The application of optimal quantum measurements to quantum channels

Matthew Day[1,*]

[1]Quantum Engineering Centre for Doctoral Training, University of Bristol
[*]matt.day@bristol.ac.uk

March 27, 2015

We explore the open problem of the most efficient way to communicate classical data across quantum channels. We find that traditional optimal measurement techniques do not necessarily maximise information transfer rates and therefore the maximisation of the mutual information must be done explicitly. This inherently non-linear problem has been solved for pure, symmetric signal states but has yet to be extended to arbitrary ensembles of mixed signal states, limiting their application to physical quantum channels.

---

1 Introduction

The ambiguity of the quantum state is simultaneously the most useful and unhelpful feature of quantum information theory. On the one hand an observer's naivety of the quantum state allows us to perform calculations in large state spaces. On the other, extracting information out of a quantum system is inherently difficult as the state is characterised by statistical quantities - most importantly, its probability distribution. As the act of extracting information from a quantum state (*measuring*) disturbs the state itself it is important to choose the best measurement possible. The choice of optimal measurements is particularly important when considering communication rates over quantum channels as it is desirable to maximise the amount of information that can be received. The problem is then to determine the choice of measurement that extracts the most amount of information out of a quantum state from a sender (Alice) by a receiver (Bob). There are other considerations, such as optimising the quantum state that is being sent - but these will not be focused on here. The other important task of sending *quantum* information over quantum channels (such as quantum teleportation) will also not be considered and we are purely interested in sending classical information using quantum states.

We begin with a review of the field of optimal measurements, before moving onto its application to the mutual information of quantum channels.

2 Optimal measurements

In this section we briefly introduce the formalism of optimal quantum measurements. It is well known that in quantum mechanics orthogonal states can be distinguished by projective measurements with certainty. However in the instances where non-orthogonal states are used (such as in quantum key distribution) then projective measurements no longer reliably distinguish between quantum states and a new strategy has to be found. There are several strategies that have been developed, and when distinguishing more than 2 states it becomes difficult to prove optimality - however there are tools that can be used to find at least *approximately* optimal measurements. Another thing of note that won't be dealt with here is the *implementation* of the optimal measurements that have been derived theoretically. In practice there may not exist experimental apparatus that can efficiently implement optimal measurements, whereas a sub-optimal measurement could be efficiently implemented and yield better results. In order to determine the set of optimal measurements, we introduce the concept of general measurements. Once we have a set of general measurements we can impose constraints on the set to find the subset of optimal measurement operators.

2.1 Generalised measurements

The act of an observer gaining information about a quantum system is governed by the *general measurement postulate*[1]

> **General measurement postulate**: The collection of operators $\{\hat{M}_i\}$ that act on the Hilbert space of a quantum system are measurement operators if the action of the operator $\hat{M}_i$ on the system state $|\Psi\rangle$ gives outcome i with

probability

$$P(i) = \left\| \hat{M}_i |\Psi\rangle \right\|^2 = \langle\Psi| \hat{M}_i^\dagger \hat{M}_i |\Psi\rangle \ , \tag{1}$$

such that the system is transformed into the state

$$|\Psi'\rangle = \frac{\hat{M}_i |\Psi\rangle}{\sqrt{\langle\Psi| \hat{M}_i^\dagger \hat{M}_i |\Psi\rangle}} \ . \tag{2}$$

The operators $\{\hat{M}_i\}$ must also be *complete*, that is they must span the Hilbert space such that

$$\sum_i \hat{M}_i^\dagger \hat{M}_i = \mathbb{I} \ , \tag{3}$$

where $\mathbb{I}$ is the identity.

