fermionic states is the characterization of the conditional state spaces for such states. This was done in Chapter 5. We highlighted some analogies with the discussion of conditional state spaces for distinguishable particles of Chapter 3 and gave some applications of this characterization.

By Chapter 6, we are then finally ready to do the entropy rate calculation for a quantum system, the free Fermions. First of all, we used the construction of Chapter 3 to construct a Fermionic Markov chain. It turns out that there is a direct way to calculate the entropy density in this case, allowing us to verify the result of the entropy rate calculation. It is in this case possible to do the entropy rate calculation by studying the behaviour of eigenvalues of large Toeplitz matrices. As a by-product we proved equality 7.1 a large class of functions without using strong sub-additivity.

There remain a number of open questions concerning the topics presented in this

thesis. For example, in Chapter 2 we have seen that the channel capacity of a quantum channel with classical memory was related to the entropy density of a classical hidden Markov process. It would be interesting to investigate whether the entropy density calculation we have carried out for quantum Markov chains can also be used to determine the channel capacity of memory channels. Also concerning conditional state spaces more research is warranted. Although we have found compelling evidence that support the use of these spaces as a means to study quantum correlations, it is as of yet unclear whether the example showing entanglement in Werner states can be extended to other quantum states. Finally, the entropy rate calculation should be extended to other finitely correlated states. Other states with high symmetry, such as Bosonic Gaussian states or SU(2) invariant states are obvious candidates.

List of publications

Journal articles

"Classical capacity of a qubit depolarizing channel with memory"
J. Wouters, I. Akhalwaya, M. Fannes, and F. Petruccione
Phys. Rev. A 79, 042303; 2009
doi: 10.1103/PhysRevA.79.042303; arXiv: 0901.2516v1

"Correlations in free Fermionic States"
M. Fannes and J. Wouters
J. Phys. A: Math. Theor. 42 465308; 2009
doi: 10.1088/1751-8113/42/46/465308; arXiv: 0906.3196v1

"Free Fermionic quantum Markov chains"
M. Fannes and J. Wouters
In preparation

Proceedings

"Non-rigid brain image registration using a statistical deformation model"
J. Wouters, E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens
Proc. SPIE Vol. 6144, 614411; 2006
doi: 10.1117/12.653081;

"Non-rigid image registration using mutual information"
F. Maes, E. D'Agostino, D. Loeckx, J. Wouters, D. Vandermeulen, and P. Suetens
Compstat 2006 - Proceedings in Computational Statistics; 2006
doi: 10.1007/978-3-7908-1709-6˙8;

"The algebraic measure of a hidden Markov quantum memory channel"
I. Akhalwaya, J. Wouters, M. Fannes, and F. Petruccione
AIP Conf. Proc. Vol. 1110, pp. 127-130; 2009
doi: 10.1063/1.3131288; arXiv: 0901.2528v1

"Quantum processes"
M. Fannes and J. Wouters
Quantum Dynamics and Information; Proc. 46th Karpacz Winter School of Theoretical Physics, World Scientific; 2010
ISBN: 978-981-4317-43-6; arXiv: 1005.3177

Bibliography

[1] I. Akhalwaya, J. Wouters, M. Fannes, and F. Petruccione. The algebraic measure of a hidden Markov quantum memory channel. *AIP Conference Proceedings*, 1110(1):127–130, apr 2009.

[2] R. Alicki and M. Fannes. *Quantum dynamical systems*. Oxford University Press, Oxford, 2001.

[3] R. Alicki, M. Fannes, and M. Pogorzelska. Quantum generalized subsystems. *Physical Review A*, 79(5):052111, 2009.

[4] H. Araki and H. Moriya. Equilibrium statistical mechanics of Fermion lattice systems. *Reviews in Mathematical Physics*, 15:93–198, 2003.

[5] E. Balslev, J. Manuceau, and A. Verbeure. Representations of anticommutation relations and Bogolioubov transformations. *Communications in Mathematical Physics*, 8(4):315–326, 1968.

[6] C. H. Bennett. Notes on Landauer's principle, reversible computation, and Maxwell's demon. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 34(3):501–510, sep 2003.

[7] H. Bethe. Zur Theorie der Metalle i. Eigenwerte und Eigenfunktionen der linearen Atomkete. *Zeitschrift für Physik*, 71:205–226, 1931.

