PROOF OF THE ORTHOGONAL MEASUREMENT CONJECTURE FOR TWO STATES OF A QUBIT

ANDREAS KEIL

NATIONAL UNIVERSITY OF

SINGAPORE

2009

PROOF OF THE ORTHOGONAL MEASUREMENT CONJECTURE FOR TWO STATES OF A QUBIT

ANDREAS KEIL
(Diplom-Physiker), CAU Kiel

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF

PHILOSOPHY

DEPARTMENT OF PHYSICS

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

I would like to thank everybody who supported me during the time of this thesis. Especially I want to thank my supervisors Lai Choy Heng and Frederick Willeboordse, their continued support was essential. For great discussions I want to thank Syed M. Assad, Alexander Shapeev and Kavan Modi. Special thanks go to Berge Englert and Jun Suzuki, without them this conjecture would still have been dormant.

Thank you!

Contents

Summary

In this thesis we prove the orthogonal measurement hypothesis for two states of a qubit. The accessible information is a key quantity in quantum information and communication. It is defined as the maximum of the mutual information over all positive operator valued measures. It has direct application in the theory of channel capacities and quantum cryptography. The mutual information measures the amount of classical information transmitted from Alice to Bob in the case that Alice either uses classical signals, or quantum states to encode her message and Bob uses detectors to receive the message. In the latter case, Bob can choose among different classes of measurements. If Alice does not send orthogonal pure states and Bobs measurement is fixed, this setup is equivalent to a classical communication channel with noise. A lot of research went into the question which measurement is optimal in the sense that it maximizes the mutual information. The orthogonal measurement hypothesis states that if the encoding alphabet consists of exactly two states, an orthogonal (von Neumann) measurement is sufficient to achieve the accessible information. In this thesis we affirm this conjecture for two pure states of a qubit and give the first proof for two general states of a qubit.

List of Figures

List of Symbols

$p(j\|r)$	conditional probability matrix of a noisy channel	5
ε_0	probability of a zero bit to flip to a one	5
ε_1	probability of a one bit to flip to a zero	5
p_{rj}	classical joint probability matrix	5
var	variance of a random variable	6
ε	relative deviation from the expected value of a sequence	7
$H_2(p)$	binary entropy of p	8
$I(\{p_{rj}\})$	mutual information of a joint probability matrix	12
$p_{\cdot j}$	column marginals of a probability distribution	12
$p_{r\cdot}$	row marginals of a probability distribution	12
$C^{classical}$	classical channel capacity	12
$\mathrm{cov}(X,Y)$	covariance of two probability distributions X,Y	15
ρ	quantum state	19
$\mathcal{H}$	finite dimensional complex Hilbert space	19
Π	positive operator valued measure (POVM)	19
Π_j	outcome of a POVM	19
$\mathbb{I}$	identity operator	20

p_{rj}	joint probability matrix given by quantum states and measurements	21
$I_{acc}(\{\rho_r\})$	accessible information of quantum states	22
χ	Holevo quantity	23
$S(\rho)$	von Neumann entropy of a state ρ	23
δI	first variation of I	33
$\delta_{(k,l)}I$	variation of I in the direction specified by k,l	35
$Q_r(t)$	auxiliary function	36
α_r	auxiliary symbol	36
ξ_r	auxiliary symbol	36
η_r	auxiliary symbol	36
$f_{(\alpha,\xi,\eta)}(t)$	auxiliary function	37
L	auxiliary function	39
Q_s	convex sum of Q_1 and Q_2	39
P	auxiliary polynomial	39
$R[p,q]$	resultant of two polynomials p and q	41
$\Delta[p]$	discriminant of a polynomial p	42
$D[g]$	domain of a family of a polynomial such that its discriminant is non zero	48
$D_0[g]$	subset of D[g] for which the highest coefficient of g vanishes	48
$D_1[g]$	complement of $D_0[g]$ in $D[g]$	48
$\mathbb{R}$	real numbers	49
$[a,b]$	closed interval from a to b	49
$\bar{\mathbb{R}}$	real numbers including plus and minus infinity	49

(a,b)	open interval from a to b	49
$C^1(M,\mathbb{R})$	set of real-valued continuous differentiable functions on the set M	49
$\lvert\cdot\rvert$	number of elements of a set	50
$C^0(M,\mathbb{R})$	set of real-valued continuous functions on the set M	50
X	difference of the auxiliary variables η and ξ^2	58

Chapter 1

Introduction

Mutual information measures the amount of classical information that two parties, Alice and Bob, share. Shannon showed in his seminal paper [1] that there always exists an encoding scheme which transmits an amount of information arbitrarily close to the mutual information per use of the channel. It was also mentioned by Shannon that it is impossible to transmit more information than the mutual information quantifies, only to be proved later [2]. An important extension to this setup is to ask what happens if Alice does not send classical states to Bob, but uses states of a quantum system instead. How much information do Alice and Bob share? This question is at the heart of quantum information and a great amount of research is devoted to it.

There are a number of possibilities to view this question. For instance we can ask how much quantum information do both parties share. Or we can ask how much classical information do Alice and Bob share if they use quantum states and measurements for communication. In this thesis we are interested in the latter question.

Assume Alice encodes a message by sending a specific quantum state ρ_r for each letter in the alphabet of the message. The rth letter in the alphabet occurs with probability $\mathrm{tr}(\rho_r)$ in the message. Bob sets up a measurement apparatus to determine which state was sent, described by a positive operator valued measure (POVM).

Alice and Bob's situation can be described by a joint probability matrix. The mutual information of the joint probability matrix tells us how much classical information on average is transmitted to Bob per transmitted state [1, 3], when Alice and Bob use an appropriate encoding and decoding scheme. If we assume the states to be fixed, Bob can try to maximize the information transmitted by choosing a POVM that maximizes the mutual information. This defines an important quantity; the so called accessible information,

$$I_{\mathrm{acc}} = \max_{\{\Pi_k\}} I(\{\rho_r\}, \{\Pi_k\}), \tag{1.1}$$

where the maximum is taken over all POVMs and I denotes the mutual information. To actually transmit this amount of information, the (Shannon-) encoding scheme has to be adjusted as well.

The question which POVM maximizes the mutual information, was raised by Holevo in 1973 [4], and is in general unanswered and usually addressed numerically [5, 6, 7]. Even the simpler question of how many outcomes are sufficient is unanswered. It has been shown [8] that an orthogonal (von Neumann) measurement, is in general not sufficient. Levitin [9] conjectured in 1995 that if Alice's alphabet consists of n states and n is smaller or equal to the dimension of the underlying Hilbert space, an orthogonal measurement is sufficient. If so, the number of outcomes would be equal to the dimension of the Hilbert space. This conjecture in general does not hold as shown by Shor [10]. A well known class of counter examples, given by states representing the legs of a pyramid, is discussed in detail by Řeháček and Englert [11]. In the same paper Shor reported that Fuchs and Peres affirmed numerically that if the alphabet consists of two states the optimal measurement is an orthogonal measurement. This is the *orthogonal measurement conjecture*. For two pure states it was proved to be true in arbitrary dimensions by Levitin [9].

This conjecture has important experimental and theoretical implications. In an experiment, orthogonal measurements are generally easier to implement than arbitrary generalized measurements. From a theoretical point, knowing the accessible information is crucial to determine the $C^{1,1}$-channel capacity [1] and for security analysis using the Csiszár-Körner theorem [12], for example the thresholds for an incoherent attack on the Singapore protocol [13] are obtained by determining the accessible information. Also part of the security analysis of the BB84 protocol for incoherent attacks relies on this conjecture [14]. Work has been done under the assumption that this conjecture is true [15]. In the sequel we will prove this conjecture for two states of a qubit.

This thesis is organized as follows, in section 1.1 we introduce the mutual information from the physical motivation of how much information can be transmitted. We have another brief look at the mutual information from the point of view of key-sharing of two parties, which is important in the modern view of security analysis. A few well known and essential mathematical properties are derived in this section as well. In the next section, section 1.2, we will introduce the quantum set-up and review some important theorems about the accessible information in this case.

The following section 1.3 is concerned with the variation of the mutual information with respect to the POVM. In the subsequent sections certain crucial features of the derivative of the mutual information are derived which allow us to prove the orthogonal measurement conjecture. In the appendix we will show how the variation equations can be derived by using a Bloch-representation of the states and POVM. Usually the Bloch-representation has advantages in dealing with qubits, but for the problem at hand it is surprisingly not the case.

1.1 Mutual Information

In this thesis mutual information is a fundamental quantity. We start in this chapter with a rather informal introduction to the physical and informational motivation of the mutual information. The results are well known and can be found in any standard textbook, e.g. [3].

The mutual information arises from the question, how much information can be sent through a noisy memoryless channel from A to B. The basic situation is depicted in figure 1.1.

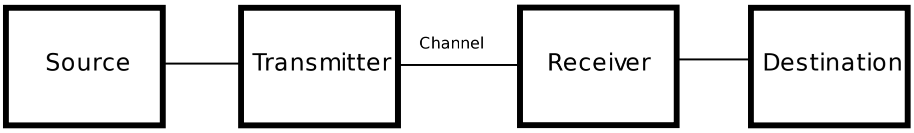

Figure 1.1: Transmitting a message from Alice to Bob through a channel

Considering a binary noisy channel, we have the following situation depicted in figure 1.2

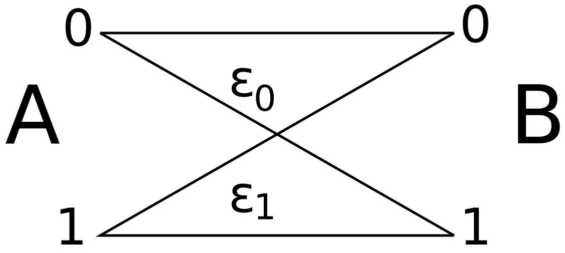

Figure 1.2: Bit-flips in a binary noisy channel

So this channel can be described by the conditional probability matrix

$$p(j|r) = \begin{pmatrix} (1-\varepsilon_0) & \varepsilon_0 \\ \varepsilon_1 & (1-\varepsilon_1) \end{pmatrix},$$

where ε_0 denotes the probability of a zero bit to flip to a one, and ε_1 the probability of the reverse case.

This determines the probability of Bob to receive outcome j under the condition that Alice sent the letter r. A channel is called *symmetric* if ε_0 equals ε_1. If the probabilities of the letters of the source are fixed to p_r we can define the joint probability matrix by

$$p_{rj} = p_r \, p(j|r).$$

To see how much information is emitted, the idea is to look at long strings of letters instead of single letters. Assume the source giving an uncorrelated string of letters with fixed probabilities. Strings of length N will follow a binomial distribution

$$P(r) = \binom{n}{r} \, p_1^r \, p_0^{n-r},$$

where $P(r)$ denotes the probability of having exactly r ones in a string of n characters. For large values of n, $P(r)$ can be approximated by a normal distribution

$$P(r) \approx \frac{1}{2\pi n\, p_0\, p_1} \exp\left(-\frac{(r - n\, p_1)^2}{2\, n\, p_0\, p_1}\right).$$

From the normal distribution we can see that, if n grows large, the distribution peaks sharply around its maximum; implying that a relative small slice contains almost the whole weight of the distribution for n growing large.

Following Shannon in his seminal paper [1] we ask the question, which sequences are typical, i.e. appear with overwhelming probability. For this we split the message into independent blocks with each block of size n. Each block is called a *sequence*. If we assign the values 0 and 1 to each of the letters, we can ask how many different sequences are in a typical block. We are interested in the random variable X,

$$X = \sum_{j=1}^{n} X_j,$$

where each random variable X_j is independent and with probability p_0 gives zero and with p_1 gives one.

We have

$$\langle X \rangle = n\, p_1, \quad \mathrm{var}(X) = \langle (X - \langle X \rangle)^2 \rangle = n\, p_0\, p_1.$$

It is known from Chebyshev's inequality that

$$P\left(|X - \langle X\rangle| \geq n\varepsilon\right) \leq \frac{p_0\, p_1}{n\,\varepsilon^2} =: \delta,$$

with ε being the relative deviation of the number of ones from the expected value. This inequality tells us that for any given, small, deviation ε we can find a (large) length n such that the probability of finding a sequence outside the typical sequences can be made arbitrary small.

So for given δ and given ε we get the minimum length n

$$n = \frac{\delta\varepsilon^2}{p_0\, p_1}$$

of a sequence such that with probability $(1-\delta)$ the number of ones in a sequence only deviates by $n\varepsilon$ from the expected value. The question is how many typical sequences are there for given ε.

The total number of sequences is given by

$$N(\text{total}) = 2^n.$$

The number typical sequences is given by the sum of the possibilities

$$N(\text{typical}) = \sum_{k=n(p_1-\varepsilon)}^{n(p_1+\varepsilon)} \binom{n}{k}$$

which can be estimated, in case $p_1 < (\frac{1}{2} - \varepsilon)$, to lie between the following bounds:

$$2n\varepsilon\binom{n}{(p_1-\varepsilon)\,n} < N(\text{typical}) < 2n\varepsilon\binom{n}{(p_1+\varepsilon)\,n}.$$

If p_1 is greater than $(\frac{1}{2}+\varepsilon)$ we have the same inequality inverted. If p_1 is exactly one-half $N(\text{typical})$ becomes arbitrarily close to $N(\text{total})$. This exhausts all possibilities, since ε can be chosen to be arbitrarily small.

We can use Stirling's series,

$$\log n! = n\log n - n + \frac{1}{2}\log(2\pi n) + O(n^{-1})$$

to approximate the binomial coefficient to get

$$\log_2\binom{n}{p_1\,n} = \frac{1}{\log 2}\left(-n\,p_1\log p_1 - n\,p_0\log p_0 - \frac{1}{2}\log(2\pi\,p_0\,p_1\,n) + O(n^{-1})\right)$$

For large n we can approximate the binomial coefficient by

$$\binom{n}{p_1\,n} \approx 2^{nH_2(p_1)-\frac{1}{2}\log(2\pi\,p_0\,p_1\,n)},$$

where $H_2(p_1)$ denotes the binary entropy of the source, i.e.

$$H_2(p) = -\left(p\log_2 p + (1-p)\log_2(1-p)\right).$$

For convenience we drop the $-\frac{1}{2}\log(2\pi\,p_0\,p_1\,n)$ term, it grows slower than order of n and will not contribute in the final result.

We have

$$2^{nH_2(p_1-\varepsilon)+\log_2(2n\varepsilon)} < N(\text{typical}) < 2^{nH_2(p_1+\varepsilon)+\log_2(2n\varepsilon)},$$

and for small ε we will reach

$$N(\text{typical}) \approx 2^{nH_2(p_1)+\log_2(2n\varepsilon)}.$$

This shows how much information is contained in the source. If we would imagine to enumerate (which is hard to do in practice) all the typical sequences we would need m-bits with

$$m = nH_2(p_1) + \log(2n\varepsilon)$$

to distinctively label the sequences, plus a few codes to signalize non-typical sequences. To determine the amount of information per original bit we need to divide by the total number n of bits in a sequence, which gives

$$C = H_2(p_1) + \frac{\log(2n\varepsilon)}{n} \approx H_2(p_1)$$

for large n. The amount of information is therefore given by the entropy of the source. This is a well established result in information theory.

Since we intend to send this information through our noisy channel we have to consider what happens to our typical sequences. Any typical sequence of Alice becomes, in the overwhelming majority of cases, a typical sequence, or close to one, on Bob's side, with a different probability distribution though.

We would like to know how much of this information can be extracted by Bob. In the case of a noisy channel there is a probability of a one flipping to a zero and vice versa. This means that Alice's typical sequences will be mapped to different typical sequences on Bob's side. In the presence of noise, these sequences on Bob's side overlap and it is not possible for Bob to determine accurately which sequence was send by Alice. The trick is Alice chooses a limited set of codewords which are separated far enough (in the sense of Hamming-distance) such that Bob can (in almost all of the cases) unambiguously determine which codeword was sent. This is illustrated in figure 1.3. To how many possible sequences does a typical sequence of Alice spread?

