Reinforcement Learning with Neural Networks for Quantum Multiple Hypothesis Testing

Sarah Brandsen[1], Kevin D. Stubbs[2], and Henry D. Pfister[2,3]

[1]Department of Physics, Duke University, Durham, North Carolina 27708, USA.
[2]Department of Mathematics, Duke University, Durham, North Carolina 27708, USA
[3]Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina 27708, USA

Reinforcement learning with neural networks (RLNN) has recently demonstrated great promise for many problems, including some problems in quantum information theory. In this work, we apply RLNN to quantum hypothesis testing and determine the optimal measurement strategy for distinguishing between multiple quantum states $\{\rho_j\}$ while minimizing the error probability. In the case where the candidate states correspond to a quantum system with many qubit subsystems, implementing the optimal measurement on the entire system is experimentally infeasible.

We use RLNN to find locally-adaptive measurement strategies that are experimentally feasible, where only one quantum subsystem is measured in each round. We provide numerical results which demonstrate that RLNN successfully finds the optimal local approach, even for candidate states up to 20 subsystems. We additionally demonstrate that the RLNN strategy meets or exceeds the success probability for a modified locally greedy approach in each random trial.

While the use of RLNN is highly successful for designing adaptive local measurement strategies, in general a significant gap can exist between the success probability of the optimal locally-adaptive measurement strategy and the optimal collective measurement. We build on previous work to provide a set of necessary and sufficient conditions for collective protocols to strictly outperform locally adaptive protocols. We also provide a new example which, to our knowledge, is the simplest known state set exhibiting a significant gap between local and collective protocols. This result raises interesting new questions about the gap between theoretically optimal measurement strategies and practically implementable measurement strategies.

1 Introduction

Optimal quantum hypothesis testing consists of finding the quantum measurement $\{\Pi_j\}|_{j=1}^m$ to optimally distinguish between m candidate states $\{\rho_j\}_{j=1}^m$ with prior probabilities $\{q_j\}_{j=1}^m$. For example, this can be used to discriminate between coherent quantum states [1] and also to decode one of m codewords that has been sent through a known noisy quantum channel [2, 3]. One important example of locally adaptive multiple hypothesis testing protocols is the Dolinar receiver, which uses an adaptive measurement scheme to distinguish between m different optical signals [4].

Sarah Brandsen: sarah.brandsen@duke.edu

Although the optimal (Helstrom) measurement has a compact expression when $m = 2$, the solution is more complicated for general non-binary state discrimination. In general, the optimal measurement can be written as the solution of a semidefinite programming problem [5, 6]. Techniques for solving semidefinite programming then can be used to find the minimal-error measurement and compute the optimal success probability [7, 8].

When the candidate states are high-dimensional (corresponding to a quantum system composed of many qubit subsystems), it can be experimentally difficult to implement operations on all subsystems at once. Thus, we also focus on finding optimal (or near-optimal) approaches that include the experimentally necessary property of locality, where only a single subsystem is measured in each round. We know that dynamic programming can be used to find an optimal local approach [9]. However, even in the simplest case where $m = 2$, the complexity grows like $O(2^n nQ)$, where n is the number of qubit subsystems and Q is the number of different local measurements considered [10].

A powerful alternative tool for developing optimal adaptive protocols is reinforcement learning with neural networks (RLNN), where an agent learns an optimized protocol through repeated interaction with an environment. While RLNN was introduced more than 20 years ago [11, 12], interest in these methods was recently rekindled by its remarkable success for Atari games [13, 14]. RLNN and other machine-learning approaches have been successfully applied to a variety of problems in quantum information theory: generating error-correcting sequences [15, 16], preparation of special quantum states [17–19], setting up experimental Bell tests [20], quantum communication [21], fault-tolerant quantum computation [22], quantum control [23–26], and nonequilibrium quantum thermodynamics [27]. Additionally, RLNN has been applied in the closely related topic of adaptive quantum metrology [28–31]. Motivated by these successes, in this work we use RLNN to find optimal locally-adaptive measurement protocols.

To demonstrate the effectiveness of reinforcement learning, we compare the RLNN performance to other locally adaptive and collective protocols. In all our numerical results, the neural network meets or exceeds the probability of success achieved by a modified locally greedy approach. Additionally, for every simulation with randomly generated candidate state sets, the RLNN scheme approximately meets an upper bound corresponding to the optimal collective success probability (i.e., the optimal measurement scheme when measurements are not restricted to be local). This upper bound is found via semidefinite programming (SDP) techniques outlined in [32]. The RLNN performance as a function of subsystem number is investigated, and we demonstrate that the neural network attains good performance for up to 20 subsystems. We additionally show that, for any locally adaptive method, the success probability is stable under small perturbations of the candidate state sets, such as rotation errors.

While our numerical tests show that RLNNs are a powerful tool for calculating an optimal or near-optimal locally-adaptive strategy, we additionally provide analytical results for some specific systems. These specific results help complete the picture regarding the optimality of locally adaptive protocols in four key regimes: pure binary state discrimination, mixed binary state discrimination, pure non-binary state discrimination, and mixed non-binary state discrimination. Prior to this work it was known that locally adaptive, projective measurement strategies are optimal for pure binary state discrimination [10, 33] and are in general are *not* optimal for both pure and mixed nonbinary state discrimination [34, 35]. In contrast to the pure state case, few analytical results are known for mixed binary state discrimination. Previous work had shown that certain fixed strategies were not optimal for mixed binary states [36] and had suggested via numerics [37] that any quantized locally adaptive strategy may be suboptimal. To our knowledge, we are the first

to prove this analytically. Moreover, while previous work discusses a potential gap in the case of multiple subsystems where both states are noisy, we demonstrate that a gap can exist for *any* nontrivial subsystem number and only one mixed state.

Still, the main significance of our paper lies in applying novel machine learning methods to quantum state discrimination. To our knowledge, we provide the first algorithm capable of successfully distinguishing between an arbitrary candidate state set via locally adaptive protocols. Our result demonstrates that not only is the optimal locally adaptive measurement strategy unknown for general state discrimination problems, but additionally the optimal locally adaptive success probability is also unknown. This further motivates the need for a reliable algorithm which can always find the optimal or close-to-optimal locally adaptive strategy, which our RLNN provides. Such an algorithm would be of broad interest not only to researchers working in the field of quantum hypothesis testing, but also to any researchers who are applying machine learning techniques to similar problems in other areas of physics.

2 Applying Reinforcement Learning to Quantum Hypothesis Testing

Each round of the reinforcement learning process involves an agent choosing one action from an allowed action space, implementing the action, and receiving a reward from the environment. For a Markov decision process, the agent can eventually learn to choose actions according to an optimal policy that maximizes the expected future reward. For the problem at hand, the agent is trained to learn the optimal adaptive measurement strategy as well as the optimal adaptive order in which subsystems should be measured.

In the context of state discrimination, the environment is a parameterized measurement protocol for the quantum system of interest. The action space (denoted by $\mathcal{A}$) is the set of allowed quantum measurements. Denote by s_t the state of the environment just before round t and let n be the total number of rounds. The agent's policy, $\pi_\theta(a_t|s_t)$, is parameterized by θ and equals the probability of selecting action $a_t \in \mathcal{A}$ in round t conditioned on the state s_t of the environment. The goal of training is for the agent to learn the optimal policy π_θ^* which maximizes a given reward function.

We consider the task of deriving the minimum-error adaptive measurement protocol to distinguish between m tensor-product quantum states $\{\rho_j\}|_{j=1}^m$ with prior probability vector $\boldsymbol{q}$ where $q_j = \Pr(\rho = \rho_j)$. To reduce the number of measurement parameters and thus the size of the action space, we restrict to the case where each candidate state is real-valued. Since each candidate state is assumed to be a tensor product of n subsystems, it can be written as

$$\rho_j = \bigotimes_{k=1}^{n} \rho_j^{(k)},$$

where $\rho_j^{(k)}$ is a qubit density matrix for all $j \in \{1, ..., m\}$ and all $k \in \{1, ..., n\}$. Thus, the quantum system ρ is composed of n unentangled qubit subsystems.

We build an OpenAI gym environment [38] capable of simulating local measurement protocols. In each round, the algorithm chooses the next subsystem j to measure as well as which measurement to implement.

The action space $\mathcal{A}$ consists of elements (ℓ, k) where $\ell \in \{1, ..., 20\}$ selects which measurement in the allowed measurement set is to be implemented and $k \in \{1, ..., n\}$ is the

subsystem to be measured. More specifically, ℓ corresponds to implementing the binary real qubit POVM,

$$\hat{\Pi}_Q(\ell) \triangleq \left\{ \begin{pmatrix} \sin^2(\frac{\ell\pi}{2Q}) & \frac{1}{2}\sin(\frac{\ell\pi}{Q}) \\ \frac{1}{2}\sin(\frac{\ell\pi}{Q}) & \cos^2(\frac{\ell\pi}{2Q}) \end{pmatrix}, \begin{pmatrix} \cos^2(\frac{\ell\pi}{2Q}) & -\frac{1}{2}\sin(\frac{\ell\pi}{Q}) \\ -\frac{1}{2}\sin(\frac{\ell\pi}{Q}) & \sin^2(\frac{\ell\pi}{2Q}) \end{pmatrix}, \right\}$$

and $Q = 20$ (unless otherwise specified). The set of allowed measurement is $\{\hat{\Pi}_Q(\ell)\}|_{\ell=1}^{Q}$ which corresponds to binary real qubit POVMs spaced evenly on the Bloch sphere. For a given Q, this action set minimizes the worst case quantization error. Increasing the quantization beyond $Q = 20$ (or allowing continuous choice of measurement) slowed the training time and did not offer observable gain in performance, so $Q = 20$ was chosen as the smallest quantization which yields near optimal results.