[8] D. Blackwell. The entropy of functions of finite state Markov chains. In *Trans. First Prague Conference on Information Theory, Decision Functions, and Random Processes, Prague*, pages 13–20, 1957.

[9] S. Bose. Quantum communication through an unmodulated spin chain. *Physical Review Letters*, 91(20):207901, 2003.

[10] O. Bratteli and D.W. Robinson. *Operator Algebras and Quantum Statistical Mechanics: Equilibrium States. Models in Quantum Statistical Mechanics*. Springer, Berlin, 1997.

[11] M.-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, jun 1975.

[12] R. Courant and D. Hilbert. *Methods of Mathematical Physics*. Wiley-VCH, first edition, 1989.

[13] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience, August 1991.

[14] A.C. (A.S.) Холево (Holevo). Квантовые системы, каналы, информация *(Kvantovye sistemy, kanaly, informacija)*. МЦНМО (MCNMO), 2010.

[15] A. D'Arrigo, G. Benenti, and G. Falci. Quantum capacity of dephasing channels with memory. *New Journal of Physics*, 9(9):310, 2007.

[16] N. Datta and T.C. Dorlas. The coding theorem for a class of quantum channels with long-term memory. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8147, 2007.

[17] N. Datta and T.C. Dorlas. Classical capacity of quantum channels with general Markovian correlated noise. *Journal of Statistical Physics*, 134(5-6):1173, 2009.

[18] J. Dereziński. Introduction to representations of the canonical commutation and anticommutation relations. In Jan Derezinski and Heinz Siedentop, editors, *Large Coulomb Systems*, volume 695, pages 63–143. Springer Berlin / Heidelberg, 2006.

[19] B. Dierckx, M. Fannes, and M. Pogorzelska. Fermionic quasifree states and maps in information theory. *Journal of Mathematical Physics*, 49(3):032109, 2008.

[20] B. G. Englert. Elements of micromaser physics. *Arxiv preprint quant-ph/0203052*, 2002.

[21] M. Fannes, B. Nachtergaele, and L. Slegers. Functions of Markov processes and algebraic measures. *Reviews in Mathematical Physics*, 4(1):39, 1992.

[22] M. Fannes, B. Nachtergaele, and R.F. Werner. Finitely correlated states on quantum spin chains. *Communications in Mathematical Physics*, 144(3):443, 1992.

[23] M. Fannes and R.F. Werner. Boundary conditions for quantum lattice systems. *Helv. Phys. Acta*, 7-8(68):635–657, 1995.

[24] M. Fannes and J. Wouters. Correlations in free Fermionic states. *Journal of Physics A: Mathematical and Theoretical*, 42(46):465308, 2009.

[25] M. Fannes and J. Wouters. Quantum processes. In *Quantum Dynamics and Information*. World Scientific, 2010.

[26] B. Gaasbeek. An introductory course on quantum mechanics. *Arxiv preprint quant-ph/1007.4184*, 2010.

[27] U. Grenander and G. Szegő. *Toeplitz Forms and Their Applications*. AMS Bookstore, 2001.

[28] P.R. Halmos. *A Hilbert space problem book*. Springer-Verlag, New York, 1974.

[29] P. Hayden, R. Jozsa, D. Petz, and A. Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics*, 246(2):359, 2004.

[30] A.S. Holevo. The capacity of quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44:269, 1998.

[31] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

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

[33] M. Karbach, K. Hu, and G. Müller. Introduction to the Bethe Ansatz ii. *Computers in Physics*, 12(6):565–573, 1998.

[34] M. Karbach and G. Müller. Introduction to the Bethe Ansatz i. *Comput. Phys.*, 11(1):36–43, 1997.

[35] A.R. Kay and B.S. Kay. Monotonicity with volume of entropy and of mean entropy for translationally invariant systems as consequences of strong subadditivity. *Journal of Physics A*, 34(3):365, 2001.

[36] C. King. The capacity of the quantum depolarizing channel. *IEEE Transactions on Information Theory*, 49(1):221, 2003.

[37] C. King and M.B. Ruskai. Minimal entropy of states emerging from noisy quantum channels. *Information Theory, IEEE Transactions on*, 47(1):192–209, 2001.

[38] D. Kretschmann and R.F. Werner. Tema con variazioni: quantum channel capacity. *New Journal of Physics*, 6:26, 2004.

[39] D. Kretschmann and R.F. Werner. Quantum channels with memory, 2005.

[40] R. Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5:183–191, 1961.