Traditionally a special class of general measurements are used: *projective* or *von Neumann* measurements. In this formalism each observable of the system is expressed as a Hermitian operator $\hat{O}$ that can be spectrally decomposed in terms of its eigenstates[2]

$$\hat{O} = \sum_i \lambda_i |\phi_i\rangle \langle\phi_i| = \sum_i \lambda_i \hat{P}_i \ , \tag{4}$$

where $\hat{P}_i$ is a *projection operator*. The von Neumann formalism has the advantage that if the quantum system being probed is in one of the eigenstates of $\hat{O}$ then a measurement using this operator will distinguish the state with certainty. As the states $\{|\phi_i\rangle\}$ form an eigenbasis they are necessarily orthogonal, and as such if the quantum system can be expressed as a linear superposition of this eigenbasis then the operator $\hat{O}$ can be used to completely determine the statistical properties of the quantum state using many projective measurements of identically prepared quantum systems. Projection operators necessarily satisfy the completeness requirement of the general measurement postulate as they are constructed from an eigenbasis and as such, span the Hilbert space. As well as satisfying the general measurement postulate, projectors are also *positive* and *orthonormal*. Orthonormality is easily demonstrated

$$\hat{P}_i \hat{P}_j = |\phi_i\rangle \langle\phi_i|\phi_j\rangle \langle\phi_j| = \delta_{ij} \hat{P}_i \ , \tag{5}$$

due to the orthonormality of the eigenbasis $\{|\phi_i\rangle\}$. Positivity automatically follows from the requirement that

$$P(i) = \langle\Psi| \hat{P}_i^\dagger \hat{P}_i |\Psi\rangle = \langle\Psi| \hat{P}_i \hat{P}_i |\Psi\rangle = \langle\Psi| \hat{P}_i |\Psi\rangle \geq 0 \ , \tag{6}$$

as a probability is necessarily positive.

Recall that we are interested in optimising the distinguishability of quantum states that are not necessarily orthonormal - and therefore the optimal set of quantum measurements may not be orthonormal. We therefore drop the orthonormality requirement and propose a set of operators $\{\hat{\Pi}_i\}$ related to a set of measurement operators $\{\hat{M}_i\}$ through

$$\hat{\Pi}_i = \hat{M}_i^\dagger \hat{M}_i \ . \tag{7}$$

The set $\{\hat{\Pi}_i\}$ is called a Positive Operator-Valued Measure (POVM) and represents a set of measurements through the relation above. The POVM naturally satisfies the general measurement postulate through its construction in that the operators are Hermitian, positive and complete. The framework of POVMs allow us to find general measurements of quantum systems such that we can derive the set of measurement operators for an optimal POVM. It is entirely possible that a POVM is in fact a set of projectors, however this will not be true in general. POVMs also allow us to accommodate for the case when the observer makes a destructive measurement on the state. For example detecting a photon using an avalanche photodiode detector necessarily destroys the photon and the quantum state is lost (it is not projected into the $|1\rangle$ photon number state for example). In this sense a POVM is most useful when the observer is interested in the statistical properties of the quantum state that is sent, rather than the state after measurement. When sending classical data over quantum channels, the classical data is encoded in the statistics of the quantum state and not the states themselves and as such the POVM lends itself naturally to our aims. Furthermore, a POVM represents a set of measurements that theoretically can be implemented in the laboratory.[2,3]

Having formulated a set of general measurements we require a way to restrict the POVM such that we can distinguish quantum states optimally. Several strategies will be explored in the following sections.

2.2 Minimising error

We begin with the strategy where finding the optimal POVM is most straightforward and aim to minimise the probability of misclassifying the state on measurement. In all generality the probability of making an error is the sum of the probability a certain state is sent weighted by the probability that you get the wrong measurement outcome for that state. More formally this is expressed[2]

$$P_{\text{error}} = \sum_i P(i) \sum_{i \neq j} P(j|i) \ , \tag{8}$$

where the probability of correct identification of outcome i is given by $P(i|i)$ such that the total probability of correct classification is given by

$$P_{\text{correct}} = 1 - P_{\text{error}} = \sum_i P(i) P(i|i) \ . \tag{9}$$

Now consider the ensemble of N states with associated known sending probabilities $\mathscr{E} = \{\hat{\rho}_i, p_i\}$. The ensemble is associated with a quantum channel from Alice (A) to Bob (B). The states $\{\hat{\rho}_i\}$ are in general mixed states, but the formalism also applies to pure states. The probability of error is now given by[2]