Let us label the possibility for a bit flip by

$$\varepsilon_0 = p(1|0), \quad \varepsilon_1 = p(0|1).$$

Since Alice has most likely $p_0 \cdot n$ zeros in her sequence, there will be

$$\binom{p_0\, n}{\varepsilon_0\, p_0\, n} \approx 2^{p_0 n H_2(\varepsilon_0)},$$

combinations with flips from zero to one and

$$\binom{p_1\, n}{\varepsilon_1\, p_1\, n} \approx 2^{p_1 n H_2(\varepsilon_1)},$$

flips from one to zero. The total number of combinations is given by the product

$$N(\text{sequences spread}) \approx 2^{n(p_0 H_2(\varepsilon_0)+p_1 H_2(\varepsilon_1))}.$$

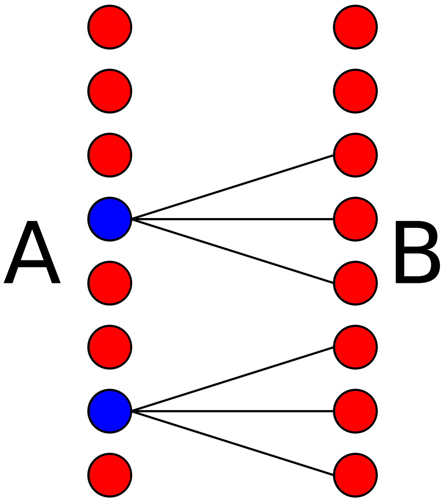

Figure 1.3: Codewords from Alice's side mapped to different codewords on Bob's side due to channel noise; blue color indicating an example set of codewords Alice chooses

The number of typical sequences on Bob's side is then given by

$$\binom{n}{(\varepsilon_0\, p_0 + (1-\varepsilon_1)\, p_1)\, n} \approx 2^{\,n H_2(\varepsilon_0\, p_0 + (1-\varepsilon_1) p_1)},$$

This implies that the number of states Alice can safely choose to transmit to Bob is given by

$$\begin{aligned} N(\text{transmit}) &= \frac{N(\text{typical Bob})}{N(\text{sequences spread})} \approx 2^{n(H_2(\varepsilon_0\, p_0 + (1-\varepsilon_1) p_1) - p_0 H_2(\varepsilon_0) - p_1 H_2(\varepsilon_1))} \\ &= 2^{\,n I(\{p_{rj}\})} \end{aligned}$$

with $I(\{p_{rj}\})$ the *mutual information* of the joint probability distribution

$$p_{rj} = \begin{pmatrix} (1-\varepsilon_0)\, p_0 & \varepsilon_0\, p_0 \\ \varepsilon_1\, p_1 & (1-\varepsilon_1)\, p_1 \end{pmatrix}$$

Explicitly the mutual information is given by

$$I(\{p_{rj}\}) = \sum_{r,j} p_{rj} \log\left(\frac{p_{rj}}{p_{r\cdot}p_{\cdot j}}\right) \tag{1.2}$$

with marginals

$$p_{\cdot j} := \sum_r p_{rj}, \;\; p_{r\cdot} := p_r = \sum_j p_{rj}.$$

So the amount of information transmitted per bit sent is given by the mutual information. This derivation works in more complicated cases with more input and outputs on Alice and Bobs side, and gives the same equation as in equation 1.2 with an adjusted range for the indices.

For a given channel $p(j|r)$, the maximization of the mutual information over all possible probabilities on Alice's side gives the *classical channel capacity*:

$$C^{classical} = \max_{\{p_r\}} I(\{p_r\, p(j|r)\}).$$

It is an interesting question, what can be considered 'mutual' in the mutual information. It is obvious that the definition for the mutual information only depends on the joint probability, it is symmetric if we exchange the roles of Alice and Bob. We will now look at the mutual information from the point of key sharing using a common source, which gives another operational meaning to the mutual information.

Consider the following scenario, depicted in figure 1.4, which is common in security analysis for quantum key distribution. A common source delivers sequences to Alice and Bob. Let's assume that this happens without any eavesdropping. The question we can ask now, is how long a secret key can Alice and Bob create by only using a public channel and not revealing any (useful) information about the key by using the channel.

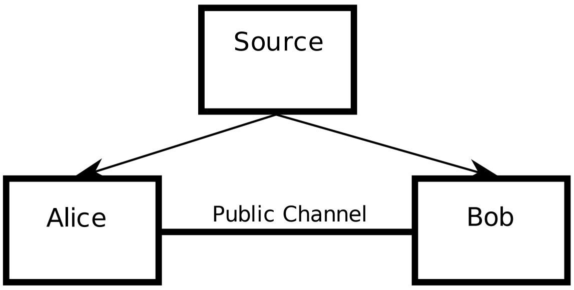

Figure 1.4: A common source for random, correlated data for Alice and Bob

The idea is a small variation to the idea laid out before. Alice and Bob agree on a number of different encoding schemes beforehand. Each typical sequence on Alice's side is part of exactly one encoding scheme, and the number of scheme is equal to the spread due to the noise. Each encoding scheme is chosen to be optimal in the sense of the transmission of signals above. Figure 1.5 shows the situation.

At each time the common source sends a sequence to Alice and Bob, Alice publicly announces into which group it fell on her side. A third party which listens to the public channel can gain no information about the content of Alice and Bob's shared string. This scheme was suggested in [16], and is called *reconciliation*. In the end, Alice and Bob share a common key of the length of the mutual information of the source, but note as outlined some information has to be directly transmitted by classical communication between Alice and Bob to achieve this.

After these physical interpretations of the mutual information we will look at more mathematical properties of the mutual information in the remainder of this

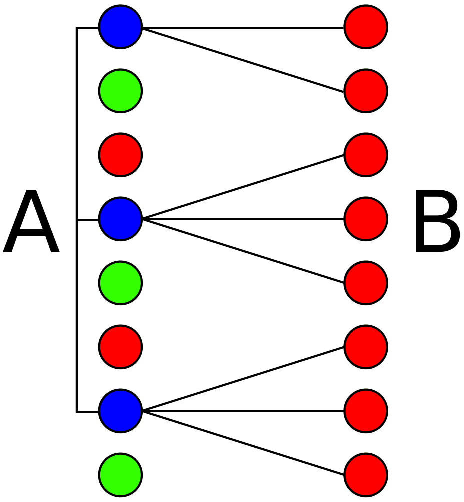

Figure 1.5: Alice announces which encoding scheme to use after each sequence received from the common source, depicted by the different colors

section.

The mutual information is non-negative and only zero if the joint probability matrix factorizes. This, and the way to prove it is well known. It can be seen by observing that $(-\log)$ is a strictly convex function, this implies

$$\begin{aligned} I = \sum_{r,j} p_{rj} \log \frac{p_{rj}}{p_{r\cdot} p_{\cdot j}} &= \sum_{r,j} p_{rj} (-\log) \left(\frac{p_{r\cdot} p_{\cdot j}}{p_{rj}} \right) \\ &\geq -\log \left(\sum_{r,j} \frac{p_{r\cdot} p_{\cdot j}}{p_{rj}} p_{rj} \right) = -\log(1) = 0. \end{aligned}$$

Equality holds *iff* for all non-zero elements of p_{rj}

$$\frac{p_{rj}}{p_{r\cdot} p_{\cdot j}} = 1.$$

This means that the probabilities factorize

$$p_{rj} = p_{r\cdot} p_{\cdot j}.$$

It is quite interesting to note at this point that zero mutual information is stronger than the covariance to be zero, which usually is called *uncorrelated*. The following gives an example

$$p_{rj} = \frac{1}{8} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 3 & 1 \end{pmatrix}$$

with the random variables taking value in $0,1$ on Alice's side and $0,1,2$ on Bobs side. The covariance is defined by

$$\mathrm{cov}(X,Y) := \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle = \langle XY \rangle - \langle X \rangle \langle Y \rangle$$

which is in this case

$$\mathrm{cov}(X,Y) = \frac{5}{8} - \frac{1}{2} \cdot \frac{5}{4} = 0.$$

The joint probability matrix does not factorize, which can be seen from the zero in the lower left entry of the matrix.

The important result by Davies [17] states that if Bob merges two outcomes, in general he loses information.

Theorem 1 (Davies [17]). *Let p_{rj} be a probability matrix, and $\tilde{p}_{rj}$ be given by replacing two columns of p_{rj} with one column representing their sum. For the* *mutual information this implies*

$$I(\tilde{p}_{rj}) \leq I(p_{rj}) \tag{1.3}$$

with equality if and only if the two columns are proportional to each other.

Proof. Label the two columns j, k and expanding both sides, we are left to show

$$\sum_r \left(\left(p_{rj} + p_{rk}\right) \log \frac{p_{rj} + p_{rk}}{p_{\cdot j} + p_{\cdot k}} - p_{rj} \log \frac{p_{rj}}{p_{\cdot j}} - p_{rk} \log \frac{p_{rk}}{p_{\cdot k}} \right) \leq 0. \tag{1.4}$$

Each term in the bracket can be rewritten as

$$p_{rk} \left(x_r \log \left(\frac{1 + \frac{1}{x_r}}{1 + y} \right) + \log \left(\frac{1 + x_r}{1 + \frac{1}{y}} \right) \right) \tag{1.5}$$

with

$$x_r = \frac{p_{rj}}{p_{rk}}, \quad y = \frac{p_{\cdot k}}{p_{\cdot j}}. \tag{1.6}$$

To show that this term is always non-positive, we observe that the term is zero for $x_r = 1/y$, and the derivative with respect to x_r is given by

$$p_{rk} \log \frac{1 + \frac{1}{x_r}}{1 + y} \tag{1.7}$$

which is positive for $x_r < 1/y$ and negative for $x_r > 1/y$. So each term in (1.4) is non-positive, and zero only if the columns are proportional. □

This theorem can also be understood as a special case of the statement that the mutual information is a convex function in the outcomes on Bob's side, more precisely

Theorem 2. *Assume we have two probability distributions p^1_{rj} and p^2_{rj} which have the same marginal probabilities on Alice's side, i.e. $p^1_{r\cdot} = p^2_{r\cdot}$, then*

$$p^\lambda_{rj} = \lambda p^1_{rj} + (1-\lambda)\, p^2_{rj},\;\; 0 \leq \lambda \leq 1$$

is a probability distribution and

$$I(p^\lambda_{rj}) \leq \lambda I(p^1_{rj}) + (1-\lambda) I(p^2_{rj}).$$

For more than two joint probability distributions we have for any probability distribution q_l and joint probability distributions p^l_{jk}

$$I\left(\sum_l q_l\, p^l_{rj}\right) \leq \sum_l q_l I\left(p^l_{rj}\right),$$

under the assumption all the probability distribution have the same marginal distributions on Alice's side.

Proof. One of the proofs for this statement was presented by Řeháček *et. al.* in [5]. One has to show that the second derivative with respect to λ is always non-negative, which can be seen by calculating

$$\begin{aligned}
\frac{dI(p^\lambda_{rk})}{d\lambda} &= \sum_{r,k} \left(p^1_{rk} - p^2_{rk}\right) \log\left(\frac{p^\lambda_{rk}}{p^\lambda_{r\cdot} p^\lambda_{\cdot k}}\right) = \sum_{r,k} \left(p^1_{rk} - p^2_{rk}\right) \log\left(\frac{p^\lambda_{rk}}{p^\lambda_{\cdot k}}\right), \\
\frac{d^2 I(p^\lambda_{rk})}{d\lambda^2} &= \sum_{r,k} \left(p^1_{rk} - p^2_{rk}\right) \left(\frac{p^1_{rk} - p^2_{rk}}{p^\lambda_{rk}} - \frac{p^1_{\cdot k} - p^2_{\cdot k}}{p^\lambda_{\cdot k}}\right) \\
&= \sum_{r,k} \left(p^1_{rk} - p^2_{rk}\right) \left(\frac{p^\lambda_{\cdot k}(p^1_{rk} - p^2_{rk}) - p^\lambda_{rk}(p^1_{\cdot k} - p^2_{\cdot k})}{p^\lambda_{rk}\, p^\lambda_{\cdot k}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{r,k} \left(p^1_{rk} - p^2_{rk}\right) \left(\frac{p^2_{\cdot k}(p^1_{rk} - p^2_{rk}) - p^2_{rk}(p^1_{\cdot k} - p^2_{\cdot k})}{p^\lambda_{rk}\, p^\lambda_{\cdot k}} \right) \\
&= \sum_{r,l,k} \left(p^1_{rk} - p^2_{rk}\right) \left(\frac{p^2_{lk} p^1_{rk} - p^2_{rk} p^1_{lk}}{p^\lambda_{rk}\, p^\lambda_{\cdot k}} \right).
\end{aligned}
$$

The trick is now to multiply the denominator and the first factor by p^λ_{lk}, thereby making the fraction anti-symmetric in r,l, and then use the anti-symmetry on the first factor, i.e.

$$
\begin{aligned}
\frac{d^2 I(p^\lambda_{rk})}{d\lambda^2} &= \sum_{r,l,k} p^\lambda_{lk} \left(p^1_{rk} - p^2_{rk}\right) \left(\frac{p^2_{lk} p^1_{rk} - p^2_{rk} p^1_{lk}}{p^\lambda_{lk}\, p^\lambda_{rk}\, p^\lambda_{\cdot k}} \right) = \sum_{r,l,k} p^2_{lk} p^1_{rk} \left(\frac{p^2_{lk} p^1_{rk} - p^2_{rk} p^1_{lk}}{p^\lambda_{lk}\, p^\lambda_{rk}\, p^\lambda_{\cdot k}} \right) \\
&= \sum_{r,l,k} \frac{(p^2_{lk} p^1_{rk} - p^2_{rk} p^1_{lk})^2}{2 p^\lambda_{lk}\, p^\lambda_{rk}\, p^\lambda_{\cdot k}} \geq 0. \qquad (1.8)
\end{aligned}
$$

The second statement follows simply by induction.

We would like to see as well when (1.8) can be zero. For this to happen each term must vanish individually, i.e.

$$(p^2_{lk} p^1_{rk} - p^2_{rk} p^1_{lk}) = 0.$$

Assume that p^2_{lk} or p^1_{lk} is non-zero for one value of l, it follows that the two columns must be proportional. □

We note theorem 1 can be obtained by choosing a second distribution p^2_{jk} with the two columns in question exchanged and setting λ to one half. We also note that merging equivalent columns does not change the mutual information.

1.2 Quantum States, POVMs and Accessible Information

In this section we will introduce communication using quantum states and measurements. Since we are interested in quantum information, let us have a look at the following scenario.

Alice wants to send her message to Bob by encoding the letters of her alphabet using quantum states. A quantum *state* ρ is described by a complex positive-semidefinite operator on a finite dimensional complex Hilbert-space $\mathcal{H}$ with unit trace, i.e.

$$\forall \psi \in \mathcal{H}: \;\; \langle\psi|\,\rho\,|\psi\rangle \in [0,\infty), \;\; \mathrm{tr}(\rho) = 1.$$

Positive-semidefiniteness implies the operator is hermitian. A state is called *pure* if there exists a vector ψ such that $\rho = |\psi\rangle\langle\psi|$.

Alice can prepare states (for example using the polarization degree of freedom of photons or the spin degree of freedom of electrons) at will and send them to Bob. After receiving a state from Alice, Bob can choose a measurement to acquire information about the received state. Since quantum mechanics is a probabilistic theory, Bob will get one of his outcome with a well-defined probability. These measurements are modeled by *POVMs* (positive operator valued measures). A POVM is defined as a collection of n positive semidefinite operators $\Pi = \{\Pi_j\}$ fulfilling the conditions

$$\Pi_j \geq 0, \;\; \sum_j \Pi_j = \mathbb{I}, \tag{1.9}$$

where $\mathbb{I}$ denotes the identity operator. The elements of the POVM are called *outcomes*. Each individual measurement gives exactly one outcome, i.e. ‘one click’ in one of the outcomes of the ideal measurement apparatus (assuming perfect detectors). The probabilities of the frequencies of the outcomes are given by the mutual trace,

$$p(\rho, j) = \mathrm{tr}(\rho \Pi_j).$$

And the condition for the Π_j to form a POVM translates to p being a probability distribution, i.e.

$$p(\rho, j) \geq 0, \ \sum_j p(\rho, j) = 1.$$

A very special kind of measurement is called *von Neumann* or *orthogonal measurement*. In this case, all the outcomes obey the following relation

$$\Pi_l \Pi_j = \delta_{l,j} \Pi_l, \text{ for all l,j.}$$

Historically this was introduced by John von Neumann [18] in terms of self-adjoint operators. In the traditional setup our collection of operators Π would be given by the projectors of the spectral-decomposition of a self-adjoint operator.

Now, since Alice wants to encode her message she translates every letter of her string labeled by r to exactly one state ρ_r. In the following we will absorb the probabilities with which Alice sends the states in the trace of the state, i.e.

$$\operatorname{tr}(\rho_r) = p_r$$

These states are now sent to Bob. Only in rare cases, i.e. when Alice sends orthogonal states, Bob can infer exactly which state was send by Alice. In the other cases we have to look at the joint probability matrix

$$p_{rj} = \operatorname{tr}(\rho_r \Pi_j).$$

This can be viewed as a classical noisy channel, with conditional probabilities

$$p(j|r) = \frac{p_{rj}}{p_r}$$

where $p(j|k)$ denotes the probability that Bob received outcome j under the condition that Alice sent state k. Observe that the order of the indices is reversed compared to the joint probability matrix.

If we restrict ourselves to transmission of classical information, we know from section 1.1 how much information can maximally be transmitted. This amount is given by the mutual information,(we repeat here due to its importance and usage in the remainder of this thesis).

$$I(\{\rho_r\},\{\Pi_j\}) = \sum_{r,j} p_{rj} \log\left(\frac{p_{rj}}{p_{r\cdot}\, p_{\cdot j}}\right),$$

with marginals

$$p_{\cdot j} = \sum_r p_{rj}, \; p_{r\cdot} = \sum_j p_{rj}.$$

Let us assume that the states sent by Alice and their probabilities are fixed. If Bob wants to improve the transmission rate, Bob will aim to choose the best measurement with respect to the mutual information. A measurement which achieves the maximum of the mutual information is called an *optimal measurement* and the maximum of the mutual information called the *accessible information*,

$$I_{acc}(\{\rho_r\}) := \max_{\{\Pi_k\}} I\left(\{\rho_r\}, \{\Pi_j\}\right).$$

Immediately the question arises, is there always an orthogonal measurement among the optimal measurements? The answer to this is in general 'no'. It has been conjectured though, that if Alice uses only two states, it is indeed the case. This is called the *orthogonal measurement conjecture*.