For a given set of candidate states, the state set $\mathcal{S}$ consists of elements $(\mathbf{p}, \mathbf{v})$ where $\mathbf{p}$ denotes the updated probabilities for each candidate state and $\mathbf{v}$ is a length-n vector which specifies which subsystems have been measured. More specifically, given starting prior $\mathbf{q}$ and measurement results $\mathbf{d}$, the updated prior is denoted by $p(\mathbf{q}, \mathbf{d})$. The list of subsystems which have been measured is given by the length-n vector $\boldsymbol{v}$ where $v_k = 1$ if subsystem k has already been measured and 0 otherwise. Thus, the overall state of the environment, given starting prior $\mathbf{q}$ and measurement history $\mathbf{d}$, may be represented as $s \triangleq (p(\boldsymbol{q}, \boldsymbol{d}), \boldsymbol{v})$. The episode is terminated when all subsystems except one have been measured, or equivalently when $\sum_i \boldsymbol{v}_i = n - 1$.

When only one subsystem remains unmeasured the optimal final measurement is automatically determined through semidefinite programming. The reward is given by the probability of successfully decoding the actual state ($\rho = \rho_{j^*}$) after the final local measurement, where a successful decoding occurs if

$$j^* = \underset{j \in \{1,\ldots,m\}}{\operatorname{argmax}} \left(p_j(\boldsymbol{q}, \boldsymbol{d})\right),$$

where $\boldsymbol{d}$ is the vector containing all previous measurement results and $\boldsymbol{p}(\boldsymbol{q}, \boldsymbol{d})$ is the updated probability given initial prior $\mathbf{q}$ and measurement results $\boldsymbol{d}$. Additionally, in each round a penalty of -0.3 is given if the agent attempts to re-measure an already measured subsystem, as for qubit subsystems re-measuring an already measured subsystem is non-informative.

3 Details of Implementation

We train the agent using the proximal policy optimization (PPO) algorithm [39]. From numerical simulations, we found that a PPO algorithm significantly outperformed a DQN-based algorithm which had poor performance and training instability. Given that the environment in our problem is not too expensive to sample from, PPO worked well and we did not try DDPG or SAC, which can be more sample efficient. PPO algorithms generally train well on problems with discrete action spaces and environments that are cheap to sample from. PPO algorithms additionally offer the benefit of relatively straightforward hyperparameter tuning and hyperparameters did not need to be re-tuned for each combination of (m, n).

Results are then generated using the default PPO algorithm from the RLlib package included in Ray version 0.7.3 [40, 41]. After hyperparameter tuning of the learning rate, we set the learning rate to be $\eta = 5 \times 10^{-5}$. For the remaining hyperparameters, we find the default parameter settings to be optimal, including the clipping parameter $\epsilon = 0.3$ and

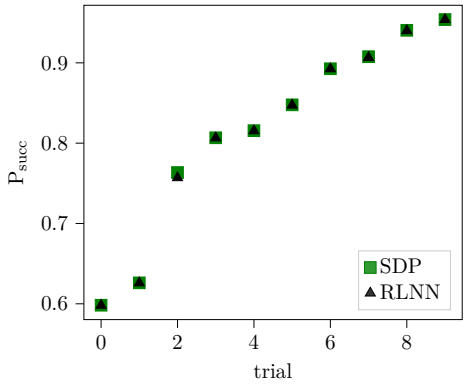

Figure 1: Performance of tuned network versus the SDP upper bound on the optimal success probability.

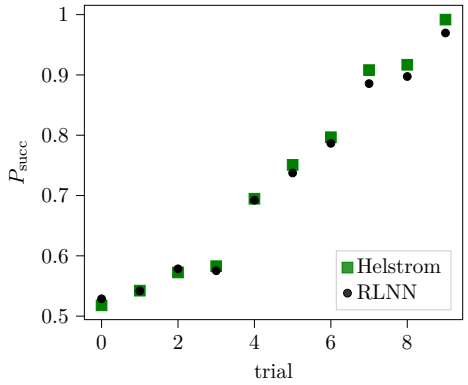

Figure 2: Performance of network before tuning versus the SDP upper bound on the optimal success probability.

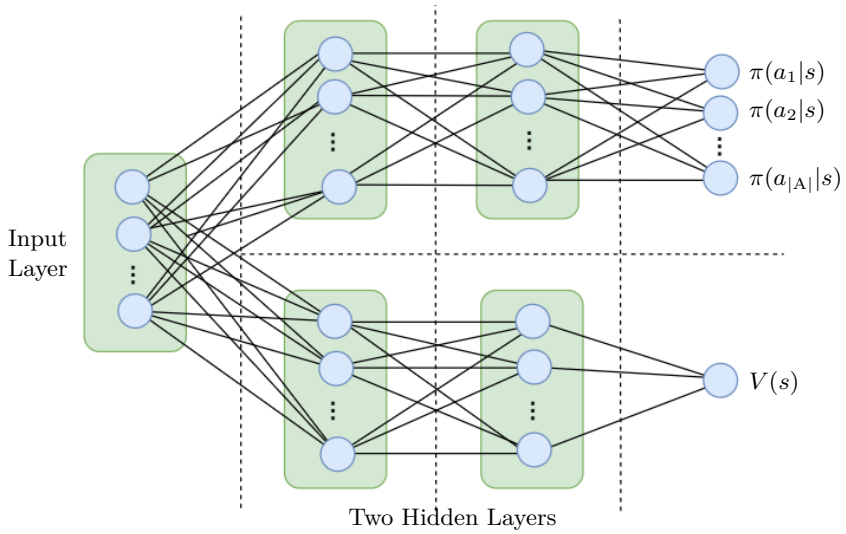

Figure 3: Neural Network configuration, consisting of one input layer, two parallel subnetworks. One subnetwork outputs an estimate of the value $V(s)$ for state s, one outputs an estimate of the policy $\pi(a|s)$.

the discount factor $\gamma = 0.99$. Comparison of the tuned versus untuned training is depicted in Figures 1 and 2.

We use a fully connected neural network where the input layer (with $m+n-1$ neurons) takes the state s as the input, as depicted in Fig. 12. This feeds into two parallel sets of subnetworks, each of which has two hidden layers of 256 neurons, tanh activation functions, and their own linear output layers. Although additional hidden layers could be added, this would increase the training time required. The output layer of the first subnetwork consists of a single neuron, and computes an estimate for the value of states. The output layer of the second subnetwork has nQ neurons (i.e. the number of allowed actions) and computes the policy, $\pi(a|s)$. See https://github.com/SarahBrandsen/RLNN-QSD for numerical results and the source code used to obtain them.

4 Numerical Results for RLNN Performance

As an initial benchmark of the RLNN performance, we compare it to known optimal results in several special cases.

In the case of binary discrimination (i.e. $m = 2$) between tensor products of pure states such that $\rho_j^{(k)} = |\psi_j^{(k)}\rangle\langle\psi_j^{(k)}|$ for all $k \in \{1, ..., n\}$ and $j \in \{1, 2\}$, it has been shown that the optimal collective success probability, $\mathrm{P}_{\mathrm{SDP}}$ can be achieved through locally-adaptive strategies [10, 33]. The collective success probability, $\mathrm{P}_{\mathrm{SDP}}$, is found using semidefinite programming techniques introduced by [32]. We randomly generate ten trials with $n = 3$ and order the trials according to increasing distinguishability measured by $\mathrm{P}_{\mathrm{SDP}}$. For each trial, we compare this success probability with the RLNN success probability, $\mathrm{P}_{\mathrm{RLNN}}$, as shown in Fig. 4. The neural network attains the correct (optimal) success probability in each case, with a very small gap that is likely due to action space quantization.

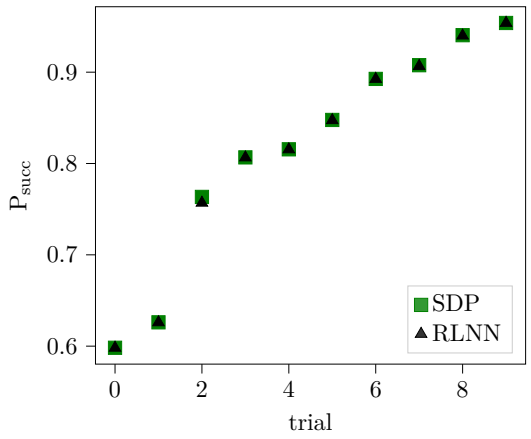

Figure 4: Probability of success for the optimal RLNN policy after 1000 training iterations vs. the optimal collective measurement for tensor-products of pure states when $m = 2$, $n = 3$. The neural network closely approximates the optimal success probability in each trial, with any gap likely arising from quantization of the action space.