$$P_{\text{error}} = \sum_{i=0}^{N-1} p_i \sum_{j \neq i} \text{Tr}(\hat{\rho}_i \hat{\Pi}_j) \ , \tag{10}$$

where $\{\hat{\Pi}_i\}$ is a POVM. We are now interested in the best POVM under this strategy. It is obvious that the POVM that corresponds to the optimal measurement in this strategy is the one that minimises the probability of error. Necessary and sufficient conditions on the POVM to be optimal exist and their derivations can be found in the literature,[4] we only quote them here. The optimal POVM $\{\hat{\Pi}_i\}$ for minimising error in distinguishing between states $\{\hat{\rho}_i\}$ with probability of being sent $\{p_i\}$ must satisfy[2]

$$\sum_i p_i \hat{\rho}_i \hat{\Pi}_i - p_j \hat{\rho}_j \geq 0 \quad \forall \ j \ , \tag{11}$$

or

$$\hat{\Pi}_i (p_i \hat{\rho}_i - p_j \hat{\rho}_j) \hat{\Pi}_j = 0 \quad \forall \ i, j \ . \tag{12}$$

These conditions are not independent and a POVM that satisfies one automatically satisfies the other.

2.2.1 Square-root measurement

The square-root measurement was developed by Hausladen and further developed with Wootters.[5] It is a well-known example of a POVM construction to minimise error and in many examples the square-root measurement satisfies the necessary and sufficient conditions to minimise error. Let us begin by considering the signal states to be pure, that is consider the signal set $\{|s_i\rangle\}$ such that $\hat{\rho} = \sum_i |s_i\rangle \langle s_i|$ (i.e. the probabilities are included in the signal state normalisation). Consider POVMs of the form

$$\hat{\Pi}_i = |m_i\rangle \langle m_i| \ , \tag{13}$$

where $\{|m_i\rangle\}$ are our measurement states. We construct the square-root measurement set using the relation

$$|\mu_i\rangle = \hat{\rho}^{-\frac{1}{2}} |s_i\rangle \ , \tag{14}$$

where $\{|\mu_i\rangle\}$ are our square-root measurement states. Our POVM elements are now given by

$$\hat{\Pi}_i = \hat{\rho}^{-\frac{1}{2}} |s_i\rangle \langle s_i| \hat{\rho}^{-\frac{1}{2}} \ . \tag{15}$$

Extending to an ensemble of mixed signal states $\mathscr{E} = \{\hat{\rho}_i, p_i\}$, equation 15 generalises to[2]

$$\hat{\Pi}_i = p_i \hat{\rho}^{-\frac{1}{2}} \hat{\rho}_i \hat{\rho}^{-\frac{1}{2}} \ . \tag{16}$$

We will now apply the square-root measurement to the quantum key distribution protocol BB84 as an example of how this strategy might be used.

Application: quantum key distribution

In the original paper by Bennett and Brassard[6] where the protocol now know as BB84 was developed it was assumed the chosen measurements were optimal. In the interest of creating a one-time pad (or *quantum key*) Alice sends one of four states, Bob then measures using the equivalent projection operators. Bob will only get a meaningful measurement half the time and when he has finished measuring all sent states he reports the basis of the successful measurements to Alice who then informs him which are correct. This forms the quantum key which both Alice and Bob share, and the key can then be used for encryption. We would like to motivate optimality of this procedure by using the square root measurement strategy.

Bennet and Brassard motivated BB84 using photon polarisations in either the linear or diagonal basis, that is the four possible states that could be sent were $\{|H\rangle, |V\rangle, |D\rangle, |A\rangle\}$ which we will rewrite as the more familiar (but entirely equivalent) set $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$ where $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. Note how the four states are neither linearly independent or form an orthonormal basis - this is where the security of BB84 comes from. Each of the four states are sent with equal probabilities and the density operator for this system is

$$\hat{\rho} = \frac{1}{4}(|0\rangle\langle 0| + |1\rangle\langle 1| + |-\rangle\langle -| + |+\rangle\langle +|) = \frac{\hat{\mathbb{I}}}{2}\,, \tag{17}$$

that is, the maximally mixed state. The only quantity to calculate to construct the square-root measurement is

