Understanding "Universal discriminative quantum neural networks"

Abhilash Neog

May 6, 2023

Abstract

The article presents a brief discussion, in the form of a technical report, on the paper titled "Universal discriminative quantum neural networks" [1]. Similar to the theme of this paper, the article focuses primarily on Quantum State Discrimination. The article introduces and discusses the field of Quantum State Discrimination (QSD), the types, challenges and strategies, before transitioning into the approach adopted by the paper and discussing its modified approach to tackling the QSD problem. The article further talks about the paper's unique elements, such as Machine learning based training strategy and the interesting experiments and results. This is followed by a brief discussion on the practical applications of the proposed technique, with a focus on Quantum repeaters as an example.

Contents

1 Introduction

Machine Learning and Quantum Computing has been receiving a lot of interest in recent times. Machine Learning algorithms usually require a lot of computation capabilities, and this is where quantum comes into play - it can significantly speed-up certain processes compared to it's classical counterparts. It can also work the other way round [[2],[3],[4]]. The latter is explored in this paper by developing a Quantum Neural Network to perform quantum data classification. Now, the ability to classify different classes of quantum data is extremely important in quantum information science. Quantum state discrimination plays a crucial role in quantum information processing protocols and find its use in many areas like quantum cryptography [5], quantum meteorology [6], quantum imaging [[7]]. etc.

However, this is not always straightforward - quantum mechanics strictly prohibits the deterministic determination of quantum states and processes. The authors in the paper tries to tackle this problem by training near-term quantum circuits to distinguish data represented by non-orthogonal quantum probability distributions. Basically, they train a quantum circuit to mimic a Positive-Operator-Value-Measure in order to classify the different distributions of quantum input. The shallow quantum circuits they develop here have a strong resemblance to the structure of quantum neural networks with non-unitary layers. Similar to neural networks, the multiple layers of generalized measurements in the circuit introduces adequate amount of non-linearities needed for performing classification.

The article is organized as follows. In Section 2 (Background), some of the relevant quantum terms are defined that is expected to aid in understanding the concepts in paper and article. Section 3 (Methodology) is divided into two parts - 3.1 Quantum State Discrimination and 3.2 Proposed Technique. Section 3.1 introduces and discusses the problem of Quantum State Discrimination in (high-level) detail. This also forms the basis for Section 3.2 in describing the proposed approach and training techniques. Section 4 (Experiments and Results) highlights the experiments conducted with the proposed classification approach and the results generated in the process. Finally, in Section 5 (Quantum Repeaters), the article explores the practical and relevant usage of this paper by attempting to integrate it with Quantum repeaters.

2 Background

This section talks about some of the quantum terms or concepts that are touched upon by the research paper, and would be useful to know to understand certain parts of the paper/article better.

Definition 2.1. A **density operator** ρ is another way to represent a quantum state. It is positive semidefinite and has unit trace. i.e., $\rho \succeq 0$ and $\mathrm{Tr}(\rho) = 1$. We can have density operators for pure as well as mixed states. We can represent any quantum state as a density operator, i.e. quantum state $|0\rangle$ can also be represented as $|0\rangle \langle 0|$. Furthermore, any generic mixture can be represented as,

$$(p_1, |\psi_1\rangle), (p_2, |\psi_2\rangle), ..., (p_n, |\psi_n\rangle)$$

where, $(p_1, p_2, ..., p_n)$ is a probability distribution

The density matrix for the same can be written as,

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

Definition 2.2. A **Positive Operator-Valued Measure (POVM)** is a mathematical framework used to describe the measurement of a quantum system. In simple terms, it is a collection of positive semidefinite matrices $P_1, P_2, ..., P_n$, such that,

$$\sum_{i=1}^{n} P_i = I$$

where, $P_i = M_i^* M_i$, and $\{M_1, M_2, ..., M_n\}$ is a measurement

Just like a density matrix corresponds to a mixture, any POVM corresponds to a measurement. It is also interesting to note that POVM does not tell anything about the post-measured states, it is only concerned with the outcome probabilities.

Definition 2.3. Fidelity $F(\rho_1, \rho_2)$ is a measure of how similar two quantum states are to each other. Technically, fidelity measures the overlap between two quantum states, and is a way to quantify how well a quantum state preparation or quantum operation approximates a desired target state or operation. Mathematically, it can be represented as,

$$F(\rho_1, \rho_2) = Tr(\sqrt{\sqrt{\rho_1}\rho_2\sqrt{\rho_2}})$$

such that, $F(\rho_1, \rho_2) = 0$, means, states are dissimilar, and $F(\rho_1, \rho_2) = 1$, means, states are very similar.