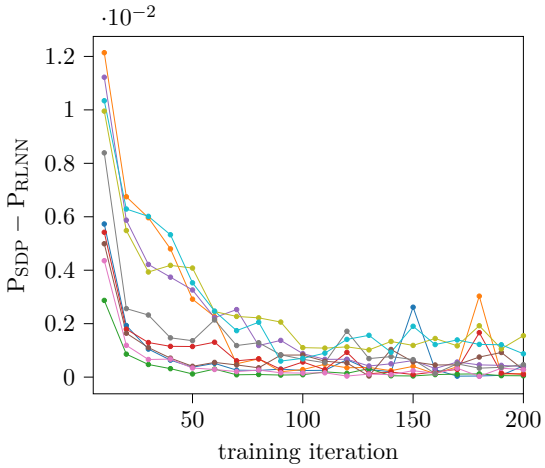

Figure 5: Difference between RLNN reward and Helstrom (SDP) success probability as a function of training iteration. We observe that the RLNN success probability stabilizes after 100 training iterations, with occasional fluctuations.

An additional case where locally adaptive protocols are strictly optimal has been found by Sasaki et. al in [42]. Consider a set of states $\mathcal{S}_1 \triangleq \{\rho_j\}|_{j=1}^m$ and associated probabilities $\{q_j\}|_{j=1}^m$. Suppose the known optimal POVM for these is $\{\Pi_j\}|_{j=1}^m$. The set of n-subsystem product states generated by $\mathcal{S}$ can be written as

$$\mathcal{S}_n \triangleq \left\{ \bigotimes_{j=1}^{n} \rho_{i_j} \;\middle|\; i \in \{1, ..., m\}^n \right\},$$

with corresponding probabilities defined as $q_{i_1 ... i_n} \triangleq q_{i_1} \times ... \times q_{i_n}$. Then, the optimal POVM candidate state set $\mathcal{S}_n$ has elements that can be written in tensor product form as:

$$\Pi_{i_1 ... i_n} = \bigotimes_{j=1}^{n} \Pi_{i_j}.$$

This provides a useful test of the neural network performance. We take the initial state set to be $\mathcal{S}_1 = \{\rho_1, \rho_2\}$, where $\rho_1 = \left(\begin{smallmatrix} 0.85 & 0 \\ 0 & 0.15 \end{smallmatrix}\right)$ and $\rho_2 = \left(\begin{smallmatrix} 0.15 & 0 \\ 0 & 0.85 \end{smallmatrix}\right)$. Since the optimal local

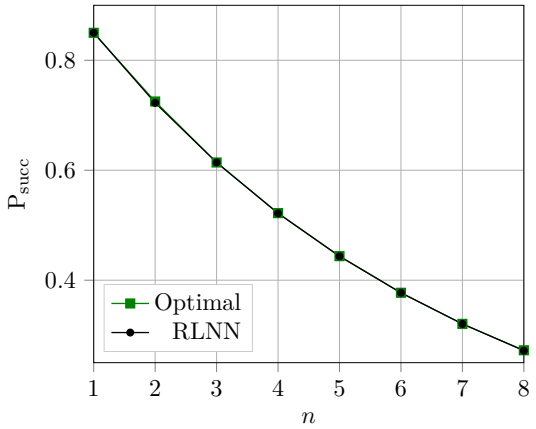

Figure 6: Performance of the RLNN policy after 150 training iterations vs. the optimal success probability as a function of the number of subsystems n. The RLNN approach converges to the optimal local approach in this example for $1 \leq n \leq 8$.

POVM belongs to the allowed action set, there should be no quantization loss. We train the neural network for 1000 iterations (using a custom learning rate schedule where the learning rate starts at 5.5×10^{-5} and decays by 0.95 every 10 iterations), and compare the neural network performance after training to the optimal success probability. For $1 \leq n \leq 8$, the neural network attains or approximately attains the exact success probability, as depicted in Fig. 6.

5 Comparison to SDP-based Locally Adaptive Strategies

Just as the collective SDP measurement provides an upper bound for the optimal locally adaptive success probability, simple locally adaptive algorithms such as locally greedy algorithms provide a lower bound. In this section, we introduce a local SDP-based approach, and demonstrate numerically that the RLNN always meets or exceeds the success probability of the local SDP-based approach.

In the case of binary state discrimination, locally greedy algorithms are optimal for pure states and close-to-optimal for mixed states. Our choice of the local SDP-based algorithm as a "good" simple strategy is motivated by the fact that it reduces to a locally greedy protocol when $m = 2$. Additionally, for $m > 2$, we found through numerical simulation that the local SDP-based based approach generally performs better than locally greedy protocols. Finally, we compare the RLNN algorithm to the local SDP-based algorithm via simulations with $n = 3$ and $n = 4$ and demonstrate that the RLNN always meets or significantly exceeds the local SDP-based success probability.

The SDP-based local algorithm selects the local measurement which maximizes the expected (collective) success probability of future rounds. Let $\mathcal{S}$ be the set of remaining unmeasured subsystems. For each round, the algorithm chooses to measure the subsystem $l \in \mathcal{S}$ and implement the measurement a such that

$$(a, l) = \underset{(a,\ell)\in\mathcal{A}\times\mathcal{S}}{\operatorname{argmax}} \sum_{d'=0}^{|a|} \Pr(d_{n-|\mathcal{S}\setminus\ell|} = d') \times \mathrm{P}_{\text{succ, coll}}\left(\{\rho_j^{\mathcal{S}\setminus l}\} \mid q, d_{n-|\mathcal{S}\setminus\ell|} = d'\right),$$

where $\mathrm{P}_{\mathrm{succ,\,coll}}(\{\rho_j^{\mathcal{S}\backslash l}\} \mid q, d_{n-|\mathcal{S}\backslash\ell|} = d')$ equals the success probability of implementing an optimal collective measurement on the remaining subsystems (with indices belonging to set $\mathcal{S}\backslash l$), given the prior for round j was q and the outcome of action a was d'.

In the special case where $m = n = 3$, all candidate states are pure states, and all subsystems identical copies, the performance of the SDP-based local algorithm and the RLNN algorithm appear to be identical for each of 5 random trials, as depicted in Fig. 7. However, we demonstrate that simpler locally adaptive strategies such as the min-entropy approach are not sufficient to find the optimal locally adaptive strategy, as when $n = 4$ and the candidate states are mixed, a significant gap appears between the RLNN results and the SDP-based local algorithm, as shown in Fig. 8.

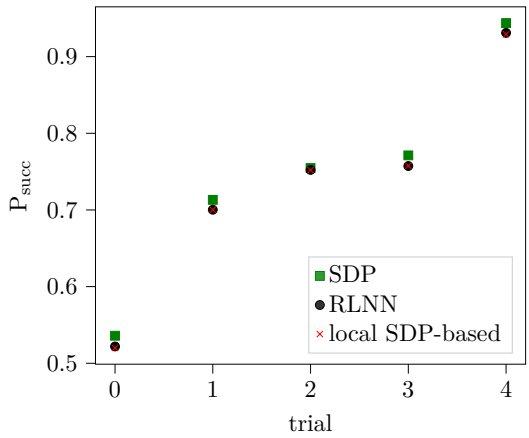

Figure 7: Plot of success probability for RLNN after 250 training iterations, the collective optimal (SDP) measurement, and the SDP-based local algorithm, over 5 trials with $m = 3$, $n = 3$ and pure states.

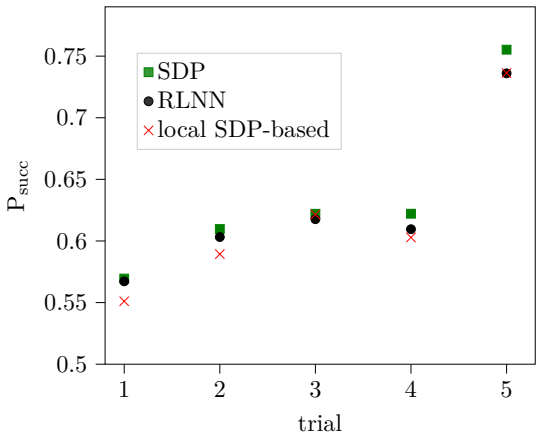

Figure 8: Plot of success probability for RLNN after 250 training iterations, the collective optimal (SDP) measurement, and the SDP-based local algorithm, over 5 trials with $m = 3$, $n = 5$.

6 Pure State Discrimination

In the special case of binary state discrimination ($m = 2$), it has been shown [10, 33] that locally-greedy algorithms are optimal for distinguishing between pure tensor product states. Thus, for pure binary state discrimination, the success probability of the optimal collective measurement can be achieved through a simple locally-greedy algorithm. It has additionally been demonstrated that for $m \geq 3$, there exist simple state sets such that the optimal locally adaptive algorithm performs worse than the optimal collective measurement [44].

We now build on these results to determine whether, for a given m and n, there exists a candidate state set such that there exists a significant gap between the optimal locally adaptive and optimal collective measurement.