Conjecture 1 (orthogonal measurement conjecture)**.** *Let* ρ_0 *and* ρ_1 *be states on a finite dimensional Hilbert space. There exists an orthogonal measurement* Π_j *such that the mutual information is equal to the accessible information, i.e.*

$$I(\{\rho_0, \rho_1\}, \{\Pi_j\}) = I_{acc}(\{\rho_0, \rho_1\}).$$

In this thesis we will prove this conjecture to be true, for states with at most a two-dimensional joint support.

For now, we continue by reviewing some of the known results about the mutual and accessible information in the quantum case.

Holevo showed [19] that the mutual information is always bounded by the so called *Holevo quantity* or Holevo χ function,

$$I_{acc}(\{\rho_r\}) \leq S\left(\sum_r \rho_r\right) - \sum_r \mathrm{tr}(\rho_r) S\left(\frac{\rho_r}{\mathrm{tr}(\rho_r)}\right) = \chi(\{\rho_r\}),$$

where S denotes the (von Neumann) entropy of the state, i.e.

$$S(\rho) = -\mathrm{tr}(\rho \log(\rho)).$$

Holevo [20], in the general case, and Hausladen *et.al* [21], in case of pure states, showed that this quantity can be achieved asymptotically if Bob is allowed to perform collective measurements on all the states sent to him by Alice. This is different from our current setup in which Bob can only probe each state individually.

The determination of the accessible information and the Holevo quantity are a sub-problem of the more general problem of channel capacities. A channel for quantum states is described by a completely positive super-operator

$$(L \otimes \mathbb{I}_d)(\rho) \geq 0,$$

for all states ρ and all d, where $\mathbb{I}_d$ denotes the identity in d dimensions. For the channel to be lossless we have to have

$$L^\dagger(\mathbb{I}) = \mathbb{I}.$$

Where $L^\dagger$ denotes the adjoint of L with respect to the Hilbert-Schmidt inner product.

For a given channel L we can define the following capacities

$$C^{1,1} = \max_{\{\rho_r\}} I_{acc}(\{\mathrm{L}(\rho_r)\})$$

$$C^{1,\infty} = \max_{\{\rho_r\}} \chi(\{\mathrm{L}(\rho_r)\})$$

For practical, experimental, implementations, the first quantity is highly relevant, since large collective measurements are extremely difficult to perform. Both quantities are important for theoretical considerations as well. A tremendous amount of work went into the question if the $C^{1,\infty}$ quantity is additive for tensor product channels; a conjecture which has been disproved only recently by Hastings [22].

Theorem 1 from the previous section allows us to show that an optimal POVM can be reached by using rank-1 outcomes. More generally, if we restrict ourself to outcomes chosen from a compact set, an optimal POVM can be reached by using extremal states of the set only.

Theorem 3. *Let M be a compact subset of positive $n \times n$ operators, then a POVM which maximizes the mutual information with the outcomes of the POVM restricted to M, can be chosen such that all outcomes are extremal points of M.*

Proof. Take any POVM which consists of elements of M, any non extremal outcome can be written as a convex sum of extremal points in M, i.e.

$$\Pi_j = \sum_l q_l \, \Xi_j^l.$$

If M were convex, this is part of the Krein-Milman theorem. Since we do not require

M to be convex, we have to work slightly harder. We have

$$M \subseteq \mathrm{hull}(M) = \mathrm{hull}(\mathrm{ex}(\mathrm{hull}(M))) = \mathrm{hull}(\mathrm{ex}(M)),$$

where *hull* denotes the convex hull, and *ex* denotes the extremal points of a set. The first equality follows from the Krein-Milman theorem.

Stringing all these extremal outcomes together creates a new POVM, and theorem 1 immediately completes the proof. □

If there exists a basis such that each state of a collection of states has a real matrix representation in this basis, we say that the states are *real*. If Alices states are real, any complex POVM can be transformed into a real one giving the same probabilities with the same number of outcomes, as the following theorem by Sasaki *et.al.* [23] shows

Theorem 4 (Sasaki et.al. [23]). *Let* ρ *be a state with real matrix representation and* Ξ *be an n-outcome POVM, then* $\Pi_j = Re(\Xi_j)$ *defines a real POVM with the same probabilities for its outcomes.*

Proof. To see that Π_j are positive operators we first note that the complex conjugate of a positive operator is positive as well, hence the real part is the sum of two positive operators, therefore positive. Since the identity matrix is real the new POVM will sum up to the identity as well. The probabilities are equal since

$$\begin{aligned}\mathrm{tr}(\rho\,\Xi_j) &= \sum_{kl}\frac{1}{2}(\rho_{kl}+\rho_{lk})\Xi_j^{lk} = \sum_{kl}\frac{1}{2}\rho_{kl}(\Xi_j^{lk}+\Xi_j^{kl}) \\ &= \mathrm{tr}\left(\frac{1}{2}\rho(\Xi_j+\Xi_j^*)\right) = \mathrm{tr}(\rho\,\Pi_j).\end{aligned}$$

□

Note that in this case , the complex POVM might consist of pure states, while the constructed real one will have in general a higher rank in each outcome. An example of this was given by Suzuki *et.al.* in section 6.4 of the paper [6].

For clarifying the structure of POVMs it is useful to look at it in the following way. Let Π_j be a POVM with all outcomes non-vanishing. We can normalize the outcomes of the POVM, i.e. define

$$\hat{\Pi}_j = \frac{\Pi_j}{\mathrm{tr}(\Pi_j)}, \;\; \mu_j := \frac{\mathrm{tr}(\Pi_j)}{d}. \tag{1.10}$$

In this case the condition for the POVM to sum up to identity becomes the statement that the trace-normalized identity is a convex combination of the normalized outcomes,

$$\sum_j \mu_j \hat{\Pi}_j = \frac{\mathbb{I}}{d},$$

and performing the trace on both sides shows that μ_j is a probability measure.

$$\sum_j \mu_j = 1, \;\; \mu_j > 0 \;\; \text{for all } j$$

Therefore the identity is in the convex hull of the normalized states. This observation, made by Davies, allows us to use a modified version of Caratheodory's theorem to show the following lemma, which will allow us to prove an important theorem found by Davies and sharpened for real states by Sasaki *et al.*.

Lemma 5. *Let $\mathcal{H}$ be a d-dimensional Hilbert space, and* Π *an n-outcome POVM* *with distinct outcomes. For* Π *to be an extremal POVM the number of non vanishing outcomes is limited to* d^2 *if* $\mathcal{H}$ *is a complex space, and limited to* $d(d+1)/2$ *if it is a real space.*

Proof. The space in which the normalized POVM live in is the convex set of all positive operators with trace one. This is a subset of a D dimensional real affine vector space, with $D = d^2 - 1$ in the complex case and $D = d(d+1)/2 - 1$ in the real case. Take any POVM $\{\Pi_j\}$ with $N > D+1$ non vanishing elements, define the normalized operators and probabilities

$$\hat{\Pi}_j = \frac{\Pi_j}{\mathrm{tr}(\Pi_j)}, \;\; \mu_j := \frac{\mathrm{tr}(\Pi_j)}{d}.$$

Fixing the first element $\hat{\Pi}_1$, the difference vectors are linearly dependent, i.e. the equation

$$\sum_{j=2}^{N} \beta_j(\hat{\Pi}_1 - \hat{\Pi}_j) = 0 \tag{1.11}$$

has nontrivial solutions for β_j. Assigning $\beta_1 = -\sum_{j=2}^{N} \beta_j$, we get

$$\sum_{j=1}^{N} \beta_j \hat{\Pi}_j = 0, \;\; \sum_{j=1}^{N} \beta_j = 0 \tag{1.12}$$

We can add any multiple of the β_j to the weights of our normalized POVM to create new weights

$$\tilde{\mu}_j^{\pm} = \mu_j \pm \alpha\beta_j \tag{1.13}$$

which will still sum up to identity, i.e.

$$\sum_j \tilde{\mu}_j^{\pm} = 1.$$

To maintain non-negativity of the new probability measure we set

$$\alpha := 1/\max_j \left\{ \frac{|\beta_j|}{\mu_j} \right\}, \tag{1.14}$$

which keeps the $\tilde{\mu}_j^{\pm}$ non-negative, since

$$\tilde{\mu}_j^{\pm} \geq 0 \;\Leftrightarrow\; \frac{1}{\alpha} \geq \mp \frac{\beta_j}{\mu_j}.$$

With this we can define two new POVMs, $\tilde{\Pi}_j^{\pm}$ whose outcomes are defined as

$$\tilde{\Pi}_j^{\pm} := \tilde{\mu}_j^{\pm} \, d \, \hat{\Pi}_j \, .$$

To check that these are POVMs, we note

$$\sum_j \tilde{\Pi}_j^{\pm} = \sum_j d \left(\mu_j \hat{\Pi}_j \pm \alpha \beta_j \hat{\Pi}_j \right) = \mathbb{I} + 0$$

and

$$\tilde{\Pi}_j^{\pm} = \tilde{\mu}_j^{\pm} \, d \, \hat{\Pi}_j \geq 0$$

since $\mu^{\pm} \geq 0$.

Observing that our original POVM is a convex combination of two POVM,

$$\Pi_j = \frac{1}{2}\tilde{\Pi}_j^+ + \frac{1}{2}\tilde{\Pi}_j^- \tag{1.15}$$

shows that any POVM with more than $D+1$ outcomes cannot be extremal. □

The following theorem goes back to the work of Davies [17] and was extended to the real case by Sasaki, Barnett, Jozsa,Osaki and Hirota in [23].

Theorem 6 (D-SBJOH)**.** *Let $\mathcal{H}$ be a d-dimensional Hilbert space, an optimal POVM* Π *can be chosen to consist of rank-1 outcomes and the number of outcomes can be limited to* d^2 *if $\mathcal{H}$ is a complex space, and limited to* $d(d+1)/2$ *real outcomes if the states have a mutual real matrix representation.*

Proof. In case the states have a real mutual matrix representation we can limit ourself to real POVMs due to theorem 4.

From theorem 3 we can always restrict ourself to POVMs whose outcomes are rank-1. The set of rank-1 outcome POVMs is a compact, but not in general convex. It is convex in the probabilities introduced in 1.10. The mutual information takes its maximum at the extremal points of this set. From the previous lemma and its proof, we see when the number of outcomes exceeds d^2 or $d(d+1)/2$ it cannot be extremal. □

The idea of the proof of theorem 4 can be generalized. Assume we have a superoperator L, such that the states are eigenstates of this operator with eigenvalue one, i.e.

$$L(\rho_r) = \rho_r.$$

This implies that the joint probability matrix is invariant as well, and

$$p_{rj} = \mathrm{tr}(\rho_r \Pi_j) = \mathrm{tr}(L(\rho_r)\Pi_j) = \mathrm{tr}(\rho_r L^\dagger(\Pi_j)), \tag{1.16}$$

where $L^\dagger$ denotes the adjoint of L with respect to the matrix scalar product. In the (rare) case where $L^\dagger$ maps every POVM to another POVM, we can restrict our search to POVMs where each outcome is an element of the image of $L^\dagger$. In the above example L was given by a projection to the real parts of the matrix, L is hermitian if its domain is restricted to the space of hermitian matrices.

The following shows that an optimal POVM for commuting states is von Neumann, which is an expected result.

Theorem 7. *An optimal POVM for mutually commuting states ρ_i is given by a von Neumann measurement which is diagonal in an eigenbasis of the states.*

Proof. Choose a basis which diagonalizes the states. Define a projector L onto the diagonal. It is clear that the image of L is convex and its extremal states are pure states which already implies that one optimal measurement is orthogonal. □

A physically intuitive but less trivial result is, that if the states can be mutually decomposed into block diagonal matrices, an optimal POVM can be constructed from an optimal POVM of the independent blocks.

Theorem 8. *Assume we have states ρ_l which are written as block diagonal matrices, and we know for each block a POVM which maximizes the mutual information. Denote the number of blocks is by M, label the outcomes by Π_j^m, where j labels the outcome and m labels the block and d_m denotes the dimension of block m. Then an* *optimal POVM is given by stringing all the outcomes together in one POVM, i.e.*

$$\{\Pi^{total}\} := \bigcup_m \{0_{(D_{m-1})} \oplus \Pi^m \oplus 0_{(D_M - D_m)}\}.$$

Where $0_{(n)}$ *denotes the zero matrix of dimension n and*

$$D_m := \sum_{j=1}^{m} d_j.$$

Proof. Choose a basis such that the matrix representation of the states is block diagonal. Define L as the projector on these blocks. The image of L is convex and closed. Since L can be written as

$$L(\rho) = \sum_k P_k^\dagger \rho P_k,$$

with P_k as the orthogonal projection on the subspace of the kth block, we have that $L^\dagger$ preserves positivity of the outcomes and since

$$L^\dagger = L, \;\; L^\dagger(\mathbb{I}) = \sum_k P_k^\dagger P_k = \mathbb{I},$$

it maps POVMs to POVMs. The extremal states of the image of L are pure states each of which are invariant under exactly one projection, and annihilated by all the other projections. The only possibility for a pure state to be block diagonal is to be zero in all of the blocks but one. □

It is also important to get two trivial cases out of the way now. It is clear that if the probability $p_1 = 0$ then no information can be transmitted and the mutual information is zero. Also, if the two states are proportional to each other the mutual information is zero. In the rest of this thesis we will not deal with these trivial cases.

1.3 Variation of POVM

We will use the Naimark extension to define our variation of the POVMs. The following theorem gives us an orthogonal extension of every POVM:

Theorem 9 (Naimark)**.** *For any POVM,* Π*, acting on a Hilbert space* $\mathcal{H}$ *there exists an Hilbert space* $\tilde{\mathcal{H}} \supseteq \mathcal{H}$ *and an orthogonal projector* $P : \tilde{\mathcal{H}} \mapsto \mathcal{H}$*, and a set of orthogonal measurements* $\tilde{\Pi}$ *such that*

$$\Pi_i = P\,\tilde{\Pi}_i\,P$$

and the dimension of $\tilde{\mathcal{H}}$ *can be chosen to be the sum of the rank of the outcomes of* Π.

Proof. To prove the theorem, we take all outcomes to be pure, otherwise we can separate them into new, pure outcomes, and define an $n \times m$ -matrix A, by writing the states as

$$\Pi_i = |q_i\rangle\langle q_i|\,,\ A_{ij} = \langle e_i|\,q_j\rangle,$$

where $|e_i\rangle$ denotes an orthonormal basis of $\mathcal{H}$. The summing to identity condition of the POVM translates to

$$\delta_{k,j} = \langle e_k|\sum_l \Pi_l\,|e_j\rangle = \sum_l A_{kl}A^*_{jl} = \left(AA^\dagger\right)_{kj},$$

where the star denotes complex conjugation of complex numbers.

This implies all the rows of A are orthonormal, which allows us to extend the matrix to a square-unitary matrix, by completing the rows to an orthonormal basis. The columns found here are the Naimark-extension and the projector is given by projecting on the first n components. □

This allows us to define a variation of a POVM, in case the POVM is given by rank-1 states; we extend it to an orthonormal basis, use an infinitesimal unitary rotation and project back on the original Hilbert-space, i.e.

$$\begin{aligned}\delta|j\rangle = \delta P|\tilde{j}\rangle &= \frac{d}{dt}P\exp(iHt)|\tilde{j}\rangle = iPH|\tilde{j}\rangle\\ &= iP\sum_m |\tilde{m}\rangle\langle\tilde{m}|H|j\rangle\\ &= i\sum_m |m\rangle\varepsilon_{mj},\ \ \varepsilon^*_{mj} = \varepsilon_{jm}\end{aligned}$$

$$\begin{aligned}\delta p_{rj} = \delta\langle j|\rho_r|j\rangle &= \sum_m \left(-i\langle m|\varepsilon^*_{mj}\rho_r|j\rangle + i\langle j|\varepsilon_{mj}\rho_r|m\rangle\right)\\ &= -2\sum_m \operatorname{Im}\left(\langle j|\rho_r|m\rangle\varepsilon_{mj}\right).\end{aligned}$$

We can look at the stationary points of the mutual information

$$\begin{aligned}\delta I &= \sum_{r,j}\delta p_{rj}\log\frac{p_{rj}}{p_{\cdot j}}\\ &= -2\sum_{r,j,m}\operatorname{Im}\left(\langle j|\rho_r|m\rangle\varepsilon_{mj}\right)\log\frac{p_{rj}}{p_{\cdot j}}.\end{aligned}$$

For each pair of outcomes (k,l) we can set ε_{mj} for $\{k,l\} \neq \{m,j\}$ to zero, except for $\varepsilon_{kl} = \frac{i}{2}$ and $\varepsilon_{lk} = -\frac{i}{2}$. Explicitly, in three dimensions the matrices are

$$(k,l) = (1,2) : \varepsilon = \frac{1}{2}\begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$(k,l) = (1,3) : \varepsilon = \frac{1}{2}\begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}$$

$$(k,l) = (2,3) : \varepsilon = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix}.$$

These matrices are equal to minus one-half of the imaginary Gell-Mann matrices.

This gives us the following set of variations

$$\delta_{(k,l)} I = \sum_r \Re\left[\langle k|\,\rho_r\,|l\rangle \log\left(\frac{p_{rk}}{p_{\cdot k}}\right) - \langle l|\,\rho_r\,|k\rangle \log\left(\frac{p_{rl}}{p_{\cdot l}}\right)\right]. \tag{1.17}$$

From now on we will focus on the case of two states of a qubit.