Definition 2.4. A **Quantum Channel** $\phi(\rho)$ refers to a mathematical model that describes the way quantum systems can be transformed from one state to another. Technically, these channels can be thought of as linear operators that maps the input state of a quantum system to an output state, where, both input and output states can be considered as density matrices. Mathematically, any quantum channel can be represented as a collection of matrices $\{A_1, ..., A_k\}$ where,

$$\phi(\rho) = \sum_{j=1}^{k} A_j \rho A_j^*$$

where $\sum_{j=1}^{k} A_j A_j^* = I$, such that,

$$\phi(\rho) \succeq 0$$
$$Tr(\phi(\rho)) = 1$$

Definition 2.5. Quantum channels can be broadly classified into **unitary channels** and **non-unitary channels**. The primary difference between a unitary and non-unitary quantum channel is that, unitary channels preserve quantum information, whereas non-unitary channels do not. A unitary quantum channel is a reversible process that can be thought of as a rotation in the space of possible quantum states.

In contrast, non-unitary quantum channels are irreversible that do not preserve quantum information, typically due to interactions with an environment that introduces noise or decoherence into the system.

Definition 2.6. Holevo's bound is an important result in quantum computing that places an upper bound on the amount of classical information that can be transmitted through a quantum channel. The bound states that the maximum amount of classical information that can be transmitted through a quantum channel is equal to the mutual information between the input and the output state of the channel. In other words, it says that given n qubits, only n classical bits of "information" can be "accessed".

3 Methodology

3.1 Quantum State Discrimination

Quantum State Discrimination (QSD) is the task of identifying an unknown quantum state ρ from a set of possible quantum states. The challenge here is that this set of possible states are believed to be non-orthogonal, which makes it difficult to distinguish them with certainty. Therefore, the goal of quantum state discrimination is to find the best measurement strategy that optimizes the probability of correctly identifying the unknown state from the set of possible states. This task is important in quantum information processing and quantum communication protocols.

QSD can be discussed based on the type of quantum states [8] - pure or mixed. Pure state discrimination is relatively simpler. Technically, it can be defined as, given a n-dimension quantum state $|\psi\rangle$, which can belong to either $|\psi_1\rangle$ or $|\psi_2\rangle$, identify which state $|\psi\rangle$ actually is. The errors associated with it are - p_1 probability that the guess is wrong given the truth is $|\psi_1\rangle$, and p_2 probability that the guess is wrong given the truth is $|\psi_2\rangle$. The aim here is to devise a technique such that the max(p_1, p_2) is minimized.

A common strategy for discriminating pure states is to perform projective measurement. Say, we have an orthogonal basis $\{|v_1\rangle, |v_2\rangle\}$ that are in the span of $|\psi_1\rangle$, $|\psi_2\rangle$ (the states we need to classify) such that $\langle v_1, v_2\rangle = 0$ and $|v_i\rangle$ is near to $|\psi_i\rangle$ for i=1,2. The strategy is, if we get the outcome $|v_i\rangle$ we guess the quantum state $|\psi\rangle$ to be $|\psi_i\rangle$. As shown in Fig. 1, the assumption behind this strategy is that the angle θ between $|\psi_1\rangle$ and $|\psi_2\rangle$ lies between 0 and $\pi/2$.

Using this strategy, the probability of achieving success is,

$$\begin{aligned} |\langle \psi_i, v_i\rangle|^2 &= \cos^2\left(\frac{\pi/2-\theta}{2}\right) \\ &= \frac{1}{2} + \frac{1}{2}\cos(\pi/2-\theta) \\ &= \frac{1}{2} + \frac{1}{2}\sin\theta \end{aligned} \tag{1}$$