Theorem 1 *Denote by $P_{\mathrm{loc}}(\{\rho_j\}, \boldsymbol{q})$ the optimal probability of success using locally adaptive projective measurements for candidate state set $\{\rho_j\}$ with prior probability vector $\boldsymbol{q}$. Likewise, denote by $P_{\mathrm{coll}}(\{\rho_j\}, \boldsymbol{q})$ the success probability for the optimal collective measurement on the full quantum system. Then for a given m and a given $n > 1$, there exists at least one set of tensor product states $\{\rho_j = \bigotimes_{k=1}^{n} \rho_j^{(k)}\}|_{j=1}^{m}$ and some starting prior $\boldsymbol{q}$ such that $P_{\mathrm{loc}}(\{\rho_j\}, \boldsymbol{q}) < P_{\mathrm{coll}}(\{\rho_j\}, \boldsymbol{q})$ if and only if at least one of the following conditions is met:*

1. *There are more than two candidate states ($m > 2$)*

2. *At least one candidate state is not a pure state.*

Proof Sketch: The full proof is listed in Appendix A. As a proof sketch, we note that the "only if" direction of the proof follows immediately from [43], where it was shown that locally adaptive methods are optimal for any binary pure state discrimination problem (including discrimination problems involving entangled states.)

To show the "if" direction, we introduce a simple binary state discrimination with two qutrit subsystems and demonstrate that a significant gap exists for the following candidate states:

$$\rho_+ \triangleq \Big(\frac{1}{2}\left|0\right\rangle\!\left\langle 0\right| + \frac{1}{2}\left|1\right\rangle\!\left\langle 1\right|\Big) \otimes \Big(\frac{1}{2}\left|0\right\rangle\!\left\langle 0\right| + \frac{1}{2}\left|1\right\rangle\!\left\langle 1\right|\Big)$$
$$\rho_- \triangleq \frac{1}{3}\Big(\sum_{j=0}^{2} \left|j\right\rangle \otimes \left|j\right\rangle\Big)\Big(\sum_{k=0}^{2} \left\langle k\right| \otimes \left\langle k\right|\Big).$$

The "if" direction is completed by previous results which demonstrate a significant gap can exist even in the case of three two-qubit candidate states [44].

7 Gap between Locally Optimal Algorithm and Collective Measurement

Finally, we use RLNN to estimate the gap between the best locally adaptive algorithm and the optimal collective (non-local) measurement in more general cases where the best locally adaptive algorithm is not otherwise known.

The simulation setup for a given m and n is as follows: for each trial, we randomly generate pure tensor product candidate states and then apply depolarizing noise with a randomly chosen noise parameter. The RLNN algorithm is independently trained 5 times over 2000 iterations, and the average final success probability is compared (with error bars) to the optimal collective success probability found via SDP. Results are plotted for $m = 2$, $n = 3$ in Figure 9 and for $m = 3$, $n = 3$ in Figure 10, and indicate that the gap between local and collective measurements increases with m.

8 Performance for a Large Number of Subsystems

We examine how the RLNN performance varies as a function of n, and demonstrate good performance for up to $n = 10$ subsystems. However, for $n \geq 20$, the RLNN begins to have suboptimal performance.

First, we consider the case of binary pure state discrimination, where a locally-greedy (LG) technique is known to be optimal. We restrict the LG algorithm to the same action space as the RLNN to remove any gap due to action space quantization, and compare the resulting success probabilities. Results are depicted in Fig. 12, and indicate that the RLNN matches or almost matches the LG algorithm when $n = 10$ but develops a performance gap for $n = 20$.

Next, we consider the performance of multiple state discrimination where $m = 3$. For $n < 8$ we can compare directly to the collective SDP. For $n \geq 8$, computing P_{succ} via SDP techniques is infeasibly slow, so we instead look at the RLNN training curve shape to determine whether the neural network converges to a steady solution.

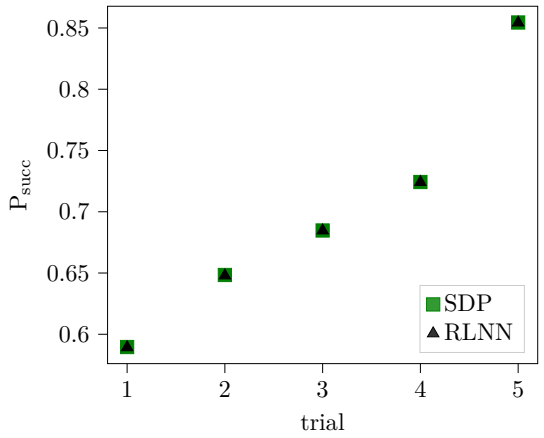

Figure 9: Probability of success for SDP and RLNN when $m = 2$ and $n = 3$. For each trial, the RLNN success probability is computed by separately training the neural network five times with 2000 iterations each. The error bars, were they visible, would represent the standard deviation in the final success probability over the five independent trainings. But in all trials, the error bars have collapsed to nothing, and the gap between local and non-local measurements is very small.

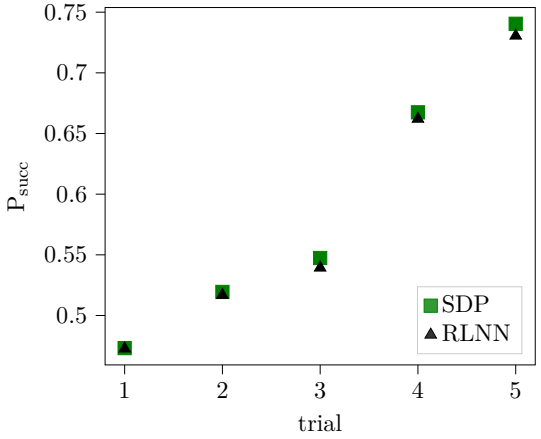

Figure 10: Probability of success for SDP and RLNN after 2000 training iterations when $m = 3$, $n = 3$. For each trial, the RLNN success probability is computed by separately training the neural network five times with 2000 iterations each. In all trials, the error bars have collapsed to nothing and the gap between local and non-local measurements is very small. Compared to the case where $m = 2$, there is a slightly larger gap between local and non-local measurements.

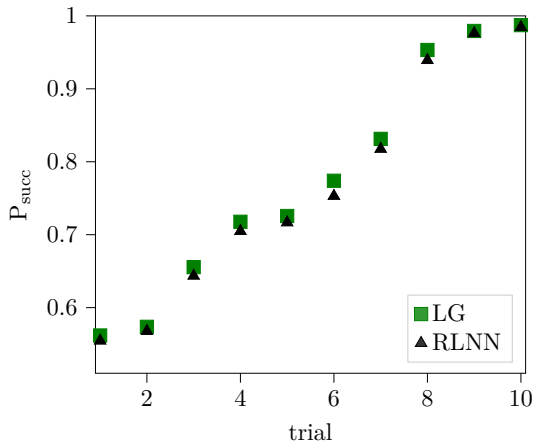

Figure 11: Success probability for $m = 2$, $n = 10$ where all candidate states are pure. Success probability for the RLNN is based on 750 training iterations for each round.

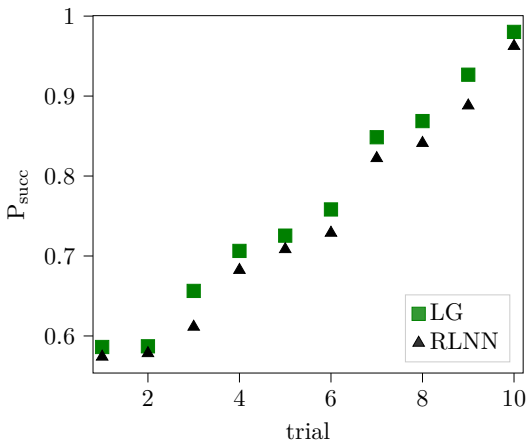

Figure 12: Success probability for $m = 2$, $n = 20$ where all candidate states are pure. Success probability for the RLNN is based on 1000 training iterations for each round.

As a first test of RLNN performance when $m = 3$, we consider the case of pure states with $n = 5$ and plot the success probability vs SDP as well as the training curves in Fig. 13.The RLNN success probability plateaus after approximately 100 training iterations, and the RLNN success probability comes close to the collective SDP in each case.

We then examine the training curves for general state discrimination when $m = 3$ as a function of n, with results shown in Fig. 17. Although stable plateaus are reached for both $n = 10$ and $n = 20$, as n increases the RLNN spends more training iterations learning not

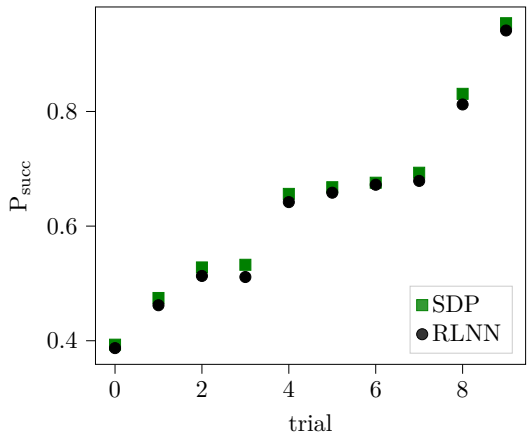

Figure 13: 10 trials with $m = 3$, $n = 5$. Success probability for RLNN after 300 training iterations compared to success probability of SDP