[41] N.P. Landsman. Lecture notes on C*-algebras, Hilbert C*-modules, and quantum mechanics, 1998.

[42] E.H. Lieb. Proof of the strong subadditivity of quantum-mechanical entropy. *Journal of Mathematical Physics*, 14(12):1938, 1973.

[43] F. Maes, E. D'Agostino, D. Loeckx, J. Wouters, D. Vandermeulen, and P. Suetens. Non-rigid image registration using mutual information. In Alfredo Rizzi and Maurizio Vichi, editors, *Compstat 2006 - Proceedings in Computational Statistics*, pages 91–103. Physica-Verlag HD, 2006.

[44] F. Maes, D. Vandermeulen, and P. Suetens. Medical image registration using mutual information. *Proceedings of the IEEE*, 91(10):1699–1722, 2003.

[45] P.A. Martin and F. Rothen. *Problèmes à N-corps et champs quantiques*. Presses polytechniques et universitaires romandes, 1990.

[46] S. Michalakis and B. Nachtergaele. Entanglement in finitely correlated spin states. *Physical review letters*, 97(14):140601, 2006.

[47] M. Miranda and P. Tilli. Asymptotic spectra of hermitian block Toeplitz matrices and preconditioning results. *SIAM J. Matrix Anal. Appl.*, 21(3):867–881, 2000.

[48] M. Mushkin and I. Bar-David. Capacity and coding for the Gilbert-Elliot channels. *IEEE Transactions on Information Theory*, 35(6):1277–1290, 1989.

[49] A. Peres. *Quantum Theory: Concepts and Methods*. Springer, 1 edition, sep 1995.

[50] D. Perez-Garcia, F. Verstraete, M.M. Wolf, and J. I. Cirac. Matrix product state representations. *Arxiv preprint quant-ph/0608197*, 2006.

[51] D. Petz. *Quantum Information Theory and Quantum Statistics*. Springer, Berlin, Heidelberg, 2008.

[52] M.B. Plenio and S. Virmani. An introduction to entanglement measures. *Arxiv preprint quant-ph/0504163*, apr 2005. Quant. Inf. Comp. 7, 1 (2007).

[53] R.T. Powers and E. Størmer. Free states of the canonical anticommutation relations. *Communications in Mathematical Physics*, 16(1):1–33, 1970.

[54] F. Rocca, M. Sirugue, and D. Testard. On a class of equilibrium states under the Kubo-Martin-Schwinger boundary condition. *Communications in Mathematical Physics*, 13(4):317, 1969.

[55] R.T. Rockafellar. *Convex Analysis*. Princeton University Press, jun 1970.

[56] W. Rudin. *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition, 1987.

[57] A. Sanpera, R. Tarrach, and G. Vidal. Local description of quantum inseparability. *Phys. Rev. A*, 58(2):826–830, Aug 1998.

[58] U. Schollwöck. The density-matrix renormalization group. *Reviews of Modern Physics*, 77(1):259, apr 2005.

[59] B. Schumacher and M. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131, Jul 1997.

[60] D. Shale and W. F. Stinespring. States of the Clifford algebra. *The Annals of Mathematics*, 80(2):365–381, 1964.

[61] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[62] W. Słomczynski, J. Kwapien, and K. Zyczkowski. Entropy computing via integration over fractal measures. *Chaos*, 10(1):180–188, 2000.

[63] G. Smith. Quantum channel capacities. *1007.2855*, jul 2010.

[64] M. Takesaki. Conditional expectations in von Neumann algebras. *Journal of Functional Analysis*, 9(3):306–321, 1972.

[65] J. Townsend. *A Modern Approach to Quantum Mechanics*. University Science Books, 2nd edition, April 2000.

[66] A. Wehrl. General properties of entropy. *Rev. Mod. Phys.*, 50(2):221–260, Apr 1978.

[67] R.F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277–4281, 1989.

[68] R.F. Werner. Remarks on a quantum state extension problem. *Letters in Mathematical Physics*, 19(4):319, 1990.

[69] J. Wouters, E. D'Agostino, F. Maes, D. Vandermeulen, and P. Suetens. Non-rigid brain image registration using a statistical deformation model. In *Proceedings of SPIE*, volume 6144, page 614411, 2006.

[70] J. Wouters, M. Fannes, I. Akhalwaya, and F. Petruccione. Classical capacity of a qubit depolarizing channel with memory. *Physical Review A*, 79(4):042303, apr 2009.