$$\hat{\rho}^{-\frac{1}{2}} = \sqrt{2}\,\hat{\mathbb{I}}\,, \tag{18}$$

and from equation 16 we find the measurement set

$$\{\hat{\pi}_i\} = \left\{\frac{1}{2}|0\rangle\langle 0|, \frac{1}{2}|1\rangle\langle 1|, \frac{1}{2}|-\rangle\langle -|, \frac{1}{2}|+\rangle\langle +|\right\}\,, \tag{19}$$

where we use a lower-case $\hat{\pi}_i$ to indicate a specific POVM. Note that $\sum_i \hat{\pi}_i = \hat{\mathbb{I}}$ as required. The derived measurement set is indeed the measurement set used in BB84 and can be written more compactly as $\hat{\pi}_i = 1/2\hat{\rho}_i$. Using the second condition on error minimisation, equation 12, we can show

$$\begin{aligned}
\hat{\pi}_i(p_i\hat{\rho}_i - p_j\hat{\rho}_j)\hat{\pi}_j &= \frac{1}{8}\hat{\rho}_i(\hat{\rho}_i - \hat{\rho}_j)\hat{\rho}_j\,, &(20)\\
&= \frac{1}{8}(\hat{\rho}_i\hat{\rho}_i\hat{\rho}_j - \hat{\rho}_i\hat{\rho}_j\hat{\rho}_j)\,, &(21)\\
&= \frac{1}{8}(\hat{\rho}_i\hat{\rho}_j - \hat{\rho}_i\hat{\rho}_j)\,, &(22)\\
&= 0\,, &(23)
\end{aligned}$$

where we have used the fact that the density matrices are pure and therefore $\hat{\rho}_i^2 = \hat{\rho}_i$. We have therefore proved that the measurement set in BB84 minimises the probability of error in Bob determining the correct state.

2.3 Maximum confidence

An alternative strategy to minimising the total error of classification is to instead maximise the probability that a measurement outcome correctly determines the sent state. While these might appear to be both sides of the same coin, they lead to quite different strategies. More precisely we are interested in maximising[2]

$$P(\hat{\rho}_i|i) = \frac{P(\hat{\rho}_i)P(i|\hat{\rho}_i)}{P(i)}\,, \tag{24}$$

We will begin by discussing linearly independent signal states as introduced by Ivanovic,[7] proved optimal by Peres[8] and further developed by Dieks,[9] before moving onto the more general case.

2.4 Unambiguous Discrimination

First consider distinguishing between N pure signal states $\{|s_i\rangle\}$. We wish to find the POVM $\{\hat{\Pi}_i\}$ such that the probability in equation 24 is unity, or equivalently[2]

$$P(i)\delta_{ij} = \langle s_i|\hat{\Pi}_i|s_i\rangle\,, \tag{25}$$

where $P(i)$ is the probability of the measurement outcome i occurring, which happens if and only if the state is $|s_i\rangle$. This can only be satisfied when the signal states are linearly independent such that we can construct the states $\{|s_i^\perp\rangle\}$ where state $|s_i^\perp\rangle$ is orthogonal to all signal states apart from $|s_i\rangle$. This can be formalised as

$$\langle s_i|s_j^\perp\rangle = \langle s_j|s_j^\perp\rangle\,\delta_{ij}\,. \tag{26}$$

It is fairly easy to motivate that this condition is only satisfied by linearly independent states: consider the linearly dependent set $\{|0\rangle,|1\rangle,|+\rangle\}$; as the state $|+\rangle$ is a linear superposition of the other two states any state that is orthogonal to both $|0\rangle$ and $|1\rangle$ will also be orthogonal to $|+\rangle$. Assuming therefore that the states are linearly independent we can construct the POVM operators

$$\hat{\Pi}_i = \frac{P(j)\,|s_j^\perp\rangle\,\langle s_j^\perp|}{|\,\langle s_j|s_j^\perp\rangle\,|^2}\,, \tag{27}$$