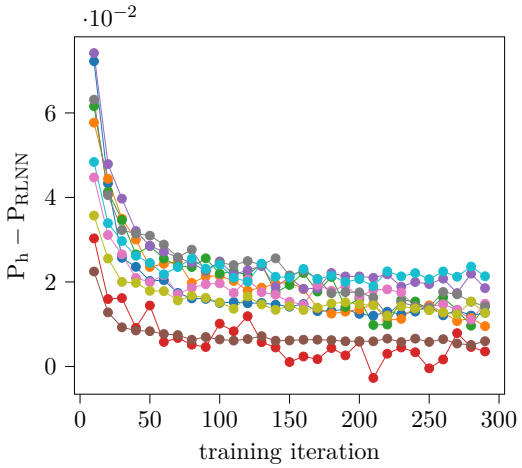

Figure 14: Difference between RLNN reward and SDP success probability as a function of training iteration. We observe that the RLNN success probability approximately plateaus after 100 training iterations

to re-measure subsystems. In the case where $n = 50$, approximately 500 training iterations are spent learning not to re-measure subsystems, leading to a negative initial reward. This leads to the question of whether it is possible to extend the RLNN performance to a larger number of subsystems by predetermining a close-to-optimal ordering,

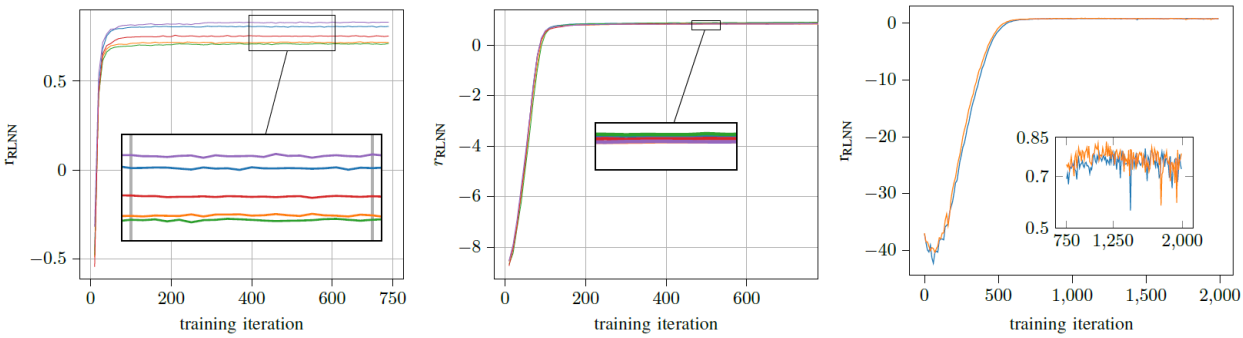

Figure 15: Training curves for five independent trials where $n = 10$ (right), $n = 20$ (centre), and one trial of $n = 50$ (left). As n increases, the training curve becomes less stable, and the number of iterations required for the reward to reach its maximal value increases. For $n = 50$, the shape of the training curve changes to include an initial dip in reward before convergence to the ideal and we also observe less stability.

9 Robustness under noise

We demonstrate that the success probability is stable when the candidate states are subject to a small perturbation. Consider an over-rotation noise model where the perturbation is parametrised by rotation angle θ with

$$\tilde{\rho}_j^{(k)}(\theta) = U(\theta)\rho_j^{(k)}U^\dagger(\theta),$$

where $\tilde{\rho}_j^{(k)}(\theta)$ is the noisy state and we set the rotation matrix as $U(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$. We prove that the error due to noise is negligible for sufficiently small enough perturbations, indicating that the RLNN adaptive method is still close-to-optimal when the candidate states are subjected to sufficiently small amounts of unitary noise.

Theorem 2 *Consider candidate set* $\{\rho_j\}|_{j=1}^m$ *with prior* $\mathbf{q}$. *Denote by* $P_{succ}\Big(\{\rho_j\}\Big)$ *the probability of success using the optimal locally adaptive method on the original state set. Likewise, let* $P_{succ}\Big(\{\tilde{\rho}_j(\theta)\}\Big)$ *be the success probability for the noisy state set. Then for all* θ,

$$\left|P_{succ}\Big(\{\rho_j\}\Big) - P_{succ}\Big(\{\tilde{\rho}_j(\theta)\}\Big)\right| \leq 4n \sin\left(\frac{|\theta|}{2}\right),$$

where n *is the number of subsystems.*

Proof: See Appendix B for a complete proof.

Finally, we generate five candidate state sets with $m = 3$, $d = 2$. For each candidate state set $\{\rho_j\}$, we train the neural network to find the optimal locally adaptive method. The original adaptive measurement scheme is then applied to the rotated state set $\{\tilde{\rho}_j(\theta)\}$, and we plot the gap in success probabilities $\text{diff}(\theta) \triangleq \text{P}_{\text{succ}}\Big(\{\rho_j\}\Big) - \text{P}_{\text{succ}}\Big(\{\tilde{\rho}_j(\theta)\}\Big)$.

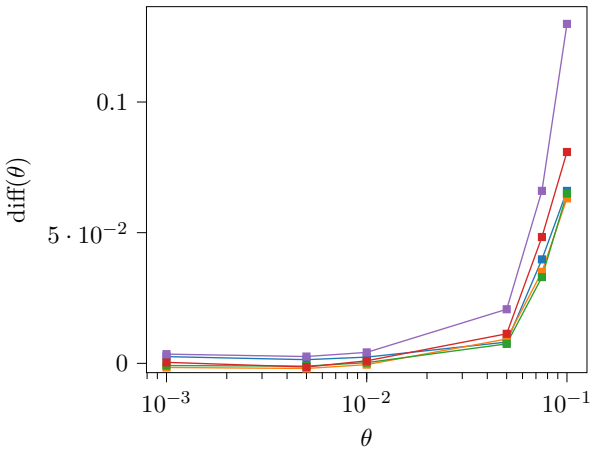

Figure 16: Gap in success probability as a function of rotation parameter θ for five trials where $m = n = 3$. We observe that the gap is negligible for sufficiently small θ.

10 Conclusion

We apply RLNN to calculate a near-optimal locally-adaptive measurement scheme for multiple state discrimination. We provide preliminary results for the neural network performance in cases where the locally-adaptive probability of success is known, and show that the network can achieve good performance when the total number of subsystems is 10 or fewer. This performance holds even when the candidate states are subjected to small perturbations. For cases where the exact locally optimal protocol is not known, we compare the RLNN performance to an SDP upper bound and find that for each trial the RLNN comes close to the upper bound. Additionally, we introduce a min-entropy based locally

adaptive approach which reduces to the optimal local approach for binary pure states, and show that the RLNN meets or exceeds this approach in every trial.

Finally, we characterize types of candidate state sets where there is a gap between optimal locally adaptive algorithms and optimal collective algorithms. While previous work has demonstrated gaps for candidate state sets with a more complex structure, we provide state sets where a gap exists for the simplest possible case of binary state discrimination with three depolarized qubits, as well as a binary state discrimination problem with an even smaller system composed of two qutrits. Future work aims to extend the RLNN performance to a larger number of subsystems, as well as to characterize the maximal gap between the optimal local and optimal collective success probability as a function of the number of candidate states and subsystems.

Competing Interests

The authors declare that there are no competing interests.

Acknowledgments

The authors would like to thank Narayanan Rengaswamy and Dhruva Sambrani for helpful discussions. The work of Brandsen and Pfister was supported in part by the National Science Foundation (NSF) under Grant No. 1908730 and 1910571. Stubbs was supported in part by a National Science Foundation Graduate Research Fellowship under Grant No. DGE-1644868. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these sponsors.

References

[1] A. Ferdinand, M. DiMario, and F. Becerra, "Multi-state discrimination below the quantum noise limit at the single-photon level," *npj Quantum Information*, vol. 3, 12 2017. https://doi.org/10.1038/s41534-017-0042-2

[2] H. Krovi, S. Guha, Z. Dutton, and M. P. da Silva, "Optimal measurements for symmetric quantum states with applications to optical communication," *Physical Review A*, vol. 92, Dec 2015. https://doi.org/10.1103/PhysRevA.92.062333

[3] N. Rengaswamy and H. D. Pfister, "Quantum advantage in classical communications via belief-propagation with quantum messages," 2020. https://doi.org/10.1038/s41534-021-00422-1

[4] A. Assalini, N. Dalla Pozza, and G. Pierobon, "Revisiting the Dolinar receiver through multiple-copy state discrimination theory," *Phys. Rev. A*, vol. 84, p. 022342, Aug 2011. https://doi.org/10.1103/PhysRevA.84.022342

[5] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problemy Peredachi Informatsii*, vol. 9, no. 3, pp. 3–11, 1973.

[6] H. Yuen, R. Kennedy, and M. Lax, "Optimum testing of multiple hypotheses in quantum detection theory," *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 125–134, 1975. https://doi.org/10.1109/TIT.1975.1055351

[7] A. H. Kiilerich and K. Mølmer, "Multistate and multihypothesis discrimination with open quantum systems," *Physical Review A*, vol. 97, May 2018. https://doi.org/10.1103/PhysRevA.97.052113

[8] R. Koenig, R. Renner, and C. Schaffner, "The operational meaning of min- and max-entropy," *IEEE Transactions on Information Theory*, vol. 55, p. 4337–4347, Sep 2009. https://doi.org/10.1109/TIT.2009.2025545

[9] R. Bellman, "The theory of dynamic programming," *Bull. Amer. Math. Soc.*, vol. 60, pp. 503–515, 11 1954. https://doi.org/10.1090/S0002-9904-1954-09848-8

[10] S. Brandsen, M. Lian, K. D. Stubbs, N. Rengaswamy, and H. D. Pfister, "Adaptive procedures for discrimination between arbitrary tensor-product quantum states," 2019.