Lemma 10. *Let* ρ_1 *and* ρ_2 *be two states of a qubit. It is always possible to find a basis such that both states have a real matrix representation. The accessible information can be reached with a measurement consisting of three real rank-1 outcomes.*

Proof. Diagonalize one of the states, say ρ_1. The state ρ_2 has in general the following matrix representation

$$\rho_2 = \begin{pmatrix} a & b \\ \bar{b} & c \end{pmatrix},$$

with a, c real numbers. The following unitary matrix transforms ρ_2 into a real matrix and keeps ρ_1 invariant,

$$U = \begin{pmatrix} \frac{b}{|b|} & 0 \\ 0 & 1 \end{pmatrix},$$

i.e. $U^\dagger \rho_2 U$ is real. From theorem 6 follows the rest of the statement. □

Specifying equation (1.17) to two states in a real representation we get

$$\delta_{(k,l)} I = \sum_{r=1}^{2} \langle k| \rho_r |l\rangle \log\left(\frac{p_{rk}}{p_{rl}} \frac{p_{\cdot l}}{p_{\cdot k}}\right). \tag{1.18}$$

Here k and l run from one to three. Since we are looking for a stationary point of the mutual information we are interested in the zeros of this function

$$\delta_{(k,l)} I = 0. \tag{1.19}$$

The function (1.18) is always well-defined if none of the states are pure. In case at least one of the states is pure we will show that this function is still continuous in section 3.3 where we focus on pure states.

Since these sets are antisymmetric in k, l we get exactly three independent pairs. Fix one of the directions, say $|1\rangle$, and one vector $|0\rangle$ orthonormal to $|1\rangle$ to complete a basis of our real Hilbert space. Any vector $|n\rangle$ can be expressed as

$$|n\rangle = \beta_0(n)\,|0\rangle + \beta_1(n)\,|1\rangle. \tag{1.20}$$

We want to see what are the restrictions from these equations (1.19) on the vectors. When β_0 is zero the conjecture is trivially true, since the vector would be proportional to $|1\rangle$. Observe that, the function (1.18) is homogeneous in the length of the vectors and therefore it is always possible for the solutions of (1.19) to divide out $\beta_0 \neq 0$ and restrict ourselves to $|n\rangle = |0\rangle + t\,|1\rangle$ with t an arbitrary real number. We get

$$\delta_{(1,n)}I = \langle 1|\,(\rho_1+\rho_2)\,|1\rangle \sum_{r=1}^{2} \alpha_r Q_r'(t) \log \frac{Q_r(t)}{\alpha_1\,Q_1(t)+\alpha_2\,Q_2(t)} \tag{1.21}$$

with

$$Q_r(t) = t^2 + 2t\frac{\langle 1|\,\rho_r\,|0\rangle}{\langle 1|\,\rho_r\,|1\rangle} + \frac{\langle 0|\,\rho_r\,|0\rangle}{\langle 1|\,\rho_r\,|1\rangle}, \quad \alpha_r = \frac{\langle 1|\,\rho_r\,|1\rangle}{\langle 1|\,(\rho_1+\rho_2)\,|1\rangle} \tag{1.22}$$

and *prime* denoting differentiation with respect to t. Introducing

$$\xi_r = \frac{\langle 1|\,\rho_r\,|0\rangle}{\langle 1|\,\rho_r\,|1\rangle}, \quad \eta_r = \frac{\langle 0|\,\rho_r\,|0\rangle}{\langle 1|\,\rho_r\,|1\rangle}, \tag{1.23}$$

the range for these variables is restricted due to positivity of the states to

$$0 \leq \xi_r^2 \leq \eta_r \leq \infty, \;\; 0 \leq \alpha_r \leq 1, \;\; r = 1,2, \;\; \alpha_1 + \alpha_2 = 1. \tag{1.24}$$

Non-negativity of the states ρ_1 and ρ_2 is reflected in the non-negativity of the polynomials Q_1 and Q_2 as the computation of their discriminant shows

$$\xi_r^2 - \eta_r = \left(\frac{\langle 1|\rho_r|0\rangle}{\langle 1|\rho_r|1\rangle}\right)^2 - \frac{\langle 0|\rho_r|0\rangle}{\langle 1|\rho_r|1\rangle} = -\frac{\det(\rho_r)}{\langle 1|\rho_r|1\rangle^2} \leq 0 \text{ for } r = 1,2.$$

In section 3.5 we will show that knowledge of the number of real roots of the function 1.21 allows us to prove the conjecture. Since the function is transcendental, analyzing its roots is not a straight forward task. We will develop some tools in the next chapter.

In particular we will prove the following key theorem

Theorem 11. *Each function defined by*

$$f_{(\alpha,\xi,\eta)}(t) = \sum_{r=1}^{2} \alpha_r Q_r'(t) \log \frac{Q_r(t)}{\alpha_1 Q_1(t) + \alpha_2 Q_2(t)} \tag{1.25}$$

with constraints given by

$$0 \leq \xi_r^2 \leq \eta_r < \infty,\ 0 \leq \alpha_r \leq 1,\ r = 1,2,\ \alpha_1 + \alpha_2 = 1,$$
$$(\xi_1, \eta_1) \neq (\xi_2, \eta_2), \tag{1.26}$$

and $Q_r(t) = t^2 + 2t\xi_r + \eta_r$ and $Q_r'(t) = 2(t + \xi_r)$, has at most two real roots. If $\xi_1 = \xi_2$ and $\eta_1 \neq \eta_2$ the function has exactly one real root. In case $\alpha_1 = 1,0$ or $(\xi_1, \eta_1) = (\xi_2, \eta_2)$ the function vanishes identically.

Proof. Here we will only consider the last two cases, all other cases will be proved in the remaining part of this thesis. If $\alpha_1 = 1,0$ or $(\xi_1, \eta_1) = (\xi_2, \eta_2)$ the function vanishes obviously. In case $\eta_1 = \eta_2$ we have $Q_1 = Q_2 + c$ or $Q_2 = Q_1 + c$ for some positive constant c. Here we consider the case $Q_1 = Q_2 + c$, the other case is shown by a similiar calculation. We have

$$f_{(\alpha,\xi,\eta)}(t) = Q_1'(t)\left[\alpha_1 \log(Q_1) + \alpha_2 \log(Q_1 + c) - \log(Q_1 + \alpha_2 c)\right],$$

since Q_1' is an affine function it has at most one root. We show now that the term in the bracket never vanishes, except for $c = 0$. For $c = 0$ the bracket vanishes. The derivative of the term in the bracket w.r.t. c is

$$-\frac{c\alpha_1\alpha_2}{(Q_1 + c)(Q_1 + \alpha_2 c)}$$

which is always negative. □

The following lemma shows how the function transforms under affine transformations

Lemma 12. *Define the following affine transformation*

$$T(t) = at + b, a \neq 0,$$

we get for a transformed f

$$f_{(\alpha,\xi,\eta)}(T(t)) = a f_{(\alpha,\xi',\eta')}(t),$$

with

$$\xi_r' = \frac{\xi_r + b}{a}, \quad \eta_r' = \xi_r'^2 + \frac{\eta - \xi^2}{a^2}, \quad r = 1,2.$$

The new variables fulfill the same constraint 1.24 as the original variables.

This lemma will help us to find the number of zeros, by reducing the parameter space.

The main idea for analyzing this class of functions is to look at the second derivative in t which is a high-order rational function.

For calculating the derivatives we introduce the abbreviations

$$L := Q_1' Q_2 - Q_2' Q_1, \;\; Q_s := \alpha_1 Q_1 + \alpha_2 Q_2. \tag{1.27}$$

For the first and second derivative with respect to t of $f_{(\alpha,\xi,\eta)}(t)$ we get

$$\begin{aligned} f' &= 2\left(\sum_{l=1}^{2} \alpha_l \log \frac{Q_l}{Q_s}\right) + \alpha_1\alpha_2 \frac{(Q_1'Q_2 - Q_2'Q_1)^2}{Q_sQ_1Q_2} \\ f'' &= \alpha_1\alpha_2 \frac{L}{(Q_1Q_2Q_s)^2} P \end{aligned} \tag{1.28}$$

with

$$P = 3L'(Q_1Q_2Q_s) - (Q_1Q_2Q_s)'L. \tag{1.29}$$

By simply counting we can see that the second derivative is a (8,12) rational function in t, i.e. has a eighth order polynomial in the numerator and a twelfth order polynomial in the denominator. Luckily its structure is graceful and a full analysis is possible. In the next chapter we will talk about the tool of the discriminant to help us analyzing the term P.

Chapter 2

Mathematical Tools

In this chapter we will develop some mathematical tools which will be used in the following chapter. We start with discussing the resultant and the discriminant of polynomials, giving us tools to the determine how many roots a class of polynomials has. In the succeeding section we develop some upper bounds on the number of roots of continuous functions.

2.1 Resultant and Discriminant

In this section we introduce the discriminant and the resultant of of polynomials. We restrict ourself to complex variables, so that every polynomial can be written as a product of linear factors. For a more detailed account we refer to van der Waerden [24]. At the end of the section we prove a key theorem, theorem 14, which will be needed in the subsequent chapter.

For a second order polynomial

$$p(x) = ax^2 + bx + c$$

the discriminant

$$\Delta = b^2 - 4ac$$

determines the number of real roots

$$\Delta \begin{cases} = 0 & \text{if p has exactly one real root} \\ < 0 & \text{if p has no real root} \\ > 0 & \text{if p has two real roots} \end{cases}$$

In general the resultant of two arbitrary polynomial $p(x)$ and $q(x)$ of degree n and m with coefficients a_j and b_j

$$p(x) = \sum_{j=0}^{n} a_j x^j$$
$$q(x) = \sum_{j=0}^{m} b_j x^j$$

is defined as the product of the differences of their roots, specifically

$$R[p,q] = a_n^m b_m^n \prod_{i=1}^{n} \prod_{j=1}^{m} (p_i - q_j) \,,$$

where p_i denotes the (complex) roots of $p(x)$ and q_j the ones of $q(x)$. All roots are counted with multiplicity. The square brackets indicate that R is a function on polynomials. Observe that the resultant is symmetric in p and q, up to a possible minus-sign. The polynomials can be decomposed into linear factors,

$$q(x) = b_m \prod_{j=1}^{m} \left(x - q_j\right).$$

By inserting the roots of one polynomial in the other one, it can be seen that the resultant can be represented by the product of p evaluated at the roots of q,

$$a_n^m \prod_{i=1}^{n} q(p_i) = a_n^m \prod_{i=1}^{n} \left(b_m \prod_{j=1}^{m} \left(p_i - q_j\right) \right) = R[p,q]. \tag{2.1}$$

From this definition it can be seen that the resultant is zero if and only if both polynomials share at least one root.

The discriminant Δ of a polynomial is given by the resultant of the polynomial and its derivative

$$R[p,p'] = (-1)^{n(n-1)/2} a_n \Delta[p]. \tag{2.2}$$

Using formula 2.1 we see that the discriminant is proportional to the product of the square of the differences of the roots of the polynomial, i.e.

$$\begin{aligned}(-1)^{n(n-1)/2} a_n \Delta[p] = R[p,p'] = a_n^{n-1} \prod_{i=1}^{n} p'(p_i) &= a_n^{2n-1} \prod_{j=1}^{n} \prod_{k \neq j} \left(p_j - p_k\right) \\ &= a_n^{2n-1} (-1)^{n(n-1)/2} \prod_{j=1}^{n} \prod_{k>j} \left(p_j - p_k\right)^2.\end{aligned}$$

Therefore

$$\Delta[p] = a_n^{2n-2} \prod_{j<k} \left(p_j - p_k\right)^2. \tag{2.3}$$

If all roots of p are real the discriminant is non-negative. In the case where complex roots are present the discriminant can be negative. There seems to be no easy way to compute the discriminant since it uses the roots of our polynomial, which in general cannot be determined if the degree of the polynomial exceeds four. Luckily there is a different way of computing resultants and discriminants using determinants.

The following holds:

$$R[p,q] = \det \left(\begin{matrix} a_n & a_{n-1} & \dots & a_1 & a_0 & \dots & 0 & 0 \\ 0 & a_n & a_{n-1} & \dots & a_1 & a_0 & \dots & 0 \\ & & & \vdots & & & & \\ b_m & b_{m-1} & \dots & b_1 & b_0 & \dots & 0 & 0 \\ 0 & b_m & b_{m-1} & \dots & b_1 & b_0 & \dots & 0 \\ & & & \vdots & & & & \\ 0 & \dots & 0 & b_m & b_{m-1} & \dots & b_1 & b_0 \end{matrix}\right) \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \end{matrix}}\right\} \text{m} \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \\ a \end{matrix}}\right\} \text{n} \end{matrix} . \qquad (2.4)$$

This matrix is called Sylvester matrix and it is an $(m+n) \times (m+n)$ quadratic matrix. It is formed by writing the coefficient of the first polynomial in the first m rows, in each row shifted by one column to the right. Afterwards the next n rows are filled with the coefficient of the second polynomial, shifted as well. The determinant of the Sylvester matrix gives the resultant of the two polynomials.

For a proof of equation (2.4) we refer to the literature, for example [24]. For our purposes it is enough to show that both terms vanish for the same polynomials. The direction that if the resultant vanishes, the determinant vanishes, is easy. Assume the resultant vanishes, i.e. p and q have a common root, say x. Form the vector

$$\vec{v} = \begin{pmatrix} x^{n+m-1} \\ x^{n+m-2} \\ \vdots \\ x^2 \\ x^1 \\ 1 \end{pmatrix},$$

a direct computation shows that $\vec{v}$ is in the kernel of the Sylvester matrix, therefore the determinant vanishes.

The reverse statement is a little trickier to show. Assume that the determinant is vanishing, this implies that the rows are linear dependent, i.e. there is a non-vanishing vector $\vec{w}$ in the kernel of the transposed matrix. We can now form two polynomials with degree $m-1$ and degree $n-1$ using the components of $\vec{w}$ as coefficients

$$\begin{aligned} r(x) &= \sum_{j=1}^{m} w_j x^{m-j} \\ s(x) &= \sum_{j=1}^{n} w_{m+j} x^{n-j}. \end{aligned}$$

The equations for $\vec{w}$ in the kernel of the transposed matrix translate to

$$r(x)\,p(x) = s(x)q(x)$$

This is only possible if $p(x)$ and $q(x)$ have a common root, as can be seen by decomposing each side into products of their linear factor.

For illustration we look at an example with two quadratic polynomials

$$\begin{aligned} p(x) &= (x-p_1)(x-p_2) = x^2-(p_1+p_2)x+p_1p_2 \\ q(x) &= (x-q_1)(x-q_2) = x^2-(q_1+q_2)x+q_1q_2, \end{aligned}$$

the Sylvester matrix of this system is given by

$$B = \begin{pmatrix} 1 & -(p_1+p_2) & p_1p_2 & 0 \\ 0 & 1 & -(p_1+p_2) & p_1p_2 \\ 1 & -(q_1+q_2) & q_1q_2 & 0 \\ 0 & 1 & -(q_1+q_2) & q_1q_2 \end{pmatrix}$$

and its determinant is equal to

$$\det(B) = (p_1-q_1)(p_1-q_2)(p_2-q_1)(p_2-q_2) = R(p,q),$$

as expected.

Looking at the discriminant of a second order polynomial p,

$$p(x) = ax^2+bx+c,$$

gives us

$$\begin{vmatrix} a & b & c \\ 2a & b & 0 \\ 0 & 2a & b \end{vmatrix} = -ab^2 + 4a^2c = -a\Delta(p)$$

in agreement with the usual definition.

We like to examine the behavior of the discriminant of a polynomial when the highest coefficient is set to zero. In this case the discriminant becomes a multiple of the discriminant of the remaining polynomial, as shown in the next theorem

Theorem 13. *Let $p(x)$ be a polynomial of degree n with coefficients a_i with discriminant $\Delta_n[p]$, $q(x)$ the same polynomial with a_n set to zero and $\Delta_{n-1}[q]$ the discriminant of $q(x)$. The following holds*

$$\lim_{a_n \to 0} \Delta_n[p] = a_{n-1}^2 \Delta_{n-1}[q]$$

Proof. We see from equations (2.2,2.4) that

$$\Delta_n[p] = (-1)^{n(n-1)/2} \times \tag{2.5}$$

$$\begin{vmatrix} 1 & a_{n-1} & \dots & a_1 & a_0 & \dots & 0 & 0 \\ 0 & a_n & a_{n-1} & \dots & a_1 & a_0 & \dots & 0 \\ & & & \vdots & & & & \\ n & (n-1)\,a_{n-1} & \dots & a_1 & a_0 & \dots & 0 & 0 \\ 0 & n\,a_n & (n-1)\,a_{n-1} & \dots & a_1 & a_0 & \dots & 0 \\ & & & \vdots & & & & \\ 0 & \dots & 0 & n\,a_n & (n-1)\,a_{n-1} & \dots & a_1 & a_0 \end{vmatrix} . \tag{2.6}$$

Expanding the determinant in the first column and setting a_n to zero we get

$$\begin{aligned} \lim_{a_n \to 0} \Delta_n[p] &= (-1)^{n(n-1)/2} \times \left\{ 1 \times (-1)^{n-2} (n-1) a_{n-1} \operatorname{Res}[q, q'] \right. \\ &\left. + (-1)^{n-1}\, n \times a_{n-1} \operatorname{Res}[q, q'] \right\} \\ &= (-1)^{n(n-1)/2} \times (-1)^{n-1}\, a_{n-1} \operatorname{Res}[q, q']. \end{aligned}$$

Using equation (2.2) again, we get

$$\begin{aligned} \lim_{a_n \to 0} \Delta_n[p] &= (-1)^{n(n-1)/2} \times (-1)^{n-1}\, a_{n-1}^2\, (-1)^{(n-1)(n-2)/2}\, \Delta[q] \\ &= a_{n-1}^2\, \Delta[q] \end{aligned}$$

□

After this introduction and the previous result we are prepared to show the main result of this section:

Theorem 14. *Let G be a family of real valued polynomials with formal degree n,* *i.e.*

$$G : \mathbb{R}^m \to \mathbb{R}[x], z \mapsto G_z = \sum_{k=0}^{n} g(z)_k x^k$$

which is continuous in the coefficients of x, and $D[g]$ a connected domain of G for which the discriminant of G does not vanish. The number of roots of G_z is constant on each of the following domains

$$\begin{aligned} D_0[g] &= \{z \in D | g(z)_n = 0\}, \\ D_1[g] &= D \backslash D_0, \end{aligned}$$

and the number of roots on D_1 is one more than the number of roots on D_0.