that satisfy the condition for maximising confidence as in equation 25. Therefore the POVM specified by equation 27 is said to *unambiguously discriminate* between the signal states $\{|s_i\rangle\}$ in that if measurement outcome i is recorded then the state $|s_i\rangle$ was definitely sent. There is a catch however in that the POVM is only complete if the signal states are orthogonal and therefore usually the POVM has to be completed with an additional operator. The most general way to do this is to define it as the operator over the remainder of the space

$$\hat{\Pi}_? = \hat{\mathbb{I}} - \sum_i \hat{\Pi}_i\,, \tag{28}$$

where the ? indicates that the measurement outcome is meaningless and provides no information. The strategy can then unambiguously discriminate between signal states with the trade-off of the occasional inconclusive result. It is important to note that while this strategy can distinguish states with certainty, often the probability of an inconclusive result can be quite high and such would often make it useless for a communication channel. However the application to quantum key distribution discussed below demonstrates its usefulness in some circumstances. The linear independence condition on the signal states is perhaps too prohibitive for any useful quantum system, however unambiguous discrimination has thankfully been generalised to linearly dependent states.

Application: quantum key distribution

In 1992, Bennett proposed a new quantum cryptography protocol that relied on non-orthogonal, linearly independent states.[10] As with BB84, Alice sends a set of signal states, this time $\{|0\rangle,|+\rangle\}$ which Bob has to be discriminate between using unambiguous discrimination. Using equation 26 and 27 we can construct the POVM where the orthogonal signal set $\{|s_i^\perp\rangle\}$ is $\{|1\rangle,|-\rangle\}$. The probability of an inconclusive measurement is found to be $1/\sqrt{2}$,[11] and therefore the probability of successful unambiguous discrimination is $1-1/\sqrt{2}\approx 29\%$. The low probability of success does not matter too much, Bob simply tells Alice which bits he failed to discriminate and Alice throws away that part of her key. Now consider an eavesdropper, Eve, who has access to the quantum states before Bob. If she uses unambiguous discrimination she will fail 71% of the time. If Eve then tries to fool Bob by sending a replacement state to Bob then Eve will be unsure what state to send if she obtains an inconclusive result. If Eve simply guesses which state to send she will therefore send the wrong state 35.5% of the time (half the time her guess will be correct). For a secure channel, if Alice and Bob compared some of their bits then they should have a zero error rate (as they are discarding all the inconclusive results). If instead there is an error rate of approximately 35% then this is a clear signal of a compromised channel and they discard the key entirely. There are other details, and it is possible for Eve to reduce her error rate, however it still stands that if there are any appreciable errors then the channel is insecure.[11]

2.5 Generalised maximum confidence measurements

The generalised form of unambiguous discrimination was developed by Croke et al[12] and still attempts to maximise the probability of equation 24 but with linearly dependent states. Now that the states are not necessarily linear independent no POVM will be able to unambiguously discriminate the signal states, however we can still construct POVMs that will maximise our probability of being correct. Rewriting equation 24 in terms of our ensemble of signal states with their associated probabilities $\mathscr{E} = \{\hat{\rho}_i, p_i\}$ and our POVM $\{\hat{\Pi}_i\}$ (each associated with outcome i) we have[12]

$$P(\hat{\rho}_i|i) = \frac{p_i \mathrm{Tr}(\hat{\rho}_i\hat{\Pi}_i)}{\mathrm{Tr}(\hat{\rho}\hat{\Pi}_i)}\,. \tag{29}$$

As with unambiguous discrimination with wish to maximise the probability of equation 29. We begin with the ansatz[12]

$$\hat{\Pi}_i = w_i\hat{\rho}^{-1/2}\hat{Q}_i\hat{\rho}^{-1/2}\,, \tag{30}$$

where the positive, trace 1 operator $\hat{Q}_i$ is to be determined and $w_i = P(i)$. It is interesting to compare equation 30 with the square-root measurement POVM in equation 16 and see that the maximum confidence case appears to be a generalised square-root measurement. The square-root measurement POVM was chosen such that the set of measurement operators would be complete, now that it has been generalised the resulting POVM will not necessarily be complete just as with unambiguous discrimination, and an inconclusive result will have to be included in general. By noting that $w_i = \mathrm{Tr}(\hat{\rho}\hat{\Pi}_i)$ and substituting in our ansatz into equation 29 we find