[11] G. Tesauro, "Practical issues in temporal difference learning," *Mach. Learn.*, vol. 8, p. 257–277, May 1992. https://doi.org/10.1007/978-1-4615-3618-5_3

[12] G. J. Gordon, "Stable fitted reinforcement learning," in *Proceedings of the 8th International Conference on Neural Information Processing Systems*, NIPS'95, (Cambridge, MA, USA), p. 1052–1058, MIT Press, 1995.

[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," 2013.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness, M. Bellemare, A. Graves, M. Riedmiller, A. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–33, 02 2015. https://doi.org/10.1038/nature14236

[15] T. Fösel, P. Tighineanu, T. Weiss, and F. Marquardt, "Reinforcement learning with neural networks for quantum feedback," *Phys. Rev. X*, vol. 8, p. 031084, Sep 2018. https://doi.org/10.1103/PhysRevX.8.031084

[16] G. D. Paparo, V. Dunjko, A. Makmal, M. A. Martin-Delgado, and H. J. Briegel, "Quantum speedup for active learning agents," *Phys. Rev. X*, vol. 4, no. 9, 2014. https://doi.org/10.1103/PhysRevX.4.031002

[17] M. Bukov, "Reinforcement learning for autonomous preparation of floquet-engineered states: Inverting the quantum kapitza oscillator," *Phys. Rev. B*, vol. 98, p. 224305, Dec 2018. https://doi.org/10.1103/PhysRevB.98.224305

[18] A. A. Melnikov, H. Poulsen Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, "Active learning machine learns to create new quantum experiments," *Proceedings of the National Academy of Sciences*, vol. 115, no. 6, pp. 1221–1226, 2018. https://doi.org/10.1073/pnas.1714936115

[19] J. Mackeprang, D. Dasari, and J. Wrachtrup, "A reinforcement learning approach for quantum state engineering," *Quantum Mach. Intell. 2, 5*, 2020. https://doi.org/10.1007/s42484-020-00016-8

[20] A. A. Melnikov, P. Sekatski, and N. Sangouard, "Setting up experimental bell tests with reinforcement learning," *Phys. Rev. Lett.*, vol. 125, p. 160401, Oct 2020. https://doi.org/10.1103/PhysRevLett.125.160401

[21] J. Wallnöfer, A. A. Melnikov, W. Dür, and H. J. Briegel, "Machine learning for long-distance quantum communication," *PRX Quantum*, vol. 1, p. 010301, Sep 2020. https://doi.org/10.1103/PRXQuantum.1.010301

[22] R. Sweke, M. S. Kesselring, E. P. L. van Nieuwenburg, and J. Eisert, "Reinforcement learning decoders for fault-tolerant quantum computation," *Machine Learning: Science and Technology*, vol. 2, p. 025005, jan 2021. https://doi.org/10.1088/2632-2153/abc609

[23] F. Schäfer, M. Kloc, C. Bruder, and N. Lörch, "A differentiable programming method for quantum control," *Machine Learning: Science and Technology*, vol. 1, p. 035009, Aug 2020. https://doi.org/10.1088/2632-2153/ab9802

[24] X.-M. Zhang, Z. Wei, R. Asad, X.-C. Yang, and X. Wang, "When does reinforcement learning stand out in quantum control? a comparative study on state preparation," *npj Quantum Inf 5, 85*, 2019. https://doi.org/10.1038/s41534-019-0201-8

[25] R. Sweke, M. S. Kesselring, E. P. L. van Nieuwenburg, and J. Eisert, "Reinforcement learning decoders for fault-tolerant quantum computation," *Machine Learning: Science and Technology*, vol. 2, p. 025005, Jan 2021. https://doi.org/10.1088/2632-2153/abc609

[26] H. Xu, J. Li, L. Liu, Y. Wang, H. Yuan, and X. Wang, "Generalizable control for quantum parameter estimation through reinforcement learning," *npj Quantum Inf 5, 82*, 2019. https://doi.org/10.1038/s41534-019-0198-z

[27] P. Sgroi, G. M. Palma, and M. Paternostro, "Reinforcement learning approach to nonequilibrium quantum thermodynamics," *Phys. Rev. Lett.*, vol. 126, p. 020601, Jan 2021. https://doi.org/10.1103/PhysRevLett.126.020601

[28] P. Palittpongarnpim, P. Wittek, and B. C. Sanders, "Single-shot adaptive measurement for quantum-enhanced metrology," *Quantum Communications and Quantum Imaging XIV*, Sep 2016. https://doi.org/10.1117/12.2237355

[29] A. Hentschel and B. C. Sanders, "Machine learning for precise quantum measurement," *Physical Review Letters*, vol. 104, Feb 2010. https://doi.org/10.1103/PhysRevLett.104.063603

[30] P. Palittapongarnpim, P. Wittek, E. Zahedinejad, S. Vedaie, and B. C. Sanders, "Learning in quantum control: High-dimensional global optimization for noisy quantum dynamics," *Neurocomputing*, vol. 268, p. 116–126, Dec 2017. https://doi.org/10.1016/j.neucom.2016.12.087

[31] P. Palittapongarnpim and B. C. Sanders, "Robustness of quantum-enhanced adaptive phase estimation," *Physical Review A*, vol. 100, Jul 2019. https://doi.org/10.1103/PhysRevA.100.012106

[32] Y. Eldar, A. Megretski, and G. Verghese, "Designing optimal quantum detectors via semidefinite programming," *IEEE Transactions on Information Theory*, vol. 49, p. 1007–1012, Apr 2003. https://doi.org/10.1109/TIT.2003.809510

[33] A. Acín, E. Bagan, M. Baig, L. Masanes, and R. Muñoz Tapia, "Multiple-copy two-state discrimination with individual measurements," *Phys. Rev. A*, vol. 71, p. 032338, 2005. https://doi.org/10.1103/PhysRevA.71.032338

[34] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, "Quantum nonlocality without entanglement," *Physical Review A*, vol. 59, p. 1070–1091, Feb 1999. https://doi.org/10.1103/PhysRevA.59.1070

[35] S. Massar and S. Popescu, "Optimal extraction of information from finite quantum ensembles," *Phys. Rev. Lett.*, vol. 74, pp. 1259–1263, Feb 1995. https://doi.org/10.1142/9789812563071_0023

[36] K. Flatt, S.M. Barnett, and S. Croke, "Multiple-copy state discrimination of noisy qubits", *Phys. Rev. A*, vol. 100, pp. 032122, Sep 2019. https://doi.org/10.1103/PhysRevA.100.032122

[37] B.L. Higgins, A.C. Doherty, S.D. Bartlett, G.J. Pryde, and H.M. Wiseman, "Multiple-copy state discrimination: Thinking globally, acting locally", *Phys. Rev. A*, vol. 81, p. 052314, 2011. https://doi.org/10.1103/PhysRevA.83.052314

[38] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI gym," 2016.

[39] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[40] R. Liaw, E. Liang, R. Nishihara, P. Moritz, J. E. Gonzalez, and I. Stoica, "Tune: A research platform for distributed model selection and training," *arXiv:1807.05118*, 2018.

[41] E. Liang, R. Liaw, P. Moritz, R. Nishihara, R. Fox, K. Goldberg, J. E. Gonzalez, M. I. Jordan, and I. Stoica, "Rllib: Abstractions for distributed reinforcement learning," 2017.

[42] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, "Quantum channels showing superadditivity in classical capacity," *Phys. Rev. A*, vol. 58, pp. 146–158, Jul 1998. https://doi.org/10.1103/PhysRevA.58.146

[43] S. Virmani, M. Sacchi, M. Plenio, and D. Markham, "Optimal local discrimination of two multipartite pure states," *Physics Letters A*, vol. 288, p. 62–68, Sep 2001. https://doi.org/10.1016/S0375-9601(01)00484-4

[44] S. Croke, S. Barnett, and G. Weir, "Optimal sequential measurements for bipartite state discrimination," *Physical Review A, vol 95, no 5*, 2017. https://doi.org/10.1103/PhysRevA.95.052308

[45] G. Weir, C. Hughes, S. M. Barnett, and S. Croke, "Optimal measurement strategies for the trine states with arbitrary prior probabilities," 2018.

[46] M. Ban, "Optimum measurements for discrimination among symmetric quantum states and parameter estimation," *International Journal of Theoretical Physics*, vol. 36, no. 6, pp. 1269–1288, 1997. https://doi.org/10.1007/BF02435921

A Proof of Theorem 1

The "only if" direction of the proof follows immediately from [43], where it was shown that locally adaptive methods are optimal for any binary pure state discrimination problem (including discrimination problems involving entangled states.)

We now demonstrate that even in the case of binary state discrimination, purity is a necessary condition for optimal state discrimination for *any* number of subsystems m. We consider the following set of two qutrit candidate states:

$$\rho_+ \triangleq \Big(\frac{1}{2}\left|0\right\rangle\!\left\langle 0\right| + \frac{1}{2}\left|1\right\rangle\!\left\langle 1\right|\Big) \otimes \Big(\frac{1}{2}\left|0\right\rangle\!\left\langle 0\right| + \frac{1}{2}\left|1\right\rangle\!\left\langle 1\right|\Big)$$

$$\rho_- \triangleq \frac{1}{3}\Big(\sum_{j=0}^{2} \left|j\right\rangle \otimes \left|j\right\rangle\Big)\Big(\sum_{k=0}^{2} \left\langle k\right| \otimes \left\langle k\right|\Big).$$

To our knowledge, this is the smallest dimensional system where binary state discrimination cannot be optimally performed via projective locally adaptive measurements.