Proof. Assume that the discriminant does not vanish. This implies that the polynomials on D_1 do not have any double root. Since the roots of the polynomials depend continuously on its coefficients this implies that the polynomials of each connected component of D_1 have the same number of real roots. The non-vanishing of the discriminant on D_0, where the highest coefficient vanishes, implies, due to theorem 13, that the second highest coefficient is non-zero and that no double root occurs. This fixes the number of real roots on each connected component of D_0, and there is exactly one root less which went off to infinity. □

2.2 Upper bounds on the number of roots of a function

Here we introduce some theorems and lemmata which will help us limit the number of real roots of a class of functions, using information from the first and second derivative. We start with the well-known Rolle theorem

Theorem 15 (Rolle)**.** *Let f be a continuous, real valued function on the closed interval $[a,b]$ with $a,b \in \mathbb{R} \cup \{-\infty,+\infty\}$, $a \neq b$ and differentiable on (a,b). If $f(a) = f(b)$ then there exists a $c \in (a,b)$ such that the derivative of f is zero at this point, i.e. $f'(c) = 0$.*

In case the function is not differentiable but still continuous the following version holds

Lemma 16. *Let f be a real valued, continuous non-constant function on the interval $[a,b]$ with $a,b \in \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty,+\infty\}$ and $a \neq b$. If $f(a) = f(b)$ there exists a $c \in (a,b)$ such that c is an extremal point of f, i.e. at c is a maximum or minimum.*

Proof. Continuous functions map compact sets to compact sets. Let $a,b \in \bar{\mathbb{R}}$ and $a \neq b$, $M := [a,b]$ is a compact set, either in the topology of $\mathbb{R}$ if $a,b \neq \infty$, or in the two point compactification of $\mathbb{R}$ in the other case. The function f maps M to a compact subset of $\mathbb{R}$, so its image is bounded and closed, therefore attains a maximum and a minimum. At least one of these extrema has to be attained between a and b, otherwise the function would be constant. □

Lemma 16 implies theorem 15 immediately.

This allows us to limit the number of zeros of a function by the number of zeros of its derivative.

Lemma 17. *Let $M = [a,b]$, $M = (a,b)$, $M = (a,b]$ or $M = [a,b)$ with $a,b \in \bar{\mathbb{R}}$ and $f \in C^1(M,\mathbb{R})$ with its derivative having a finite number of roots. The number of* *roots of f is at most one more than the number of roots of the derivative*

$$|f^{-1}(0)| \leq |(f')^{-1}(0)| + 1,$$

where $|\cdot|$ denotes the number of elements of a set.

Proof. According to theorem 15 between two successive roots of f must be one root of the derivative. □

In case that the function is not everywhere differentiable we can substitute extrema instead of roots of the derivative.

Lemma 18. *Let $M = [a,b]$, $M = (a,b)$, $M = (a,b]$ or $M = [a,b)$ with $a,b \in \bar{\mathbb{R}}$ and $f \in C^0(M,\mathbb{R})$ with finite number of extrema N. The number of roots of f on the interval is at most one more than the number of extrema*

$$|f^{-1}(0)| \leq N + 1.$$

where $|\cdot|$ denotes the number of elements of a set.

Proof. Same proof as lemma 17 except that lemma 16 is used instead of theorem 15. □

In a later section we will be mostly concerned with functions which converge to zero at infinity.

Lemma 19. *Let $f \in C^1(\mathbb{R})$ which converges to zero at plus and minus infinity and its first derivative has a finite number of roots. Then we have*

$$|f^{-1}(0)| \leq |(f')^{-1}(0)| - 1$$

on $\mathbb{R}$.

Proof. Extend f to the two point compactification of $\mathbb{R}$, apply lemma 17 and subtract one root for plus infinity and one for minus infinity. □

Chapter 3

The Proof

In the following section we will present the proof of the orthogonal measurement conjecture for states of a qubit. The main focus is on the variation of the mutual information, given in equation (1.18) and the associated function $f_{\alpha,\xi,\eta}(t)$ given in equation (1.25). We start in the next section by analyzing the asymptotic behavior of $f_{\alpha,\xi,\eta}(t)$ and its derivatives. We prove theorem 11 for two mixed states in section 3.2, for two pure states in section 3.3, and for one mixed and one pure state in section 3.4. In section 3.5 we present the proof of the orthogonal measurement conjecture for two states of a qubit. We conclude this chapter with a short discussion about the von Neumann measurement that maximizes the mutual information.

3.1 Asymptotic Behavior

To complete our analysis, we have to look at the asymptotic behavior of the class of functions (1.25). We observe

$$\lim_{t\to\infty} \frac{Q_r}{Q_s} = 1, \ \ r = 1,2,$$

and the terms in front of the logarithm going to infinity, so we have to expand the logarithm to find the right asymptotic behavior. We have

$$\begin{aligned}
\log\left(\frac{Q_1}{Q_s}\right) &= \log\left(\frac{Q_1}{\alpha_1 Q_1 + \alpha_2 Q_2}\right) = -\log\left(1+\alpha_2\left(\frac{Q_2-Q_1}{Q_1}\right)\right) \\
&\approx -\left(\alpha_2\frac{Q_2-Q_1}{Q_1} - \frac{\alpha_2^2}{2}\left(\frac{Q_2-Q_1}{Q_2}\right)^2\right) + O(t^{-3}) \\
&= 2\alpha_2(\xi_1-\xi_2)\frac{1}{t} + O(t^{-2}) \\
\log\left(\frac{Q_2}{Q_s}\right) &\approx -\left(-\alpha_1\frac{Q_2-Q_1}{Q_2} - \frac{\alpha_1^2}{2}\left(\frac{Q_2-Q_1}{Q_1}\right)^2\right) + O(t^{-3}) \\
&= 2\alpha_1(\xi_2-\xi_1)\frac{1}{t} + O(t^{-2})
\end{aligned}$$

Therefore

$$\lim_{t\to\pm\infty} f(t) = \lim_{t\to\pm\infty} 2t\left(\alpha_1 2\alpha_2(\xi_1-\xi_2)\frac{1}{t} + \alpha_2 2\alpha_1(\xi_2-\xi_1)\frac{1}{t}\right) + O(t^{-1}) = 0, \tag{3.1}$$

and for the first and second derivative of f given by

$$f' = 2\left(\sum_{l=1}^{2}\alpha_l \log\frac{Q_l}{Q_s}\right) + \alpha_1\alpha_2\frac{(Q_1'Q_2 - Q_2'Q_1)^2}{Q_s Q_1 Q_2}$$

$$f'' = \alpha_1\alpha_2 \frac{L}{(Q_1Q_2Q_s)^2}P. \tag{3.2}$$

We immediately see that the first and second derivative of f converge to zero at plus and minus infinity by observing that $Q_1Q_2Q_s$ is a sixth order polynomial and LP at most an eighth order polynomial. We have

$$\begin{aligned} \lim_{t\to\pm\infty} f'(t) &= 0, \\ \lim_{t\to\pm\infty} f''(t) &= 0. \end{aligned} \tag{3.3}$$

To get a better feeling for the functions involved we will now examine their asymptotic behavior in greater detail.

For the remainder of this section we will use a rescaled and translated version of f with $\xi_1 = 0$ and $\eta_1 = 1$, and label the rescaled and translated values of ξ_2 by ξ and η_2 by η. As usual $\alpha_1 = 0, 1$ is excluded.

Expanding the logarithms in function (1.25) to fourth order in t around infinity leads to

$$\begin{aligned} f(t) \approx{} & \frac{2}{3}\alpha_2\,\alpha_1\,\xi\,\left((4-2\,\alpha_1\,)\xi^2 - 3\,\eta + 3\right)\frac{1}{t^2} - \alpha_2\,\alpha_1\,\left[(8\,\alpha_1^2 + 24\,\alpha_2)\,\xi^4\right. \\ & \left.+(12\,\alpha_1\,\eta - 24\,\eta + 12\,\alpha_2)\xi^2 + 3(\eta-1)^2\right]\frac{1}{3t^3} + O(t^{-4}) \end{aligned} \tag{3.4}$$

In most cases the function approaches zero at infinity as t^{-2}, in some cases the prefactor of t^{-2} will vanish and the function will behave asymptotically as t^{-3}. We now show that if the coefficient of t^{-2} vanishes it is not possible for the coefficient of t^{-3} to vanish as well.

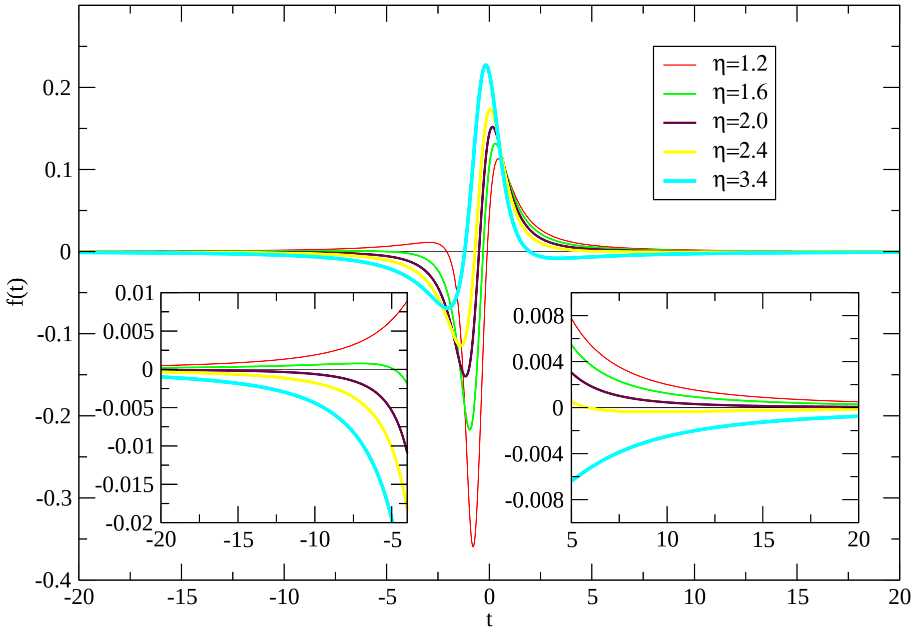

Figure 3.1: Function f with parameters $\alpha_1 = 1/2$, $\xi = 1$ and various values for η as indicated in the graph. Insets show two magnified region. Observe the transition of asymptotic behavior at $\eta = 2$ (i.e. maroon colored curve).

The first term of (3.4), proportional to t^{-2}, is zero iff

$$\xi = 0, \text{ or } \alpha_1 = \frac{4\xi^2 - 3\eta + 3}{2\xi^2}. \tag{3.5}$$

In the case that ξ equals zero, the prefactor of $(3t)^{-3}$ which is the second term in (3.4) reads

$$-\alpha_2\alpha_1 (\eta - 1)^2,$$

and this being zero implies that the two states are proportional to each other.

In case $\xi \neq 0$, the constraint on α_1 to be bounded by zero and one turns into:

$$\eta \in (1+\frac{2}{3}\xi^2, 1+\frac{4}{3}\xi^2), \tag{3.6}$$

keeping in mind the general constraint of $\eta \geq \xi^2$. Substituting α_1 of equation (3.5) into the coefficient proportional to t^{-3} of (3.4), gives

$$-\frac{1}{12\xi^4}\left(2\xi^2-3\eta+3\right)\left(-4\xi^2+3\eta-3\right)\left(8\xi^4-12\eta\xi^2+3\eta^2-6\eta+3\right). \tag{3.7}$$

Ignoring the numerator, the second factor is equal to $-\alpha_1$ so it is strictly negative. Notice that the first factor of (3.7) is strictly negative as well, since η is greater than $1+\frac{2}{3}\xi^2$ (3.6), giving

$$2\xi^2-3\eta+3 < 2\xi^2-3-2\xi^2+3 = 0.$$

For the third and last factor, we have

$$8\xi^4-8\eta\xi^2-3\eta+3-4\eta\xi^2+3\eta^2-3\eta < 8\,(\xi^2-\eta)\,\xi^2-3\eta+3 < 0,$$

by realizing that the last three terms of the first line sum up to η times negative α_1 and $\eta \geq 1$ because of (3.6).

3.2 Two Mixed States

In this section we deal with the case that both states are mixed, which avoids us having to consider any poles in the first- or second derivative of $f_{(\alpha,\xi,\eta)}(t)$ with respect to t. Since the number of roots in t is translation and scaling invariant, it is always possible to set $\xi_1 = 0$ and $\eta_1 = 1$. It will be convenient to label the translated and rescaled values of ξ_2 simply as ξ and η_2 as η .

The second derivative of $f_{(\alpha,\xi,\eta)}(t)$ is, as stated in equation (1.28),

$$f'' = \alpha_1\alpha_2 \frac{L}{(Q_1 Q_2 Q_s)^2} P,$$

with

$$L = 2(\xi t^2 + (\eta - 1)t - \xi),$$

and P

$$P = 3L'(Q_1 Q_2 Q_s) - (Q_1 Q_2 Q_s)' L$$

from equation (1.29) is given by a sixth order polynomial in t

$$P = \sum_{l=0}^{6} P_l(\xi, \eta, \alpha_1)\, t^l \tag{3.8}$$

with

$$\begin{aligned} P_6 &= -2\left(3(\eta-1) + 2(\alpha_1 - 2)\xi^2\right), \qquad (3.9)\\ P_5 &= -4\xi\left(\alpha_1 - 8 + (2-\alpha_1)\eta - 4\alpha_2\xi^2\right),\\ P_4 &= 2(1+\eta-2\eta^2+4(9+2\eta)\xi^2\\ &\quad +\alpha_1((\eta-1)^2 - 2(7+4\eta)\xi^2)), \end{aligned}$$

$$
\begin{aligned}
P_3 &= -8\xi\left(\alpha_1(\eta+\eta^2-2+8\xi^2)-1-\eta(4+\eta)-8\xi^2\right), \\
P_2 &= 2\left[(\eta-1)(\eta(2+\eta)+\alpha_1(1-\eta^2))\right. \\
&\qquad \left.+2(4+7\alpha_1+18\alpha_2\eta)\xi^2\right], \\
P_1 &= -4\xi(\alpha_1+2\eta-9\alpha_1\eta-8\alpha_2)\eta^2-4\alpha_2\xi^2), \\
P_0 &= -6(\alpha_1(\eta-1)-\eta)(\eta-1)\eta+4(2\alpha_2\eta+\alpha_1)\xi^2.
\end{aligned}
$$

We will give the result of the discriminant of this polynomial in the next lemma 20; it is helpful though, to introduce the ‘defect’, i.e. the difference between η and ξ^2, and denote it by X

$$X := \eta - \xi^2 > 0, \tag{3.10}$$

which is positive because of the constraints (1.26).

Lemma 20. *The discriminant $\Delta(P,t)$ of P is non-vanishing for all $0 \leq \alpha_1 \leq 1$ and in the case*

$$
\begin{gathered}
\xi^2 > 0 \textit{ and } X > 0 \\
\textit{or} \\
\xi^2 = 0 \textit{ and } 0 < X \neq 1.
\end{gathered}
\tag{3.11}
$$

Proof. The discriminant of P is given by

$$
\begin{aligned}
\Delta(P,t) = &-589824X\left[(1-X-\xi^2)^2+4\xi^2\right]^7 \\
&\times\left\{\left[\alpha_1(\alpha_2\xi^2+1)+\alpha_2 X\right]\right\}\left[Y(\alpha_1,X,\xi^2)\right]^2
\end{aligned}
$$

All factors except the last are obviously nonzero, so we take a closer look at the last factor, which is a fourth order polynomial in X,

$$Y(\alpha_1, X, \xi^2) = \sum_{k=0}^{4} Y_k(\alpha_1, \xi^2) X^k. \tag{3.12}$$

We are now going to show that each of the coefficients is non-negative and at least one of them is non-vanishing, giving us a positive polynomial. The coefficient

$$Y_4 = \alpha_2^2(16\alpha_2 + \alpha_1^2) \tag{3.13}$$

is zero for $\alpha_1 = 1$ and positive otherwise. The coefficient

$$\begin{aligned} Y_3 = &-4(\alpha_2)^2(3\alpha_1^2 + 4\alpha_1 - 8)\xi^2 \\ &+ 4\alpha_2\left(-3\alpha_1^3 + 67\alpha_1^2 - 196\alpha_1 + 136\right) \end{aligned}$$

is affine in ξ^2. To show that this coefficient is greater than zero, we use that

$$-3\alpha_1^2 - 4\alpha_1 + 8 \geq -3 \cdot 1^2 - 4 + 8 = 1$$

and

$$\begin{aligned} &-3\alpha_1^3 + 67\alpha_1^2 - 196\alpha_1 + 136 > -3 + 66\alpha_1^2 - 198\alpha_1 + 136 \\ &= 66(\alpha_1^2 - 3\alpha_1 + 2) + 1 = 66(2 - \alpha_1)(1 - \alpha_1) \geq 1, \end{aligned}$$

and we get

$$Y_3 \geq 4\,\alpha_2\,(\alpha_2 \xi^2 + 1) \geq 0$$

with $Y_3 = 0$ iff $\alpha_1 = 1$.