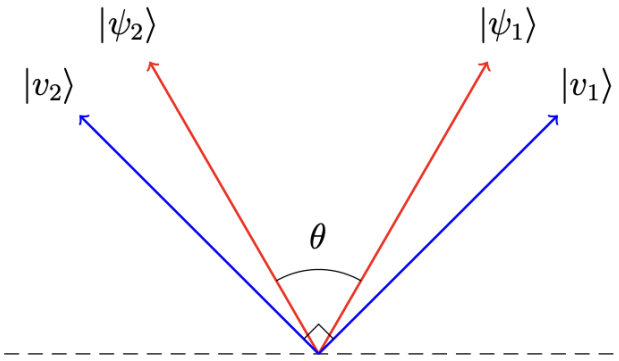

Fig. 1. Discriminating Pure states through projection

With mixed states, the goal is same, except, here we have mixed states ρ_1 and ρ_2 with probabilities P_1 and P_2, where any ρ_i can be represented as $\{p_{ij}, |\psi_{ij}\rangle\}_{j=1}^{d}$. Mixed states discrimination represents a more general case, and it's success probability is given by the Holevo-Helstrom theorem as defined below. This also explains the optimality of Eq. 1

Theorem 3.1. The optimal probability of successfully discriminating two mixed states, represented by ρ_1 and ρ_2, is generally given by the expression $\frac{1}{2} + \frac{1}{2}||\rho_1 - \rho_2||_1$.

The proof of Theorem 3.1 is simple yet quite interesting. Let's define a POVM $\{P_1, P_2\}$ s.t. $P_1 \succeq 0$, $P_2 \succeq 0$ and $P_1 + P_2 = I$. The probability of success in guessing ρ_i given outcome i is,

$$\frac{1}{2}Tr(P_1\rho_1) + \frac{1}{2}Tr(P_2\rho_2)$$

$$= \frac{1}{4}Tr((P_1 + P_2)(\rho_1 + \rho_2)) + \frac{1}{4}Tr((P_1 - P_2)(\rho_1 - \rho_2))$$

$$= \frac{1}{2}X_1 + \frac{1}{2}X_2 \text{ (let)}$$

where $X_1 = \frac{1}{2}Tr((P_1 + P_2)(\rho_1 + \rho_2))$ and $X_2 = \frac{1}{2}Tr((P_1 - P_2)(\rho_1 - \rho_2))$

As $P_1 + P_2 = I$ and $Tr(\rho_i) = 1$, we have $X_1 = 1$. To solve for X_2, we can use the following inequality: given any two hermitian matrices, M_1 and M_2,

$$Tr(M_1M_2) \leq ||M_1||_p||M_2||_q, \forall p, q \in [1, \infty] \quad s.t. \frac{1}{p} + \frac{1}{q} = 1$$

Using this, we have,

$$X_2 = \frac{1}{2}Tr((P_1 - P_2)(\rho_1 - \rho_2)) \leq \frac{1}{2}||P_1 - P_2||_\infty||\rho_1 - \rho_2||_1$$

As, $0 \preceq P_1, P_2 \preceq 1$, we have $||P_1 - P_2||_\infty \leq 1$. Hence, the probability of success,

$$= \frac{1}{2}X_1 + \frac{1}{2}X_2 \leq \frac{1}{2} + \frac{1}{2}||\rho_1 - \rho_2||_1$$

This is an interesting theorem. However, the paper does not delve much into this side of QSD, rather it focuses on the two strategies of dealing with QSD - *1. Minimum-error state discrimination* and *2. Unambiguous state discrimination*. The first strategy is a deterministic approach that aims to minimize the errors in classification. In this approach, the number of outcomes equal the number of possible states. The second strategy, however, takes a non-deterministic approach that primarily tries to identify the correct state with 100% certainty. The idea is interesting yet can be confusing. What this approach does is that, it includes an extra outcome state - inconclusive outcome. This is introduced to account for the lack of knowledge. Say, we need to identify a state ρ. If we are not fully certain, we output inconclusive outcome, instead of guessing and introducing errors. This way any sort of ambiguity is nullified. Hence, in this approach, the number of outcomes will always be greater than the number of possible input states, or one plus the number of input states.

3.2 Proposed Technique

To provide a high-level overview of what the authors have done in the paper, they train quantum circuits to distinguish data represented by non-orthogonal quantum probability distributions, with the aim of learning the generalized structure of a POVM such that, the quantum neural network (or the trained circuit) show maximum generalization performance.