The collective success probability is $P_{\text{succ}} = \frac{1}{48}(33+\sqrt{129}) \approx 0.924121$. We now utilize a computer-assisted proof to demonstrate that the gap between the collective and local success probability is strictly positive. To this aim, we introduce a set of closely quantized measurements, find the optimal measurement strategy from the quantized set, and then demonstrate that the success probability is smooth enough that *any* projective local measurement strategy must be strictly worse than the optimal collective measurement.

The set of projective qutrit measurements with parameters ϕ, θ, and ω is given by:

$$\Big\{ |u_1(\phi,\theta,\omega)\rangle\langle u_1(\phi,\theta,\omega)| \, , |u_2(\phi,\theta,\omega)\rangle\langle u_2(\phi,\theta,\omega)| \, , |u_3(\phi,\theta,\omega)\rangle\langle u_3(\phi,\theta,\omega)| \Big\}$$

where

$$\begin{aligned}
|u_1(\phi,\theta,\omega)\rangle &\triangleq \begin{pmatrix} -\cos(\omega)\sin(\phi)+\cos(\phi)\cos(\theta)\sin(\omega) \\ \cos(\phi)\cos(\omega)+\cos(\theta)\sin(\phi)\sin(\omega) \\ -\sin(\theta)\sin(\omega) \end{pmatrix} \\
|u_2(\phi,\theta,\omega)\rangle &\triangleq \begin{pmatrix} \cos(\phi)\cos(\theta)\cos(\omega)+\sin(\phi)\sin(\omega) \\ \cos(\theta)\cos(\omega)\sin(\phi)-\cos(\phi)\sin(\omega) \\ -\cos(\omega)\sin(\theta) \end{pmatrix} \\
|u_3(\phi,\theta,\omega)\rangle &\triangleq \begin{pmatrix} \cos(\phi)\sin(\theta) \\ \sin(\phi)\sin(\theta) \\ \cos(\theta) \end{pmatrix}.
\end{aligned}$$

Any locally adaptive strategy will then consist of a sequence of measurements, such that $\hat{\Pi}(\phi,\theta,\omega)$ is implemented on the first subsystem and, given measurement outcome d_1 from the first measurement, $\hat{\Pi}'(d_1) = \{\Pi'_+(d_1), \Pi'_-(d_1)\}$ is implemented on the second subsystem. Evidently, the optimal measurement on the second subsystem will always be the Helstrom measurement given the updated prior and candidate states.
Thus, finding the optimal locally adaptive strategy is equivalent to finding the optimal first measurement $\hat{\Pi}(\phi,\theta,\omega)$. Using the state set above with starting prior $q = \frac{1}{2}$, the success probability is given by

$$\begin{aligned}
\mathrm{P_s}(\phi,\theta,\omega) = \sum_{j=1}^{3} \mathrm{Tr}\Big[\Big(\Pi_j(\phi,\theta,\omega)\otimes\mathbb{I}\Big)\Big(\frac{1}{2}\rho_+ + \frac{1}{2}\rho_-\Big)\Big]\Bigg(\frac{1}{2} + \frac{1}{2}\Big\|\Pr(\rho=\rho_+|\phi,\theta,\omega,d_1=j)\rho_+(d_1=j) \\
- \Big(1-\Pr(\rho=\rho_+|\phi,\theta,\omega,d_1=j)\Big)\rho_-(d_1=j)\Big\|_1\Bigg),
\end{aligned}$$

where $\rho_\pm(d_1 = j)$ is the post-measurement state for $\rho_\pm$ given that the measurement outcome d_1 is observed to correspond to $\Pi_j(\phi,\theta,\omega)$ on the first subsystem. We now simplify the above expression so that it can be computed directly in Mathematica. First, we note that

$$\begin{aligned}
\rho_+(d_1=j) &\triangleq |u_j(\phi,\theta,\omega)\rangle\langle u_j(\phi,\theta,\omega)| \otimes \Big(\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|\Big) \\
\rho_-(d_1=j) &\triangleq |u_j(\phi,\theta,\omega)\rangle\langle u_j(\phi,\theta,\omega)| \otimes |u_j(\phi,\theta,\omega)\rangle\langle u_j(\phi,\theta,\omega)|
\end{aligned}$$

and likewise

$$\begin{aligned}\mathrm{Tr}\Big[\Big(\Pi_j(\phi,\theta,\omega)\otimes\mathbb{I}\Big)\Big(\frac{1}{2}\rho_+ + \frac{1}{2}\rho_-\Big)\Big] &= \frac{1}{2}\left(\frac{1}{3} + \mathrm{Tr}\Big[\Big(\Pi_j(\phi,\theta,\omega)\otimes\mathbb{I}\Big)\rho_+\Big]\right)\\ &= \frac{1}{2}\left(\frac{1}{3} + \mathrm{Tr}\Big[\Pi_j(\phi,\theta,\omega)\Big(\frac{1}{2}\left|0\right\rangle\!\left\langle 0\right| + \frac{1}{2}\left|1\right\rangle\!\left\langle 1\right|\Big)\Big]\right),\end{aligned}$$

where the first line follows from noting that all measurement outcomes are equally likely on the maximally entangled state. Finally, we can rewrite the success probability as

$$\begin{aligned}\mathrm{P_s}(\phi,\theta,\omega) &= \sum_{j=1}^{3}\frac{1}{2}\left(\frac{1}{3} + \mathrm{Tr}\Big[\Pi_j(\phi,\theta,\omega)\frac{|0\rangle\!\langle 0| + |1\rangle\!\langle 1|}{2}\Big]\right)\left(\frac{1}{2} + \frac{1}{2}\Big\|\Pr(\rho=\rho_+|\phi,\theta,\omega,d_1=j)\right.\\ &\quad\left.\times\frac{|0\rangle\!\langle 0| + |1\rangle\!\langle 1|}{2} - \Pr(\rho=\rho_-|\phi,\theta,\omega,d_1=j)\,|u_j(\phi,\theta,\omega)\rangle\!\langle u_j(\phi,\theta,\omega)|\,\Big\|_1\right)\\ &= \frac{1}{2} + \sum_{j=1}^{9} g_k(\phi,\theta,\omega)|f_k(\phi,\theta,\omega)|,\end{aligned}$$

where in the last step we rewrite the success probability in terms of functions $g_k(\phi,\theta,\omega)$, which correspond to the probability of observing measurement outcome $\lfloor\frac{k}{3}\rfloor$ when measuring the first subsystem, and $f_k(\phi,\theta,\omega)$ which correspond to eigenvalues arising from the trace norm. More specifically, for a given $j\in\{1,2,3\}$, let $\lambda_{j,1},\lambda_{j,2},\lambda_{j,3}$ denote the three eigenvalues of the operator

$$\Pr(\rho=\rho_+|\phi,\theta,\omega,d_1=j)\frac{|0\rangle\!\langle 0| + |1\rangle\!\langle 1|}{2} - \Pr(\rho=\rho_-|\phi,\theta,\omega,d_1=j)\,|u_j(\phi,\theta,\omega)\rangle\!\langle u_j(\phi,\theta,\omega)|\,.$$

The functions $\{f_k(\phi,\theta,\omega)\}_{k=1}^{9}$ are then defined as $f_{3(j-1)+\ell}(\phi,\theta,\omega) := \lambda_{j,\ell}$ where $j,\ell\in\{1,2,3\}$. Likewise,

$$g_k(\phi,\theta,\omega) = \frac{1}{4}\Big(\frac{1}{3} + \mathrm{Tr}\Big[\Pi_{\lfloor\frac{k}{3}\rfloor}(\phi,\theta,\omega)\frac{|0\rangle\!\langle 0| + |1\rangle\!\langle 1|}{2}\Big]\Big)$$

The full expression for $P_s(\phi,\theta,\omega)$ is independent of ϕ and therefore can be denoted as $P_s(\theta,\omega)$. While $P_s(\phi,\theta,\omega)$ can be computed directly via Mathematica, the closed form expression is extremely lengthy. For completeness, we provide the code used to generate and plot $P_s(\phi,\theta,\omega)$ at https://github.com/SarahBrandsen/RLNN-QSD. In Figure 17, we plot $P_s(\theta,\omega)$ to show the continuity of the function and demonstrate graphically that $P_s(\theta,\omega)$ is upper bounded by 0.87.