Coefficient Y_2: Y_2 is a quadratic polynomial in ξ^2 :

$$\begin{aligned} Y_2 = {} & 2(-13\alpha_1^4 + 34\alpha_1^3 - 21\alpha_1^2 - 8\alpha_1 + 8)\xi^4 \\ & - 2(122\alpha_1^4 - 636\alpha_1^3 + 914\alpha_1^2 - 384\alpha_1 - 16)\xi^2 \\ & - 2\left(13\alpha_1^4 - 26\alpha_1^3 + 405\alpha_1^2 - 392\alpha_1 - 8\right) \end{aligned}$$

in which all terms can be shown to be non-negative.

Coefficient Y_1:

$$\begin{aligned} Y_1 = {} & -4\alpha_1(-4(1+\xi^2)^3 + \alpha_1(1+\xi^2)^2(-71 + 11\xi^2) \\ & + \alpha_1^3(-3 + 61\xi^2 - 61\xi^4 + 3\xi^6) - 2\alpha_1^2(29 - 37\xi^2 - 61\xi^4 + 5\xi^6)) \\ = {} & 4\alpha_1 \Big((-3\,\alpha_1^3 + 10\,\alpha_1^2 - 11\,\alpha_1 + 4)\xi^6 + (61\,\alpha_1^3 - 122\,\alpha_1^2 + 49\alpha_1 + 12)\xi^4 \\ & + (-61\,\alpha_1^3 - 74\,\alpha_1^2 + 131\,\alpha_1 + 12)\xi^2 + 3\,\alpha_1^3 + 58\,\alpha_1^2 + 71\,\alpha_1 + 4 \Big)\,, \end{aligned}$$

which is a third order polynomial in ξ^2. All the coefficients are positive for $\alpha_1 \in (0,1)$.

Coefficient Y_0: The last coefficient is given by

$$\begin{aligned} Y_0 = {} & \alpha_1^2(1+\xi^2)^2 \\ & \times \left[\xi^4(\alpha_2^2 + \xi^2\, 2(1 - \alpha_1^2 + 6\,\alpha_1\alpha_2) + 1 + \alpha_1^2 + 14\alpha_1\right] \geq 0. \end{aligned}$$

This shows that the discriminant is non-zero. □

To determine the number of roots we need to look at one polynomial for conveniently chosen parameters. Choose

$$\xi = 2, \ X = 1, \ \alpha_1 = \frac{1}{2}. \tag{3.14}$$

This choice lets the highest coefficient (3.9) of the polynomial P (3.8) vanish, and gives us

$$\begin{aligned} P &= -64(1+t)(7+8t+8t^2+4t^3+t^4) \\ &= -64(1+t)\left(\left((t+1)^2+2\right)(t+1)^2+3\right) \end{aligned} \tag{3.15}$$

which has exactly one real root.

Lemma 21. *The class of polynomial P defined in (3.8) has at most two real roots for* $\alpha_1 \in [0,1]$ *and X and* η *constraint as in (3.11).*

Proof. Choosing the parameters such as in equation (3.14) gives us a polynomial with one real root, as is shown in equation (3.15). Since the parameters were chosen such that the highest coefficient of the polynomial was vanishing and the discriminant of the polynomial is always non-zero we use theorem 14 to infer that P has at most two real roots. □

We are now prepared to prove theorem 11, which we restate here for mixed states.

Theorem 22. *The class of functions*

$$f_{(\alpha,\xi,\eta)}(t) = \sum_{r=1}^{2} \alpha_r Q_r'(t) \log \frac{Q_r(t)}{\alpha_1 Q_1(t) + \alpha_2 Q_2(t)} \tag{3.16}$$

with parameters constraint by

$$0 \le \xi_r^2 < \eta_r < \infty, \;\; 0 < \alpha_r < 1, \;\; r = 1,2, \;\; \alpha_1 + \alpha_2 = 1.$$

$$(\xi_1, \eta_1) \neq (\xi_2, \eta_2), \tag{3.17}$$

has at most two real roots.

Proof. The second derivative of $f_{(\alpha,\xi,\eta)}(t)$ with respect to t is given by

$$f'' = \alpha_1 \alpha_2 \frac{L}{(Q_1 Q_2 Q_s)^2} P,$$

and Q_1,Q_2 and Q_s remain positive since $\eta_r > \xi_r^2$ for $r = 1,2$. L is a second order polynomial and has therefore at most two real roots. From lemma 21 we know that P has at most two real roots. Since f and f' converge to zero (3.1,3.3) at plus and minus infinity we can apply lemma 19 twice

$$|f^{-1}(0)| \le |(f')^{-1}(0)| - 1 \le |(f'')^{-1}(0)| - 2 \le 2.$$

□

This completes the proof of theorem 11 for mixed states.

3.3 Two Pure States

If one or two states are pure our life gets surprisingly more difficult due to possible discontinuities and non-differentiability. Here we have a closer look at our function (1.18).

The function $\delta_{(k,l)}I$ is given by

$$\delta_{(k,l)}I = \langle k|\,\rho_1\,|l\rangle \log\left(\frac{p_{1k}}{p_{1l}}\frac{p_{\cdot l}}{p_{\cdot k}}\right) + \langle k|\,\rho_2\,|l\rangle \log\left(\frac{p_{2k}}{p_{2l}}\frac{p_{\cdot l}}{p_{\cdot k}}\right). \tag{3.18}$$

This function is well-defined, if p_{1k}, p_{1l}, p_{2k} and p_{2l} are each non zero. It is not possible for p_{1k} and p_{2k} to be simultaneously zero, otherwise the states would be proportional to each other. The same reasoning applies to p_{1l} and p_{2l}. Also, since $|l\rangle$ and $|k\rangle$ are assumed to be distinct, p_{1k} and p_{1l} cannot vanish at the same time, and vice versa for p_{2k} and p_{2l}. Therefore at most either p_{1k} and p_{2l} is zero, or p_{1l} and p_{2k}.

For continuity it is sufficient to show that each term is continuous by itself, in particular that

$$g(\,|k\rangle,\,|l\rangle) := \langle k|\,\rho_1\,|l\rangle \log \langle k|\,\rho_1\,|k\rangle$$

is continuous on the line defined by $\langle k|\,\rho_1\,|k\rangle = 0$ with $|k\rangle \neq 0$, and has a limit of zero.

Since ρ_1 is non negative, this implies $\rho_1\,|k\rangle = 0$. We write

$$\rho_1 = p_1\,|\psi\rangle\langle\psi|$$

and have

$$g(|k\rangle, |l\rangle) = p_1 \langle k|\psi_1\rangle\langle\psi_1|l\rangle \log p_1 |\langle k|\psi_1\rangle|^2$$

which is a composition of the continuous function

$$h(x,y) = xy \log |x|^2$$

and the scalar products with ψ_1. Therefore the function is continuous, albeit it is not everywhere differentiable.

We have

Lemma 23. *In the case p_{1k} is zero p_{2k} has to be non zero and the variation of I (1.18) is*

$$\delta_{(k,l)} I = \langle k|\rho_2|l\rangle \log\left(1 + \frac{p_{1l}}{p_{2k}}\right).$$

This expression is only zero if p_{2l} is zero. The same statement holds if we reverse the role of k and l, or switch ρ_1 and ρ_2.

For the rest of this section we only look at the case that p_{1k} and p_{2k} are both non-zero. In this case we have

$$\xi_r^2 = \eta_r, \ \ r = 1,2.$$

Giving us

$$Q_r(t) = (t + \xi_r)^2, \ \ r = 1,2.$$

Set $\xi_2 = 0$ by using translation invariance, and label the translated value of ξ_1 by ξ. We have for the first and second derivative of (1.25)

$$\begin{aligned} f'(t) &= 2\left(\sum_{l=1}^{2} \alpha_l \log \frac{Q_l}{Q_s}\right) + \alpha_1\alpha_2 \frac{(Q_1'Q_2 - Q_2'Q_1)^2}{Q_sQ_1Q_2} \\ &= 2\left(\sum_{l=1}^{2} \alpha_l \log \frac{Q_l}{Q_s}\right) + 4\alpha_1\alpha_2\xi^2 Q_s^{-1}, \qquad (3.19) \\ f''(t) &= -\frac{2\alpha_1\alpha_2\xi^3\left((\alpha_1-\alpha_2)t^2 + 2\alpha_1\xi t + \alpha_1\xi^2\right)}{t\,(t+\xi)\,Q_s^2}. \qquad (3.20) \end{aligned}$$

The first derivative (3.19) has precisely two poles due to the argument of the logarithm approaches zero. These poles are located at $t = 0$ and $t = -\xi$.

The denominator of the second derivative (3.20) has exactly two distinct simple zeros, at $t = 0$ and $t = -\xi$. To see if the location of the poles can coincide with the location of the roots of the numerator of (3.20), we define

$$h(t) := (\alpha_1 - \alpha_2)t^2 + 2\alpha_1\xi t + \alpha_1\xi^2,$$

and observe that

$$\begin{aligned} h(0) &= \alpha_1\xi^2 \\ h(-\xi) &= -\alpha_2\xi^2, \end{aligned}$$

which would imply ξ^2=0 in the case of h vanishing at one of these points, which is excluded since otherwise the states would be proportional to each other.

On a side note, we notice that there is a mild symmetry in the the parameters of the function, the function with parameters $\xi_{new} = -\xi$ and $\alpha_{new} = 1 - \alpha$ is given by a simple translation in the t-variable by

$$f_{\alpha,\xi}(t-\xi) = f_{1-\alpha,-\xi}(t).$$

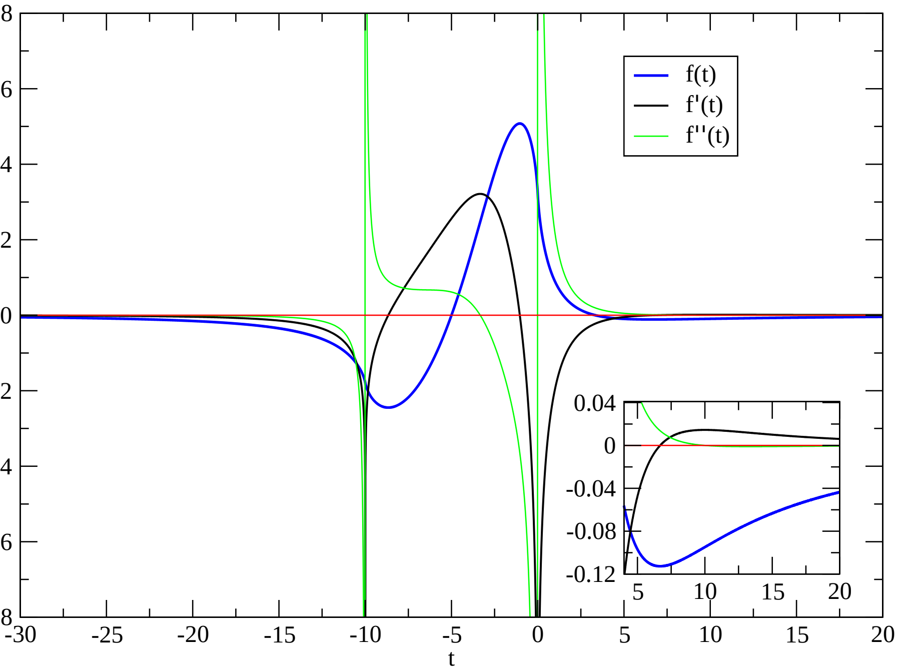

Figure 3.2: Function f defined in (1.25) in blue, its first derivative in black and its second derivative in green for $\alpha_1 = 0.2$ and $\xi = 10$. The inset shows a magnified region.

The approach to infinity is given by

$$\lim_{t\to\pm\infty} f''(t) = 0_{sign(\xi(\alpha_2-\alpha_1))},$$

in case that $\alpha_1 \neq \alpha_2$. In case that $\alpha_1 = \alpha_2$, we have

$$\lim_{t\to\pm\infty} f''(t) = 0_{\mp sign(\xi)}.$$

Since

$$h(0) = \alpha_1 \xi^2 > 0 \text{ and } h(-\xi) = -\alpha_2 \xi^2 < 0$$

one of the zeros is between the poles and another (if it exists) outside.

Lemma 24. *The derivative of f with respect to t has two poles and at most three real roots.*

Proof. From direct inspection we see that f' has one pole at $t = -\xi$ and one at $t = 0$. We know from the considerations above, that f'' has at most two real roots and two poles at $t = -\xi$ and $t = 0$, where exactly one root is between the poles and the other root is outside if it exists. From (3.3) we know that f' converges to zero in the limit of plus and minus infinity. From lemma 17 we know f' has at most two real roots between the poles, and from the same lemma at most one outside the poles. □

We are in good shape to prove theorem (11) in the case of two pure states, which reads in this case

Theorem 25. *The class of functions*

$$f_{(\alpha,\xi,\eta)}(t) = \sum_{r=1}^{2} \alpha_r Q_r'(t) \log \frac{Q_r(t)}{\alpha_1 Q_1(t) + \alpha_2 Q_2(t)} \tag{3.21}$$

with parameters constraint by

$$\begin{gathered} 0 \le \xi_r^2 < \infty, \;\; 0 < \alpha_r < 1, \;\; r = 1,2, \\ \alpha_1 + \alpha_2 = 1, \;\; \xi_1 \neq \xi_2, \end{gathered} \tag{3.22}$$

and $\eta_r = \xi_r^2$ for $r = 1,2$, has at most two real roots.

Proof. In this proof we work with the translated function f, i.e. we only need to consider the case $\xi_2 = 0$. The function $f_{(\alpha,\xi,\eta)}(t)$ converges to zero at plus and minus infinity. The function is continuous, and at the poles of its derivative has values

$$f(0) = -2\alpha_1 \xi \log(\alpha_1),$$
$$f(-\xi) = 2\alpha_2 \xi \log(\alpha_2).$$

These are not maxima nor minima since the left and the right limit both converge to minus infinity, i.e.

$$\lim_{t\to-\xi^{\pm}} f'(t) = -\infty,$$
$$\lim_{t\to0^{\pm}} f'(t) = -\infty.$$

With help of lemma 24 we see that f has at most three extrema and therefore by lemma 18 at most two real roots. □

3.4 One Pure State and One Mixed State

In this case, a similar analysis of continuity as in the case of two pure states holds.

Choose ρ_1 to be pure. We have

$$\xi_1^2 = \eta_1 \rightarrow Q_1 = (t+\xi_1)^2.$$

Using translation and scale invariance, we can set

$$Q_2 = t^2 + 1,$$

and get for the first and second derivative of (1.25), after labeling the translated and rescaled value of ξ_1 as ξ

$$\begin{aligned} L &= 2(t+\xi)(1-t\xi), \\ f'(t) &= 2\left(\sum_{l=1}^{2}\alpha_l \log\frac{Q_l}{Q_s}\right) + \alpha_1\alpha_2\frac{(Q_1'Q_2 - Q_2'Q_1)^2}{Q_sQ_1Q_2} \\ &= 2\left(\sum_{l=1}^{2}\alpha_l \log\frac{Q_l}{Q_s}\right) + 4\alpha_1\alpha_2(1-t\xi)^2 Q_s^{-1}Q_2^{-1}, \\ f''(t) &= 4\alpha_1\alpha_2\frac{(1-t\xi)P}{Q_2^2(t+\xi)Q_s^2}, \\ P &= \left((2\alpha_1-1)\xi^2+3\right)t^4 + 2\xi\left(\alpha_1(\xi^2+1)+4\right)t^3 \\ &\quad + \left(\alpha_1(\xi^4-1)+2(6\alpha_1+1)\xi^2+2\right)t^2 \\ &\quad + 2\xi\left(\alpha_1\left(5\xi^2-3\right)+4\right)t + \alpha_1 \\ &\quad + \xi^2\left(\alpha_1\left(3\xi^2-2\right)+3\right) - 1. \end{aligned} \tag{3.23}$$

In this case we have one simple pole for the second derivative at $t = -\xi$ and an obvious zero at $t = 1/\xi$. Figure 3.3 shows f and its first derivative, while figure 3.4 shows the first and second derivative for illustration for one typical value of α_1 and ξ. To see that the poles and the roots of f'' cannot coincide we evaluate for the numerator of f''

$$(1+\xi^2)P(-\xi) = -\alpha_2\left(\xi^2+1\right)^4$$

which is strictly non zero.