$$P(\hat{\rho}_i|i) = \frac{p_i \mathrm{Tr}(\hat{\rho}_i w_i \hat{\rho}^{-1/2} \hat{Q}_i \hat{\rho}^{-1/2})}{w_i} , \tag{31}$$

$$= p_i \mathrm{Tr}(\hat{\rho}_i \hat{\rho}^{-1/2} \hat{Q}_i \hat{\rho}^{-1/2}) , \tag{32}$$

$$= p_i \mathrm{Tr}(\hat{\rho}_i \hat{\rho}^{-1}) \mathrm{Tr}\left(\frac{\hat{\rho}^{-1/2} \hat{\rho}_i \hat{\rho}^{-1/2}}{\mathrm{Tr}(\hat{\rho}_i \hat{\rho}^{-1})} \hat{Q}_i \right) , \tag{33}$$

$$= p_i \mathrm{Tr}(\hat{\rho}_i \hat{\rho}^{-1}) \mathrm{Tr}(\hat{\sigma}_i \hat{Q}_i) , \tag{34}$$

where we have used the cyclic properties of the trace and defined

$$\hat{\sigma}_i = \frac{\hat{\rho}^{-1/2} \hat{\rho}_i \hat{\rho}^{-1/2}}{\mathrm{Tr}(\hat{\rho}_i \hat{\rho}^{-1})} , \tag{35}$$

which is also a positive, trace 1 operator. The probability $P(\hat{\rho}_i|i)$ is therefore maximised when $\hat{Q}_i$ is the projector onto the eigenvector of $\hat{\sigma}_i$ with the largest eigenvalue, that is

$$\hat{Q}_i = |\lambda_i^{\max}\rangle \langle \lambda_i^{\max}| , \quad \hat{\sigma}_i |\lambda_i^{\max}\rangle = \lambda_i^{\max} |\lambda_i^{\max}\rangle , \tag{36}$$

such that

$$\max[P(\hat{\rho}_i|i)] = p_i \mathrm{Tr}(\hat{\rho}_i \hat{\rho}^{-1}) \lambda_i^{\max} , \tag{37}$$

and our POVM becomes

$$\hat{\Pi}_i = w_i \hat{\rho}^{-1/2} |\lambda_i^{\max}\rangle \langle \lambda_i^{\max}| \hat{\rho}^{-1/2} . \tag{38}$$

Different choices of $\{w_i\}$ give distinct maximum confidence POVMs and are often chosen with the desire to minimise the probability of an inconclusive measurement (ideally, zero probability). As a comparison to the square-root measurement, consider the case of three symmetric qubit states[12]

$$|\psi_0\rangle = \cos\theta |0\rangle + \sin\theta |1\rangle , \tag{39}$$

$$|\psi_1\rangle = \cos\theta |0\rangle + e^{2\pi i/3} \sin\theta |1\rangle , \tag{40}$$

$$|\psi_2\rangle = \cos\theta |0\rangle + e^{-2\pi i/3} \sin\theta |1\rangle . \tag{41}$$

It can be found using the formalism above that the maximum confidence probability for these states is

$$P_{mc}(\hat{\rho}_i|i) = 2/3 . \tag{42}$$

Calculating the same value for the square-root measurement we find

$$P_{sr}(\hat{\rho}_i|i) = \frac{1}{3}(1 + \sin 2\theta) , \tag{43}$$

such that $P_{sr}(\hat{\rho}_i|i) \leq P_{mc}(\hat{\rho}_i|i)$ and the maximum confidence is better for all θ apart from $\pi/4$ where the two strategies are equal. This was experimentally demonstrated by Croke et al.[13]

3 Quantum channels and mutual information

We now turn our attention to the application of optimal measurements to quantum channels. As with classical channels we are interested in maximising the amount of information that can be decoded by our receiver, Bob. We start by making use of