As discussed in the QSD section, there are two different strategies in terms of having or not having errors. However, achieving pure unambiguous quantum state discrimination (QSD) is impractical due to the high cost of inconclusive outcomes, and that it is necessary to allow for small non-zero errors. To solve the QSD problem, the authors adopt a numerical optimization approach [[9],[10]]. However, instead of solely optimizing the figure of merit, they propose a

hybrid strategy that combines both the strategies using machine learning. The strategy looks something like this :

$$P_{succ} = \text{Probability of guessing it right}$$

$$P_{err} = \text{Probability of guessing it wrong}$$

$$P_{inc} = \text{Probability of guessing it as inconclusive}$$

if $P_{inc} == 0$ **then**
Minimize P_{err} -> it becomes the standard minimum-error state discrimination strategy
else if $P_{err} == 0$ **then**
Minimize P_{inc} -> it becomes the standard unambiguous state discrimination strategy
end if

To further explain their idea, the authors talk about the input distribution for the quantum circuit. In this paper, they consider the QSD problem of discriminating between an ensemble of families of pure states. Each state in the ensemble is drawn from a specific family of states that is characterized by a parametric distribution. Specifically, the input is given as follows:

$$\rho = \sum_{i=1}^{m} \lambda_i \left|\psi_i(\alpha_i)\right\rangle \left\langle\psi_i(\alpha_i)\right|$$

where $\sum_{i=1}^{m} \lambda_i = 1$ and α_k is any discrete classical probability distribution, and we can further write the above quantum state as,

$$\left|\psi_i(\alpha_i)\right\rangle := \left|\psi_i(a_i)\right\rangle_{a_i \sim \alpha_i}$$

The above shorthand notation means that, a is randomly drawn from the distribution α_i. Let $\beta(a)$ represent some distribution for the parameterised value a. Basically, the idea is that a_i is drawn from the distribution α_i, which is then used to specify the distribution $\left|\psi_i\right\rangle = \sum_{j=1}^{N} \beta_i(a) \left|i\right\rangle$. Now, from here, $\left|\psi_i(\alpha_i)\right\rangle$ is used as input to the circuit with certain probability.

The authors also walk us through an example to show how they set the benchmark result for comparison. Let's consider the QSD problem among 2 families of non-orthogonal 2-qubit quantum states. For $a, b \in [0, 1]$, define,

$$\psi_a(a) = (\sqrt{1-a^2}, 0, a, 0)$$

$$\psi_{2/3}(b) = (0, \pm\sqrt{1-b^2}, b, 0)$$

For simplicity, only the first family of states is kept learnable, and the second set is fixed by setting $b = \frac{1}{\sqrt{2}}$. Therefore, the states now become,

$$\psi_a(a) = (\sqrt{1-a^2}, 0, a, 0)$$

$$\psi_{2/3}(b) = (0, \pm\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$$

In comparison with [11], setting $_1$ and λ_2 values, the fidelity comes out to be equal to $a/\sqrt{2}$, which implies that the difficulty in discriminating increases with the value of a. Again, in reference to the experiments conducted by [11], values of a are set to a = 0.25 and a = 0.5 which is employed in benchmarking the results in the paper.

Implementation The circuit designs primarily comprise gates from the universal gate set which includes gates like C-NOT and single-qubit gates. Given, these gates are most widely used in experiments, the authors have thus, decided to go with them. An interesting point to note here is that it is highly important to have an optimal number of C-NOT gates for implementing on near-term devices. In this case, the authors adapt the technique proposed in [[12],[13]] for finding near optimal circuit topologies. Apart from this, the quantum circuit is built as a generic POVM with four possible measurement outcomes. The measurement outcomes are denotes as $m_{i_2 i_1}$ where $i_1, i_2 \in 0, 1$. Here, i_1 and i_2 are the outcomes of first and second qubits respectively. So, the big picture looks something like this - we have three inputs $\psi_1(a)$, $\psi_{2/3}(b)$ and inconclusive outcome, and four possible outcomes. To solve the discrepancy, any outcome is arbitrarily chosen to form the correspondence to an input. The mapping order chosen in the paper is: m_{00} or m_{10} for $\psi_1(a)$; m_{01} for $\psi_{2/3}(b)$; and m_{11} for inconclusive outcome.

Training The proposed training procedure is a hybrid quantum-classical method that uses both a classical device and a simulated quantum computer. The steps can be delineated as follows:

1. A classical device generates a set of training data, which consists of pairs of non-orthogonal quantum states and their corresponding labels.
2. The classical device sends the training data to a quantum processor.
3. The quantum processor performs the POVM optimization and a series of measurements on the training data.
4. The classical device collects the measurement results, assess them and then uses them to update the parameters of the quantum circuit.