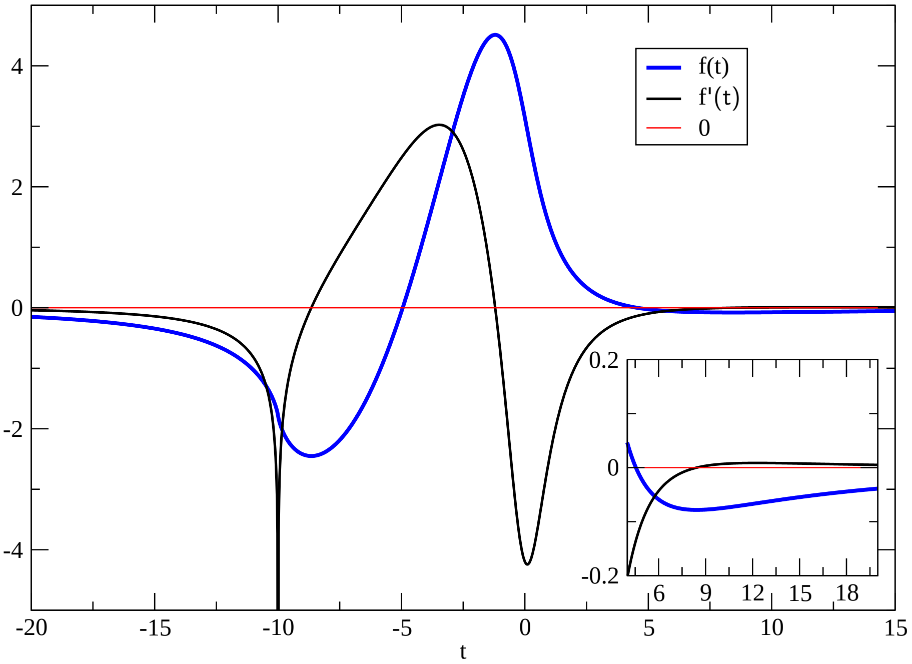

Figure 3.3: Function f defined in (1.25) in blue and its first derivative in black for $\alpha = 0.2$ and $\xi = 10$. The inset shows a magnified region.

The discriminant of P is

$$\Delta(P,t) = -48\,\alpha_2\left(\xi^2+1\right)^6\left(\alpha_1\xi^2+1\right) \\ \times\left(\alpha_1^2\xi^4+\alpha_1(16-14\alpha_1)\xi^2+\alpha_1^2+16\alpha_2\right)^2$$

which is always smaller than zero. Setting $\alpha_1 = 1/3$ and $\xi = 3$ we get

$$P = \frac{4}{3}\left(33t^3+62t^2+81t+76\right) = 44\left(t+\frac{4}{3}\right)\left(\left(t+\frac{3}{11}\right)^2+\frac{191}{121}\right) \qquad (3.24)$$

which has exactly one real root. With help from the next lemma we conclude that the second derivative f has at most three roots and one pole.

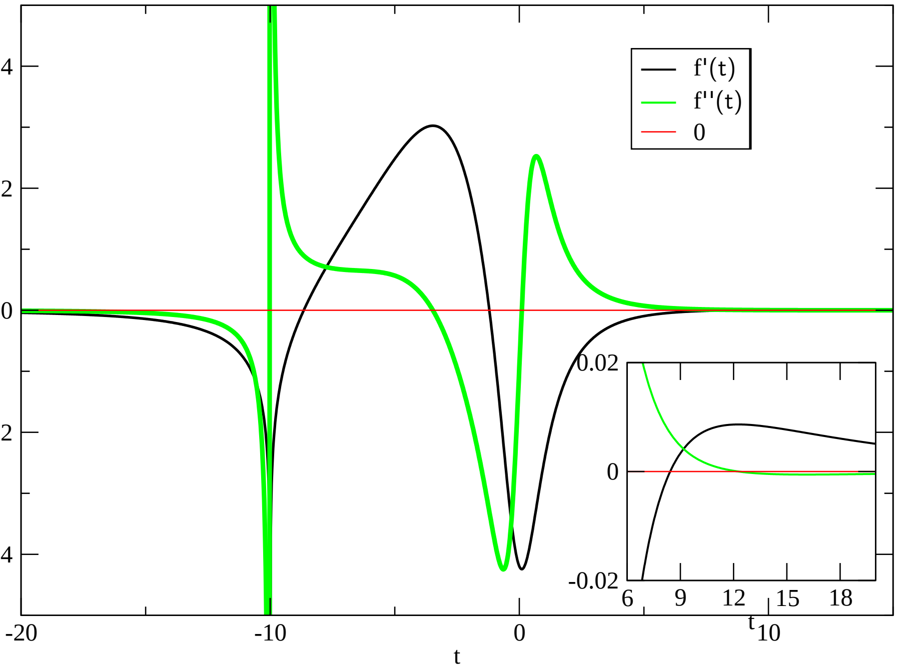

Figure 3.4: First derivative of f shown in (3.23) in black and second derivative in green for $\alpha = 0.2$ and $\xi = 10$. The inset shows a magnified region

Lemma 26. *The class of polynomial P defined in (3.8) has at most two real roots.*

Proof. Choosing the parameters $\alpha = 1/3$ and $\xi = 3$ gives us a polynomial with one real root, as is shown in equation (3.24). Since the parameters were chosen such that the highest coefficient of the polynomial vanishes and the discriminant of the polynomial is always non-zero we use theorem 14 to infer that P has at most two real roots. □

Lemma 27. *The derivative of f has one pole and at most three real roots.*

Proof. From direct inspection of equation (3.23) we see that f' has exactly one pole at $t = -\xi$. We know from lemma 26 and the considerations above, that f'' has at most three real roots and one pole at $t = -\xi$ which cannot coincide with one of the roots. From equation (3.3) we know that f' converges to zero in the limit of plus and minus infinity. Since our discussion does not change if we change t to $-t$ we are left with two cases.

1. All roots are on the right side of the pole. Considering the interval $(-\infty,-\xi)$ there cannot be any roots due to lemma 17. Looking at the interval $(-\xi,\infty)$ there are at most three roots due to the same lemma.

2. One root is on the left side of the pole. A similar analysis as in the previous case shows that there are is at most one root on the left and two on the right side of the pole.

□

Theorem 28. *The class of functions*

$$f_{(\alpha,\xi,\eta)}(t) = \sum_{r=1}^{2} \alpha_r Q_r'(t) \log \frac{Q_r(t)}{\alpha_1 Q_1(t) + \alpha_2 Q_2(t)} \tag{3.25}$$

with parameters constraint by

$$0 \le \xi_r^2 < \infty, \ \ 0 < \alpha_r < 1, \ \ r = 1,2, \ \ \alpha_1 + \alpha_2 = 1.$$
$$\xi_1 \neq \xi_2, \tag{3.26}$$

and $\eta_r = \xi_r^2$ *for* $r = 1,2$*, has at most two real roots.*

Proof. In this proof we work with the translated and rescaled function f, i.e. we only need to consider with $\xi_2 = 0$ and $\eta_2 = 1$. From the preceding discussion we see that

The function $f_{(\alpha,\xi,\eta)}(t)$ converges to zero at plus and minus infinity as shown in 3.1. Although f' has a pole at $t=-\xi$, f is finite and non zero at $t=-\xi$ with value

$$f(-\xi)=2\alpha_2\xi\log(\alpha_2).$$

This is not a maximum of minimum since the left and the right limit both converge to minus infinity, i.e.

$$\lim_{t\to-\xi^{\pm}} f'(t)=-\infty.$$

Therefore f has at most three extrema and by lemma 18 at most two real roots. □

3.5 The Proof

In this part we finally show our central result. Let us recall, the following equations have to be solved simultaneously:

$$\begin{aligned}
\delta_{(1,2)}I &= \sum_{r=1}^{2}\langle 1|\rho_r|2\rangle\log\left(\frac{p_{r1}}{p_{r2}}\frac{p_{\cdot 2}}{p_{\cdot 1}}\right)=0,\\
\delta_{(1,3)}I &= \sum_{r=1}^{2}\langle 1|\rho_r|3\rangle\log\left(\frac{p_{r1}}{p_{r3}}\frac{p_{\cdot 3}}{p_{\cdot 1}}\right)=0,\\
\delta_{(2,3)}I &= \sum_{r=1}^{2}\langle 2|\rho_r|3\rangle\log\left(\frac{p_{r2}}{p_{r3}}\frac{p_{\cdot 3}}{p_{\cdot 2}}\right)=0.
\end{aligned}\tag{3.27}$$

From our previous analysis we know that if we keep one state fixed, we have at most two solutions for the second state for each individual equation (plus the trivial one that both vectors are proportional). It is important to note, that if any two rank-1 outcomes are proportional, the third one must be orthogonal to them to form a POVM and our system would be equivalent to an orthogonal measurement.

One of the solutions can actually be found by hand, and it is given when both logarithms vanish simultaneously. The argument of the logarithm has the peculiar property that if one of them is one, the other one is as well,

$$\frac{p_{11}}{p_{12}}\frac{p_{\cdot 2}}{p_{\cdot 1}} = 1 \quad \leftrightarrow \quad \frac{p_{21}}{p_{22}}\frac{p_{\cdot 2}}{p_{\cdot 1}} = 1.$$

Theorem 29. *If the alphabet consists of two states of a qubit, then every stationary point of the mutual information which is not a minimum, is a von Neumann measurement.*

Proof. Assume that the mutual information is stationary and that POVM is not von Neumann. We start by analyzing the special case that $p_{1k} = 0$, or $p_{2k} = 0$ for $k = 1,2$, or 3 which only happens if at least one of the states is pure. Say $p_{11} = 0$, it follows from lemma 23 that p_{22} and p_{23} must be zero as well. This is only possible if $|3\rangle$ is proportional to $|2\rangle$, which implies it must be an orthogonal measurement.

For all the other cases we can assume that $p_{rk} \neq 0$ for all r,k. Observe that in (3.27) if one logarithm is zero, automatically the other is zero as well. Since $|2\rangle$ and $|3\rangle$ have to be distinct, Theorem 11 tells us that one of these states must set the logarithm to zero, say $|3\rangle$. This means that

$$\frac{p_{\cdot 1}p_{13}}{p_{\cdot 3}p_{11}} = 1 \leftrightarrow \frac{p_{21}}{p_{11}} = \frac{p_{23}}{p_{13}},$$

so outcome one and outcome three are equivalent. Since the same reasoning is applicable to $\langle 2|$ instead of $\langle 1|$ in equations (1.19,1.21) we find that all outcomes are equivalent and we are in a minimum. □

Corollary 30. *The orthogonal measurement conjecture is true for all states* ρ_0 *and* ρ_1 *if they can be mutually diagonalized apart from a qubit, i.e. if a basis exists such that* ρ_0 *is diagonal and* ρ_1 *diagonal except on a two dimensional subspace.*

In particular this includes the case that both states are states of a qubit.

Proof. Using theorem 8 we can build an optimal measurement by using optimal measurements for the independent blocks. Theorem 7 tells us, that for the commuting part an orthogonal measurement is sufficient. For the qubit part any maximum must be a stationary point of the mutual information, and from theorem 29 we know this is only possible for an orthogonal measurement. □

3.6 Finding the Maximum

Now that the type of POVM which maximizes the mutual information is found, we ask the question where this maximum is. Since the equation in question is transcendental it is in general not possible to find analytical solutions. For special cases a solution was found by Fuchs and Caves [25]. See also section 11.6.1 in Suzuki *et al.* [6] about this matter.

Since we established that the optimal measurement is a von Neumann measurement we have to look for the condition that the variation of the mutual information (1.21) is zero at $t = 0$, i.e.

$$\delta I = 2\sum_{r=1}^{2} \alpha_r \xi_r \log\left(\frac{\eta_r}{\alpha_1\eta_1 + \alpha_2\eta_2}\right) = 0.$$

Let us express that in terms of the matrix coefficients

$$\delta I = 2\sum_{r=1}^{2} \langle 1|\,\rho_r\,|0\rangle \log\left(\frac{\frac{\langle 0|\,\rho_r\,|0\rangle}{\langle 1|\,\rho_r\,|1\rangle}}{\frac{\langle 0|\,(\rho_1+\rho_2)\,|0\rangle}{\langle 1|\,(\rho_1+\rho_2)\,|1\rangle}}\right). \tag{3.28}$$

Parametrizing $|0\rangle$ and $|1\rangle$ by

$$|0\rangle = \begin{pmatrix} 1 \\ s \end{pmatrix}, \quad |1\rangle = \begin{pmatrix} s \\ -1 \end{pmatrix},$$

we get for the right hand side of (3.28), assuming a real matrix representation

$$\begin{aligned} &2\left[(s(\rho_1^{00}-\rho_1^{11})+(s^2-1)\rho_1^{01})\times\right.\\ &\left(\log\left(\frac{s^2\rho_2^{00}-2s\rho_2^{01}+\rho_2^{11}}{s^2\rho_1^{00}-2s\rho_1^{01}+\rho_1^{11}}+1\right)-\log\left(\frac{s^2\rho_2^{11}+2s\rho_2^{01}+\rho_2^{00}}{s^2\rho_1^{11}+2s\rho_1^{01}+\rho_1^{00}}+1\right)\right)\\ &\left.+\rho_1\leftrightarrow\rho_2\right], \end{aligned}$$

where the upper indices denote the matrix element in the standard basis.

The structure of this function is quite complicated as figure 3.5 indicates. From the graph we see that there are two maxima and two minima, which allows for more roots according to our analysis. This situation can be traced back to the fact that we did not normalize the outcomes $|0\rangle$ and $|1\rangle$, i.e. we are missing a factor of $(s^2+1)^{-1}$; if we include this factor the function does not have superfluous extrema. Though, if we include this factor, multiple differentiation of the function does not get rid of the logarithm. Our approach does not seem to be viable for this problem.

From numerical experiments we know that there do not exist more than two solutions. Unfortunately this cannot be shown by our method, thus giving a clearer

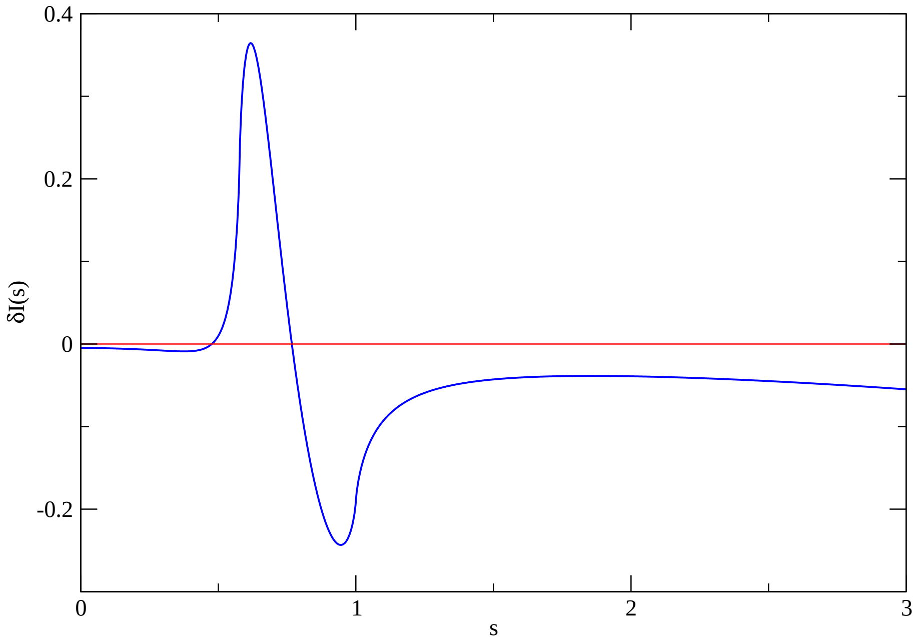

Figure 3.5: Variation of the mutual information for von Neumann measurements as described in the text. Both states are pure and we have $p_1 = 0.2$, $\rho_1^{00} = 0.1$ and $\rho_2^{00} = 0.6$.

view on its limitations.

We close this chapter with a conjecture about the number of stationary points of the mutual information when restricted to von Neumann measurements.

Conjecture 2. *For two states of a qubit, there exists only two stationary points of the mutual information if the the number of outcomes of the measurements is restricted to two and both lie in the same plane as the states in the Bloch representation. One of the stationary points is the global minimum and the other one is the global maximum.*

Chapter 4

Outlook

In this thesis we have proved the orthogonal measurement conjecture for states of a qubit. This gives immediate rise to a couple of questions. Firstly, since the proof has been very technical, the proof sheds not much light on the question *why* the theorem is true. It almost seems accidental for the theorem to be true. We do not believe in an accident for this case, so the question is, *is there a simpler proof which reveals more about the underlying structure of the problem?* We were not able to answer this question, but it could be that the following formula might give a hint to the right direction

$$\frac{d}{dt}\left(\alpha_1 Q_1 \log\frac{Q_1}{Q_s} + \alpha_2 Q_2 \log\frac{Q_2}{Q_s}\right) = f_{(\alpha,\xi,\eta)}(t).$$

The second question, is how to show only one maximum and one minimum exists if we restrict ourselves to orthogonal measurements. This result would be extremely valuable since it would allow to turn numerical results into rigorous estimates. Also, it would allow us to conclude that the cases in the 'solvable' case are actually the true solutions.