Shannon's information theory which states that the observer's knowledge of a system that has probability $P(i)$ of being in state i is given by the Shannon entropy

$$H = -\sum_i P(i)\log_2 P(i)\ , \tag{44}$$

and any measurement of the system will alter the probabilities to *conditional* probabilities, altering the entropy and the observer's knowledge of the system. The resulting information gain of the observer is given by the difference between the initial and final entropy of the system.[3] In the context of channels, this information gain is known as the *mutual* information, defined as[14]

$$I(X:Y) = H(X) - H(X|Y)\ , \tag{45}$$

where X is all the possible signal values $\{x_i\}$ sent by Alice, and Y is the set of values $\{y_i\}$ determined by Bob such that

$$H(X|Y) = -\sum_i P(y_i)\sum_i P(x_i|y_i)\log_2 P(x_i|y_i)\ . \tag{46}$$

Generalising the above formalism to an ensemble of signal states $\mathscr{E} = \{\hat{\rho}_a, p_a\}$, each denoting a message a sent by Alice to Bob, then the mutual information between Alice and Bob is given by the Holevo bound[14]

$$I(A:B) \leq H(\hat{\rho}) - \sum_a p_a H(\hat{\rho}_a)\ . \tag{47}$$

where $H(\hat{\rho}) = -\mathrm{Tr}(\hat{\rho}\ln\hat{\rho})$ is the von Neumann entropy. In general, the equality does not hold and there are cases where no choice of Bob's POVM will allow the mutual information to reach its maximum value. However, it is a useful quantity to maximise nonetheless and makes comparison of different POVMs easier in the context of quantum channels.

There are two distinct ways that mutual information is used. The two optimal measurement strategies that have been studied so far have both essentially been minimising the error of misclassifying the signal state (or, *minimising the'Bayes cost'*), which is a well-behaved and easily solvable case - as has been demonstrated. Entropy, on the other hand, is a nonlinear quantity and in turn the maximum mutual information is difficult to solve for. Therefore the two distinct cases are

1. A POVM is found by minimising the Bayes cost of the signal state ensemble. The mutual information is then found for the POVM and compared to other strategies.

2. The fact that the mutual information is a convex function is utilised to find the POVM that maximises the mutual information.

As already stated, the latter is much harder than the former, though is more likely to find the optimal quantum measurements that closest approach the Holevo bound. An example of the first case is presented first by Hausladen et al[14] and then extended to mixed states by Schumacher and Westmoreland.[15] In these two papers the square-root measurement is used to find a POVM for which the mutual information is then calculated.

Davies was the first to tackle the second case using a group-theoretic approach[16] and later extended for symmetric ensembles (such as equations 39-41) by Sasaki et al.[17] More recently, numerical methods have been developed to tackle this difficult problem.[18] However a general solution to find the optimal POVM that maximises the mutual information for a given quantum ensemble and channel is still an open problem. Despite the difficulty of finding a generalised formalism for maximising the mutual information, it should be noted that the strategy developed by Sasaki et al[17] is perhaps sufficient for an optimal (noiseless) quantum channel. Sasaki et al consider the set of M qubit states symmetrically distributed around a great circle on the Bloch sphere, which we can represent as the ensemble

$$\mathscr{E}_M = \left\{ |\psi_k\rangle = \cos\left(\frac{k\pi}{M}\right)|0\rangle + \sin\left(\frac{k\pi}{M}\right)|1\rangle, p_k = \frac{1}{M} \;\middle|\; k = 0,...,M-1 \right\}\ . \tag{48}$$

Sasaki et al then go on to show that in order to maximise the mutual information only three POVM elements are required for such an ensemble, no matter how large M is. This allows the construction of an alphabet of states to efficiently send classical information via a quantum channel using only three measurement settings. For two non-orthogonal qubit states ($M = 2$) it was predicted[19] and then shown[20] that the minimum error POVM coincides with the maximisation of mutual information. However when using a binary (also $M = 2$) ensemble of mixed states is has been found that the optimal POVM that maximises the mutual information differs from the POVM that minimises the Bayes cost.[21] For $M > 2$, such as the triad in equations 39-41 it has been demonstrated that the minimum error POVM does not maximise the mutual information[22] and therefore more careful measurements must be used in these cases. It is of importance to note that all of these experimental demonstrations occurred in optics and therefore translate themselves well to the consideration of quantum communications.