The training procedure is repeated for a number of epochs (basically the steps 2-4), where an epoch is a complete pass through the training data. The number of epochs is typically chosen to be large enough to ensure that the quantum circuit has converged to a good solution. Moreover, in Step 4, the overall optimization of the circuit is performed by the classical machine using the Adam [15] optimization algorithm. Similar to machine learning techniques, the authors also perform a train and test split. The training data is used for the following - computing the loss using the cost function (described in the below paragraph) and optimizing the quantum circuit (Step 4 above). The testing data is used for evaluating the performance of the QSD circuit.

Cost Function Cost function is what dictates the learning phase in any machine learning technique. The quality of a machine learning model highly depends upon the cost function or optimization algorithm used. The cost function used in the paper is defined as,

$$J_1 = \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} (1 - P_{succ}(\psi_i(a_i)))$$
$$+\alpha_{err} \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} P_{err}(\psi_i(a_i))$$
$$+\alpha_{inc} \sum_i \frac{1}{|S_i|} \sum_{a_i \in S_i} P_{inc}(\psi_i(a_i))$$

where S_i is a sample drawn from the distribution α_i, $|S_i|$ is the sample size, α_{err} is a penalty term for erroneous discrimination, α_{inc} is a penalty term for inconclusive outcomes. $P_{succ}(\psi)$, $P_{err}(\psi)$, $P_{inc}(\psi)$ are the output probabilities for success, error and inconclusive scenarios respectively, for any input state ψ.

4 Results

There are some interesting experiments being conducted in the paper. Besides verifying the claims of the proposed approach, the results provide great insight into how the quantum systems behave. Throughout all the experiments, the authors employ the values P_{succ}, P_{err}, and P_{inc}, to have a uniform comparison of the results. Moreover, the Adam [15] machine learning optimization algorithm is used in all the experiments, owing to the observation that stochastic gradient descent algorithm frequently gets stuck in the saddle points present in the training data.

One interesting observation from the experiment trying to classify a single data point was that, if we relax the constraints on P_{err} by allowing some non-zero error probability, it is possible to achieve a P_{inc} value close to the theoretical minima. This basically implies that there exists a trade-off between the error rate P_{err} and the inconclusive rate P_{inc} during training, and a little constraint relaxation on P_{err} can result in a significant reduction in the inconclusive rate.

In a follow-up (to the above observations) experiment, the authors tried to discriminate the inputs from two different classes of quantum states. In this, they varied the penalty parameters (α_{err} and α_{inc}), and observed that the penalties act like Machine learning regularization terms when compared to zero penalty values, that can be adjusted accordingly to increase P_{succ} or lower P_{inc}. Moreover, with the varying of penalty values, a gradual transition from unambiguous QSD to minimal error QSD was observed, which was expected yet quite exciting.

Is fidelity a good metric to measure the generalization ability of the quantum circuit? A very useful question in general. The authors carried out an experiment to verify their yes hypothesis, and it revealed the same observation as discussed in the Section 3.2 i.e. the difficulty in distinguishing $\psi_1(a)$ and $\psi_1(a)$ increases as value of a gets closer to 1. This corroborates their hypothesis that fidelity is indeed a good measure.

The most confusing experiment however, is the one where the authors test their quantum circuit on data not seen during optimization. They randomly draw parameters a and b from a probability distribution during the training process. The circuit's performance is tested by using data also sampled from the same distribution as the training data. It is observed that the trained model can accurately classify data drawn from various probability distributions such as the normal, uniform, or a mix of the two distributions. This is not a great experiment to verify the generalizability of the network, as the data, even if different, is still being drawn from the same distribution.

5 Quantum Repeaters

The idea here is mainly on the application of the approach discussed in the paper. That is, the practical implications of this work. The thought process is that, does the integration of the proposed approach into something like a Quantum repeater system fit well?

What are Quantum repeaters? They are just like the classical repeaters used in network communication, except it is the quantum version of it. Classical repeaters are inserted between two nodes and are used to improve signal strength by copying the signal from one node and retransmitting at a higher amplitude to the other node. Similar to this, with quantum repeaters we can have quantum internet transmitting data over long distances as well.

However, the fundamental problem with designing quantum repeaters in the same way as classical repeaters, is that, it cannot copy the data, owing to the no-cloning theorem. Instead, quantum repeaters use a technique called "entanglement swapping" [14] to operate. The basic idea is that this technique uses teleportation in the creation of a long-distance entanglement chain by utilizing multiple short-distance entanglements.