The next question is if the conjecture is also true in case the states are qutrits or qunits. It is illustrative to see where mimicking the proof for qubits fails in case of qutrits. For two general states of a qutrit it is not always possible to choose a common basis such that both states have a real matrix representation. Setting this problem aside, and just assuming that both states are real, the D-SBJOH theorem tells us we need at most $d(d+1)/2$ outcomes, which in the case of qutrits means six. The same equation as (1.18) can be derived, i.e.

$$\delta_{(k,l)} I = \sum_{r=1}^{2} \langle k| \rho_r |l\rangle \log\left(\frac{p_{rk}}{p_{rl}} \frac{p_{\cdot l}}{p_{\cdot k}}\right) = 0.$$

But the parametrization of the vectors would be significantly different

$$|n\rangle = \beta_0(n)\,|0\rangle + \beta_1(n)\,|1\rangle + \beta_2(n)\,|2\rangle.$$

Again, one of these parameters is superfluous, but the remaining parameters will lead to a one-dimensional family of solution on a two-dimensional surface. In our proof of the qubit case we had zero-dimensional solutions on a one-dimensional curve, which allows us to use real analysis to determine the number of solutions and then make statements about mutual roots of the equations. In the present case we are in deeper trouble. A great deal of mathematical work has been devoted to mutual roots of algebraic curves in the field of algebraic geometry, far less is known about transcendental curves. This road does not seem to be feasible to follow.

In a broader perspective, this work is also a tiny step to the more general question of *how many outcomes do we need*. In a setting with m-qunits, how many outcomes are sufficient to achieve the accessible information?

Lastly, but not least, we would like to state a conjecture, which might help to proof the general orthogonal measurement conjecture and which would be paramount for gaining confidence in numerical results. The question is, what if we vary the allowed number of outcomes, if we are below the optimal number, we believe that the accessible information is strictly increasing with the number of outcomes:

Conjecture 3. *The maximal information is strictly increasing in the numbers of outcomes for fixed states until the global accessible information is reached.*

$$\max_{\{\Pi_i\}_{i\leq N}} I = \max_{\{\Pi_i\}_{i\leq N+1}} I \rightarrow \max_{\{\Pi_i\}_{i\leq N}} I = I_{acc}$$

This would be an extremely convenient statement. The general problem for large Hilbert-spaces is that the maximum number of outcomes according to the D-SBJOH theorem increases with d^2 so the total memory needed increases with d^3 for pure outcomes, and computation times usually scale worse. This conjecture might also offer advantages for a general proof of the orthogonal measurement hypothesis.

With this we conclude this thesis. We hope reading it was as enjoyable as obtaining the result was, and that the reader might be able to contribute to these open questions.

Bibliography

[1] Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[2] Jacob Wolfowitz. The coding of messages subject to chance errors. *Illinois Journal of Mathematics*, 1:591–606, 1957.

[3] Michael Aaron Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[4] Alexander S. Holevo. Statistical detection theory for quantum systems. *Journal of Multivariate Analysis*, 3:337–394, 1973.

[5] Jaroslav Řeháček, Berthold-Georg Englert, and Dagomir Kaszlikowski. Iterative procedure for computing accessible information in quantum communication. *Phyical Review A*, 71:054303, 2005.

[6] Jun Suzuki, Syed M. Assad, and Berthold-Georg Englert. Accessible information about quantum states: An open optimization problem. In Goong Chen, Louis Kauffman, and Samuel J. Lomonaco, editors, *Mathematics of Quantum Computation and Quantum Technology*. Chapman Hall, 2007.

[7] Kean Loon Lee, Wee Kang Chua, Shiang Yong Looi, and Berthold-Georg Englert. Somim: An open-source program code for the numerical search for optimal measurements by an iterative method. arXiv:0805.2847.

[8] Alexander S. Holevo. Information-theoretical aspects of quantum measurement. *Problems of Information Transmission*, 9(2):110–118, 1973.

[9] Lev B. Levitin. Optimal quantum measurements for two pure and mixed states. In V.P. Belavkin, O Hirota, and R.L. Hudson, editors, *Quantum Communications and Measurement*, pages 439–447, 1995.

[10] Peter W. Shor. On the number of elements in a POVM attaining the accessible information. 2000, arXiv:quant-ph/000907.

[11] Jaroslav Řeháček and Berthold-Georg Englert. How well can you know the edge of a quantum pyramid? 2009, arXiv:0905.0510.

[12] Imre Csiszár and János Körner. Broadcast channels with confidential messages. *IEEE Transactions on Information Theory*, 24:339–348, 1978.

[13] Berthold-Georg Englert, Dagomir Kaszlikowski, Hui Khoon Ng, Wee Kang Chua, Jaroslav Řeháček, and Janet Anders. Efficient and robust quantum key distribution with minimal state tomography. arXiv:quant-ph/0412075.

[14] Christopher A. Fuchs, Nicolas Gisin, R.B. Griffiths, Chi-Sheng Niu, and Asher Peres. Optimal eavesdropping in quantum cryptography. i. information bound and optimal strategy. *Physical Review A*, 56(2):1163–1172, 1997.

[15] Christopher A. Fuchs. Distinguishability and accessible information in quantum theory, 1996, arXiv:quant-ph/9601020.

[16] Gilles Brassard and Louis Salvail. Secret-key reconciliation by public discussion. In *Lecture Notes in Computer Science*. Springer, 1994.

[17] E. Brian Davies. Information and quantum measurement. *IEEE Transactions on Information Theory*, 24(5):596–599, 1978.

[18] John von Neumann. *Mathematical foundations of quantum mechanics*. Princeton University Press, Princeton, NJ, 1955.

[19] Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9(3):177–183, 1973.

[20] Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998.

[21] Paul Hausladen, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William Wooters. Classical information capacity of a quantum channel. *Physical Review A*, 54:1869–1876, 1996.

[22] Matthew B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5:255–257, 2009.

[23] Masahide Sasaki, Stephen M. Barnett, Richard Jozsa, Masao Osaki, and Osamu Hirota. Accessible information and optimal strategies for real symmetrical quantum sources. *Physical Review A*, 59:3325–3335, 1999.

[24] Bartel Leendert van der Waerden. *Algebra I & II*. Springer-Verlag, New York, 1990.

[25] Christopher A. Fuchs and Carlton M Caves. Ensemble-dependent bounds for accessible information in quantum mechanics. *Physical Review Letters*, 73:3047–3050, 1994.

Appendix A

Variation Equations in Bloch Representation

In the following we will derive the variation equations (1.19) by using the Bloch-representation for qubit states. In two dimensions we have the Pauli-matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

These matrices are hermitian and trace-free. Together with the identity they form a real basis of the space of all hermitian two-by-two matrices. Any state of a qubit ρ can be expanded

$$\rho = \frac{1}{2}\left(\mathbb{I} + r_1\sigma_1 + r_2\sigma_2 + r_3\sigma_3\right) =: \frac{1}{2}(\mathbb{I} + \vec{r}\cdot\vec{\sigma}),$$

where $\vec{r}$ denotes a real, three dimensional vector. The condition that states have unit trace is already implemented. The positivity condition translates to

$$|\vec{r}| \leq 1$$

and we have a pure state iff $|\vec{r}| = 1$. For a POVM we also use the Bloch vector representation. We use a three rank-1 outcome POVM, which by the D-SBJOH theorem (6) is sufficient. Define

$$\begin{aligned}
\Pi_1 &:= a\,(\mathbb{I} + \vec{n}_1 \cdot \vec{\sigma})\,, \\
\Pi_2 &:= b\,(\mathbb{I} + \vec{n}_2 \cdot \vec{\sigma})\,, \\
\Pi_3 &:= c\,(\mathbb{I} + \vec{n}_3 \cdot \vec{\sigma})\,, \;\; a,b,c > 0.
\end{aligned}$$

For this to be a POVM the following has to hold

$$\Pi_3 = \mathbb{I} - \Pi_1 - \Pi_2 = (1-a-b)\left(\mathbb{I} - \frac{a\vec{n}_1 + b\vec{n}_2}{1-a-b} \cdot \vec{\sigma}\right)$$

Where $\vec{n}_1$ and $\vec{n}_2$ denote unit vectors and

$$1-a-b \geq 0 \text{ and } \left|\frac{a\vec{n}_1 + b\vec{n}_2}{1-a-b}\right|^2 = 1$$

has to hold. The second condition is equaivalent to

$$2\,ab\,\vec{n}_1 \cdot \vec{n}_2 = 1 - 2a - 2b + 2ab. \tag{A.1}$$

We also have for Π_j to be a POVM

$$\vec{n}_3 = -\frac{a\vec{n}_1 + b\vec{n}_2}{1-a-b}.$$

The mutual information is given by

$$I = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{p_{\cdot j} p_{i\cdot}}$$

and its variation

$$\delta I = \sum_{i,j} \delta p_{ij} \log \frac{p_{ij}}{p_{\cdot j} p_{i\cdot}} = \sum_{i,j} \delta p_{ij} \log \frac{p_{ij}}{p_{\cdot j}}$$

The joint probability matrix is given by

$$\begin{aligned}
p_{11} &= ap_1(1+\vec{r}_1\cdot\vec{n}_1),\\
p_{12} &= bp_1(1+\vec{r}_1\cdot\vec{n}_2),\\
p_{13} &= p_1\left(1-a(1+\vec{r}_1\cdot\vec{n}_1)-b(1+\vec{r}_1\cdot\vec{n}_2)\right),\\
p_{21} &= ap_2(1+\vec{r}_2\cdot\vec{n}_1),\\
p_{22} &= bp_2(1+\vec{r}_2\cdot\vec{n}_2),\\
p_{23} &= p_2\left(1-a(1+\vec{r}_2\cdot\vec{n}_1)-b(1+\vec{r}_2\cdot\vec{n}_2)\right).
\end{aligned}$$

We are using the method of Lagrange multipliers to implement the constraint (A.1). The variation is restricted by

$$\underbrace{(b\vec{n}_1\cdot\vec{n}_2+1-b)}_{X}\delta a + \underbrace{(a\vec{n}_1\cdot\vec{n}_2+1-a)}_{Y}\delta b + ab\delta\vec{n}_1\cdot\vec{n}_2 + ab\vec{n}_1\cdot\delta\vec{n}_2 = 0 \qquad \text{(A.2)}$$

Observe, that

$$X = (1-2b)/(2a), \qquad Y = (1-2a)/(2b)$$

leading to

$$XY = \frac{1}{2}\left(\vec{n}_1 \cdot \vec{n}_2 + 1\right) \qquad 0 \leq XY \leq 1.$$

For the unrestricted variation we would get

$$\begin{aligned}
\delta I = {} & \delta p_{11} \log\left(\frac{p_{11}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{13}}\right) + \delta p_{12} \log\left(\frac{p_{12}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{13}}\right) + \delta p_{21} \log\left(\frac{p_{21}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{23}}\right) \\
& + \delta p_{22} \log\left(\frac{p_{22}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{23}}\right), \\
\delta_a I = {} & \left(p_1(1+\vec{r}_1\cdot\vec{n}_1)\log\frac{p_{11}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{13}} + p_2(1+\vec{r}_2\cdot\vec{n}_1)\log\frac{p_{21}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{23}} \right)\delta a, \\
\delta_{\vec{n}_1} I = {} & a\left(p_1\vec{r}_1\cdot\delta\vec{n}_1\log\frac{p_{11}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{13}} + p_2\vec{r}_2\cdot\delta\vec{n}_1\log\frac{p_{21}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{23}} \right), \\
\delta_b I = {} & \left(p_1(1+\vec{r}_1\cdot\vec{n}_2)\log\frac{p_{12}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{13}} + p_2(1+\vec{r}_2\cdot\vec{n}_2)\log\frac{p_{22}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{23}} \right)\delta b, \\
\delta_{\vec{n}_2} I = {} & b\left(p_1\vec{r}_1\cdot\delta\vec{n}_2\log\frac{p_{12}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{13}} + p_2\vec{r}_2\cdot\delta\vec{n}_2\log\frac{p_{22}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{23}} \right).
\end{aligned}$$

Solving the differential constraints (A.2) for δa and expressing δI, the restricted variation is:

$$\begin{aligned}
\delta I = {} & \left(p_1(1+\vec{r}_1\cdot\vec{n}_2)\log\frac{p_{12}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{13}} + p_2(1+\vec{r}_2\cdot\vec{n}_2)\log\frac{p_{22}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{23}} \right)\delta b \\
& - \left[\left(p_1(1+\vec{r}_1\cdot\vec{n}_1)\log\frac{p_{11}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{13}} + p_2(1+\vec{r}_2\cdot\vec{n}_1)\log\frac{p_{21}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{23}} \right) \right. \\
& \left. \cdot \left(\frac{Y}{X}\delta b + \frac{ab}{X}\vec{n}_1\delta\vec{n}_2 + \frac{ab}{X}\vec{n}_2\delta\vec{n}_1 \right) \right]
\end{aligned}$$

$$+a\left(p_1\vec{r}_1\log\frac{p_{11}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{13}}+p_2\vec{r}_2\log\frac{p_{21}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{23}}\right)\cdot\delta\vec{n}_1+$$
$$+b\left(p_1\vec{r}_1\log\frac{p_{12}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{13}}+p_2\vec{r}_2\log\frac{p_{22}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{23}}\right)\cdot\delta\vec{n}_2. \tag{A.3}$$

Define

$$\vec{v}:=p_1(\vec{n}_1+\vec{r}_1)\log\overbrace{\frac{p_{11}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{13}}}^{q_{11}}+p_2(\vec{n}_1+\vec{r}_2)\log\overbrace{\frac{p_{21}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{23}}}^{q_{21}}, \tag{A.4}$$
$$\vec{w}:=p_1(\vec{n}_2+\vec{r}_1)\log\underbrace{\frac{p_{12}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{13}}}_{q_{12}}+p_2(\vec{n}_2+\vec{r}_2)\log\underbrace{\frac{p_{22}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{23}}}_{q_{22}}. \tag{A.5}$$

Since the variations of n_1 are restricted to orthogonal transformations, we have

$$\delta n_1 = n_1\times\delta n.$$

So we get from setting the variation to zero and (A.3)

$$\delta\vec{n}_1\cdot\left(\vec{v}-\vec{v}\cdot\vec{n}_1\frac{b}{X}\vec{n}_2\right)=0, \tag{A.6}$$
$$\delta\vec{n}_2\cdot\left(\vec{w}-\vec{v}\cdot\vec{n}_1\frac{a}{X}\vec{n}_1\right)=0, \tag{A.7}$$
$$\delta b\cdot(X\vec{w}\cdot\vec{n}_2-Y\vec{v}\cdot\vec{n}_1)=0. \tag{A.8}$$

To solve these equations we write

$$\vec{v}=v^1\vec{n}_1+v^2\vec{n}_2,\quad \vec{w}=w^1\vec{w}_1+w^2\vec{w}_2$$

applying (A.6) shows $v^2=(\vec{v}\cdot\vec{n}_1)\frac{b}{X}$, substituting this and computing $\vec{n}_1\cdot\vec{v}=v^1+\vec{n}_1\cdot\vec{n}_2\frac{\vec{v}_1\cdot\vec{n}_1 b}{X}$, leading to $v^1=(\vec{v}\cdot\vec{n}_1)\frac{1-b}{X}$. Now expanding $\vec{w}$ and applying (A.7) leads to $w^1 = (\vec{v}\cdot\vec{n}_1)\frac{a}{X}$, and computing $\vec{w}\cdot\vec{n}_2$ in conjuncture with (A.8) leads us to the solution of these equations

$$\vec{w} = \frac{\vec{v}\cdot\vec{n}_1}{X}\left(a\vec{n}_1 + (1-a)\vec{n}_2\right), \tag{A.9}$$

$$\vec{v} = \frac{\vec{v}\cdot\vec{n}_1}{X}\left((1-b)\vec{n}_1 + b\vec{n}_2\right). \tag{A.10}$$

Observe the following:

$$\begin{aligned}(\vec{n}_2+\vec{n}_3)\cdot\vec{w} &= 0,\\ (\vec{n}_1+\vec{n}_3)\cdot\vec{v} &= 0,\\ (\vec{n}_1+\vec{n}_2)\cdot(\vec{v}-\vec{w}) &= 0.\end{aligned}$$

leading to

$$\sum_j p_j\left(1+\vec{n}_2\cdot\vec{n}_3+\vec{r}_j\cdot(\vec{n}_2+\vec{n}_3)\right)\log\left(\frac{p_{j2}}{p_{\cdot 2}}\frac{p_{\cdot 3}}{p_{j3}}\right) = 0, \tag{A.11}$$

$$\sum_j p_j\left(1+\vec{n}_1\cdot\vec{n}_3+\vec{r}_j\cdot(\vec{n}_1+\vec{n}_3)\right)\log\left(\frac{p_{j1}}{p_{\cdot 1}}\frac{p_{\cdot 3}}{p_{j3}}\right) = 0, \tag{A.12}$$

$$\sum_j p_j\left(1+\vec{n}_1\cdot\vec{n}_2+\vec{r}_j\cdot(\vec{n}_1+\vec{n}_2)\right)\log\left(\frac{p_{j1}}{p_{\cdot 1}}\frac{p_{\cdot 2}}{p_{j2}}\right) = 0. \tag{A.13}$$

The following identity holds

$$\begin{aligned}\langle 1|\,\rho\,|2\rangle\langle 2|\,1\rangle = \mathrm{tr}\left(|1\rangle\langle 1|\,\rho\,|2\rangle\langle 2|\right) &= \frac{1}{8}\mathrm{tr}\left((\mathbb{I}+\vec{n}_1\cdot\vec{\sigma})(\mathbb{I}+\vec{r}\cdot\vec{\sigma})(\mathbb{I}+\vec{n}_2\cdot\vec{\sigma})\right)\\ &= \frac{1}{4}\left(1+\vec{n}_1\cdot\vec{n}_2+\vec{r}\cdot(\vec{n}_1+\vec{n}_2)\right);\end{aligned}$$

applied to (A.11,A.12,A.13) gives us (1.18) and (1.19).