4 Summary

We have introduced the formalism of optimal quantum measurements. The main strategy is to use a generalised measurement that minimises the Bayes cost. If considering the maximum amount of information that can be extracted from a given ensemble sent across a quantum channel then the optimal quantum measurements changes for sets of 3 or more pure signal states - however currently these states are restricted to being symmetrically distributed on the Bloch sphere. The formalism of Sasaki et al has yet to be extended to mixed states and therefore cannot model the effect of noise in a quantum channel for more than two signal states, nor have the ensemble probabilities been optimised for an optimal POVM. It is these two problems that hold back full implementation of optimal quantum measurements in order to maximise the sending of classical information over quantum channels.

References

1. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, New York, NY, USA, 2011), 10th edn.
2. Barnett, S. M. & Croke, S. Quantum state discrimination. *Advances in Optics and Photonics* **1**, 238 (2009).
3. Peres, A. & Wootters, W. Optimal detection of quantum information. *Physical Review Letters* **66**, 1119–1122 (1991).
4. Barnett, S. M. & Croke, S. On the conditions for discrimination between quantum states with minimum error. *Journal of Physics A: Mathematical and Theoretical* **42**, 062001 (2009).
5. Hausladen, P. & Wootters, W. K. A 'Pretty Good' Measurement for Distinguishing Quantum States. *Journal of Modern Optics* **41**, 2385–2390 (1994).
6. Bennett, C. H. & Brassard, G. Quantum cryptography: Public key distribution and coin tossing. *Theoretical Computer Science* (1984).
7. Ivanovic, I. How to differentiate between non-orthogonal states. *Physics Letters A* **123**, 257–259 (1987).
8. Peres, A. How to differentiate between non-orthogonal states. *Physics Letters A* **128**, 19 (1988).
9. Dieks, D. Overlap and distinguishability of quantum states. *Physics Letters A* **126**, 303–306 (1988).
10. Bennett, C. Quantum cryptography using any two nonorthogonal states. *Physical Review Letters* **68**, 3121–3124 (1992).
11. Bergou, J. a. Quantum state discrimination and selected applications. *Journal of Physics: Conference Series* **84**, 012001 (2007).
12. Croke, S., Andersson, E., Barnett, S. M., Gilson, C. R. & Jeffers, J. Maximum confidence quantum measurements. *Physical Review Letters* **96**, 1–4 (2006).
13. Croke, S., Mosley, P. J., Barnett, S. M. & Walmsley, I. A. Maximum confidence measurements and their optical implementation. *The European Physical Journal D* **41**, 589–598 (2006).
14. Hausladen, P., Jozsa, R., Schumacher, B., Westmoreland, M. & Wootters, W. Classical information capacity of a quantum channel. *Physical Review A* **54**, 1869–1876 (1996).
15. Schumacher, B. & Westmoreland, M. Sending classical information via noisy quantum channels. *Physical Review A* **56**, 131–138 (1997).
16. Davies, E. B. Information and Quantum Measurement **l**, 596–599 (1978).
17. Sasaki, M., Barnett, S. M., Jozsa, R., Osaki, M. & Hirota, O. Accessible information and optimal strategies for real symmetrical quantum sources **59**, 16 (1998).
18. Dall'Arno, M., D'Ariano, G. M. & Sacchi, M. F. Informational power of quantum measurements. *Physical Review A* **83**, 062304 (2011).
19. Mizuno, J. *et al.* Optimum detection for extracting maximum information from symmetric qubit sets **65**, 10 (2001).
20. Barnett, S. M. Optical demonstrations of statistical decision theory for quantum systems. *Quantum Information & Computation* **4**, 450–459 (2004).
21. Tomassoni, N. & Paris, M. G. Quantum binary channels with mixed states. *Physics Letters A* **373**, 61–64 (2008).
22. Clarke, R. B. M. *et al.* Experimental realization of optimal detection strategies for overcomplete states **64**, 14 (2000).