Now, say, we have a quantum repeater that is inserted between Alice and Bob on the two ends of a noisy quantum channel. Alice wants to send some information to Bob through this channel. Given, a noisy channel, one needs to distinguish the background messages in the noise from the actual message being sent, before amplifying the signal. Hence, one needs to repeatedly distil the messages as they start overlapping. This is now similar to a QSD problem. In a scenario like this, if we can integrate a generalized quantum circuit or model that can distinguish the incoming quantum data into a quantum repeater system, it would bring in significant improvements in quantum communication protocols. Either we develop a very generalized quantum circuit or train a circuit to perform such message distillations.

Although, the circuit training needs a very secured environment, once implemented, it can open up doors to many other application areas as well.

6 Conclusions

The authors develop a method for training near-term quantum circuits to classify data represented by non-orthogonal quantum probability distributions. The main point of the paper is to devise a mechanism to train circuits to simulate the structure of a generalized POVM. The method is based on a hybrid quantum-classical algorithm that involves both, a classical and quantum device. This is actually a very interesting idea despite not being a purely quantum system. Although quantum algorithms show better speedup compared to classical algorithms, especially in algebra related operations, they have limitations when applied to classical data, specifically in the preparation of quantum states that encode classical information. This process can have a worst-case polynomial scaling in the input dimension of the data, which ultimately reduces the advantage of quantum learning over classical learning. Hence, a hybrid method is a very unique way to approach the discrimination problem.

Furthermore, the authors show that their method can be used to train quantum circuits to classify data from various different probability distributions, approaching the optimal theoretical performance when such bounds are available. The authors also focus on the generalization ability of their circuit, i.e., they train the circuit on a specific range of the parameters with the goal of maximizing its generalization performance.

Overall, their method seems to be a very promising new approach for training quantum circuits for a variety of tasks.

References

[1] Hongxiang Chen, Leonard Wossnig, Simone Severini, Hartmut Neven, and Masoud Mohseni. "Universal discriminative quantum neural networks." Quantum Machine Intelligence 3 (2021): 1-11.

[2] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, "Quantum machine learning: a classical perspective," Proc. R. Soc. A, vol. 474, no. 2209, p. 20170551, 2018.

[3] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, "Quantum machine learning," Nature, vol. 549, no. 7671, p. 195, 2017.

[4] M. Schuld, I. Sinayskiy, and F. Petruccione, "An introduction to quantum machine learning," Contemporary Physics, vol. 56, no. 2, pp. 172–185, 2015.

[5] C. H. Bennett, "Quantum cryptography using any two nonorthogonal states," Physical review letters, vol. 68, no. 21, p. 3121, 1992.

[6] V. Giovannetti, S. Lloyd, and L. Maccone, "Advances in quantum metrology," Nature Photonics, vol. 5, pp. 222 EP –, 03 2011.

[7] J. Chen, J. L. Habif, Z. Dutton, R. Lazarus, and S. Guha, "Optical codeword demodulation with error rates below the standard quantum limit using a conditional nulling receiver," Nature Photonics, vol. 6, pp. 374 EP –, 05 2012.

[8] https://www.cs.cmu.edu/ odonnell/quantum15/lecture17.pdf

[9] Y. Eldar, "A semidefinite programming approach to optimal unambiguous discrimination of quantum states," IEEE Transactions on Information Theory, vol. 49, pp. 446–456, feb 2003.

[10] Y. Eldar, A. Megretski, and G. Verghese, "Designing optimal quantum detectors via semidefinite programming," IEEE Transactions on Information Theory, vol. 49, pp. 1007–1012, apr 2003.

[11] M. Mohseni, A. M. Steinberg, and J. A. Bergou, "Optical Realization of Optimal Unambiguous Discrimination for Pure and Mixed Quantum States," Physical Review Letters, vol. 93, p. 200403, nov 2004.

[12] R. Iten, R. Colbeck, I. Kukuljan, J. Home, and M. Christandl, "Quantum Circuits for Isometries," Physical Review A - Atomic, Molecular, and Optical Physics, jan 2015.

[13] R. Iten, R. Colbeck, and M. Christandl, "Quantum Circuits for Quantum Channels," Physical Review A - Atomic, Molecular, and Optical Physics, vol. 93, p. 052316, sep 2016.

[14] https://www.aliroquantum.com/blog/what-are-quantum-repeaters

[15] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).