To prove this, we quantize θ and ω into 10,000 discrete values such that $\theta,\omega\in\{\frac{2\pi j}{10000}\}|_{j=1}^{10000}$, and find that the best success probability achieved is

$$\frac{1}{384}\left(240 + \sqrt{1054 - 42\sqrt{5}} + \sqrt{2302 + 630\sqrt{5}}\right) \approx 0.864325$$

Then, we demonstrate that any error due to quantization is sufficiently small (i.e., the gap between the local and collective measurement strategy cannot be due to quantization.) Consider a fixed θ and ω and for any θ' and ω' satisfying $|\theta'-\theta|\le\epsilon$ and $|\omega-\omega'|\le\epsilon$, then

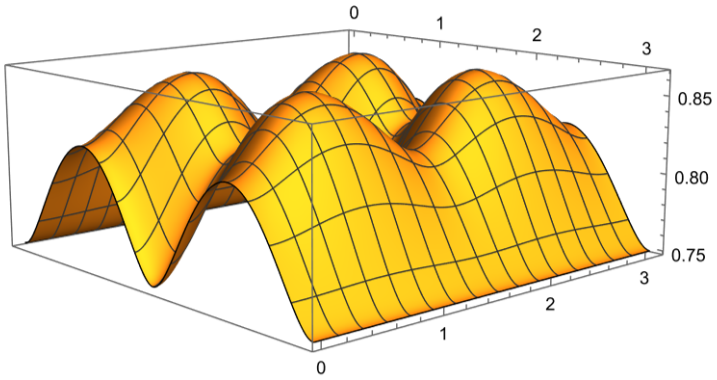

Figure 17: $P_{\text{succ}}(\theta,\omega)$ plotted over the range $0 \le \theta \le \pi$ and $0 \le \omega \le \pi$.

$$
\begin{aligned}
&\Big|P_s(\theta,\omega) - P_s(\theta',\omega')\Big| \\
&= \Big|\sum_j g_j(\theta,\omega)|f_j(\theta,\omega)| - g_j(\theta',\omega')|f_j(\theta',\omega')|\Big| \\
&= \Big|\sum_j \Big(g_j(\theta,\omega) - g_j(\theta',\omega') + g_j(\theta',\omega')\Big)|f_j(\theta,\omega)| - g_j(\theta',\omega')|f_j(\theta',\omega')|\Big| \\
&\le \sum_j |g_j(\theta,\omega) - g_j(\theta',\omega')||f_j(\theta,\omega)| + |g_j(\theta',\omega')||f_j(\theta,\omega) - f_j(\theta',\omega')| \\
&\le \Big(\max_j |g_j(\theta,\omega) - g_j(\theta',\omega')|\Big)\sum_j |f_j(\theta,\omega)| \\
&\qquad + \Big(\max_j |f_j(\theta,\omega) - f_j(\theta',\omega')|\Big)\sum_j |g_j(\theta',\omega')|
\end{aligned}
$$

By mean-value theorem, there exists a point $(\theta_{r_j},\omega_{r_j})$ along the line connecting (θ,ω) and (θ',ω') so that

$$
\begin{aligned}
|g_j(\theta,\omega) - g_j(\theta',\omega')| &= |\nabla g_j(\theta_{r_j},\omega_{r_j})\cdot(\theta-\theta',\omega-\omega')| \\
&\le \epsilon\left(\Big|\frac{\partial g_j}{\partial\theta}(\theta_{r_j},\omega_{r_j})\Big| + \Big|\frac{\partial g_j}{\partial\omega}(\theta_{r_j},\omega_{r_j})\Big|\right)
\end{aligned}
$$

where the last line follows from Holder's inequality. Similarly, there exists a point $(\theta_{r'_j},\omega_{r'_j})$ so that

$$
|f_j(\theta,\omega) - f_j(\theta',\omega')| \le \epsilon\left(\Big|\frac{\partial f_j}{\partial\theta}(\theta_{r'_j},\omega_{r'_j})\Big| + \Big|\frac{\partial f_j}{\partial\omega}(\theta_{r'_j},\omega_{r'_j})\Big|\right)
$$

Using the explicit formula for $g_j(\theta,\omega)$ and $f_j(\theta,\omega)$ we computed with Mathematica, it can be checked that

$$
\max_{\theta,\omega}\sum_j |f_j(\theta,\omega)| \le 9 \qquad \max_{j,\theta,\omega}\left(\Big|\frac{\partial f_j}{\partial\theta}(\theta,\omega)\Big| + \Big|\frac{\partial f_j}{\partial\omega}(\theta,\omega)\Big|\right) \le 12 \tag{1}
$$

$$
\max_{\theta,\omega}\sum_j |g_j(\theta,\omega)| \le 3 \qquad \max_{j,\theta,\omega}\left(\Big|\frac{\partial g_j}{\partial\theta}(\theta,\omega)\Big| + \Big|\frac{\partial g_j}{\partial\omega}(\theta,\omega)\Big|\right) \le \frac{1}{4} \tag{2}
$$

Hence, we finally conclude that [1]

$$\max_{\theta',\omega'} \left| P_s(\theta,\omega) - P_s(\theta',\omega') \right| \leq \frac{1}{4}\epsilon \times 9 + 12\epsilon \times 3 = \frac{153}{4}\epsilon$$

Setting $\epsilon = \frac{\Pi}{10000}$ then gives

$$\max_{\theta',\omega'} \left| P_s(\theta,\omega) - P_s(\theta',\omega') \right| \leq 0.015.$$

It follows that the observed gap between the locally adaptive and collective measurement scheme (with an approximate magnitude of 0.06) persists even after accounting for quantisation.
We likewise demonstrate that for the candidate states

$$\rho_+ \triangleq \begin{pmatrix} 0.85009903 & 0.1343714 \\ 0.1343714 & 0.14990097 \end{pmatrix}^{\otimes 3}$$
$$\rho_- \triangleq \begin{pmatrix} 0.58134943 & 0.36607003 \\ 0.36607003 & 0.41865057 \end{pmatrix}^{\otimes 3}$$

with $q = \frac{1}{2}$, then $P_{\text{coll}} - P_{\text{local}} \geq 0.011$ even when the quantization for the qubit projective measurements is set to $10,000$. (It follows from the proof of Theorem 2, discussed in Appendix B, that the largest error due to quantization in this case would be upper bounded by $3 \times 2^3 \times \frac{\pi}{10000}$ and thus is negligible.)

Finally, we utilize the result found in [44] to demonstrate that even for pure state state discrimination, if $n \geq 3$, then for any m there exists a candidate state with a gap. We restate the result found in [44] and provide a simplified proof which draws on their later work in [45]. Consider as an example the case where $n = 2$, $m = 3$ and where each candidate state set consists of the trine ensemble, defined to be symmetric with

$$\rho_j \triangleq \left(U^j |0\rangle\langle 0| (U^j)^\dagger \right)^{\otimes 2} = (U \otimes U)^j |00\rangle \langle 00| \left((U \otimes U)^j \right)^\dagger,$$

where $U \triangleq \begin{pmatrix} \cos(\frac{2\pi}{3}) & -\sin(\frac{2\pi}{3}) \\ \sin(\frac{2\pi}{3}) & \cos(\frac{2\pi}{3}) \end{pmatrix}$ and $\boldsymbol{q} = [1/3, 1/3, 1/3]$. Since $(U \otimes U)^m = \mathbb{I}$, and the starting prior is balanced, the PGM is optimal [46], with a corresponding collective success probability of $P_{\text{coll}} = \frac{1}{6}(3 + 2\sqrt{2}) \approx 0.971$.

We now demonstrate that the optimal local strategy is to measure the first subsystem with an "anti-trine" measurement, defined as $\hat{\Pi}_{\text{AT}} = \{\frac{2}{3} U^{\frac{1}{2}} \rho_j (U^{\frac{1}{2}})^\dagger\}|_{j=1}^3$, such that each measurement outcome is orthogonal to one of the candidate states. After obtaining the measurement outcome for the first subsystem, the updated prior is a permutation of $\mathbf{q} = [\frac{1}{2}, \frac{1}{2}, 0]$, and the second subsystem is measured according to the optimal measurement for the remaining two candidate states.
The most general local approach is to implement measurement $\hat{\Pi}_1 = \{\Pi_{1,j}\}|_{j=1}^m$ on the first subsystem. We may label the result of the first subsystem out_1. Then the second and

[1] While the current proof shows an analytical bound of $\frac{153}{4}$, computer-assisted proofs may potentially show tighter bounds.

last measurement can be chosen as $\hat{\Pi}_2(\text{out}_1) = \{\Pi_{2,j}(\text{out}_1)\}|_{j=1}^3$ and is in general allowed to depend on the outcome out_1. It is conventional to label the elements of the second measurement such that state ρ is decoded as ρ_j if measurement element j is obtained in the final round. Then

$$\begin{aligned}\mathrm{P}_{\text{succ}}\Big(\hat{\Pi}_1, \{\hat{\Pi}_2(d_1)\}\Big) &= \sum_{d_1=1}^{m} \Pr[\text{out}_1 = \Pi_{d_1}^{(1)}]\mathrm{P}_{\text{succ}}(\hat{\Pi}^{(2)}(d_1)\Big|\text{out}_1 = \Pi_{d_1}^{(1)}) \\ &\leq \max_{\hat{\Pi}_1, \{\hat{\Pi}_2(d_1)\}} \mathrm{P}_{\text{succ}}\Big(\hat{\Pi}^{(2)}(d_1)\Big|\text{out}_1 = \Pi_{d_1}^{(1)}\Big)\end{aligned}$$

Thus, the second line presents an upper bound on the success probability of *any* locally adaptive strategy, as it gives the success probability of the best possible measurement sequence.

It follows that a sufficient condition for the optimality of the anti-trine based method is

$$\mathrm{P}_{\text{succ}}\Big(\hat{\Pi}_{\text{AT}}, \{\hat{\Pi}_2^*(d_1)\}\Big) \geq \max_{\hat{\Pi}_1, \{\hat{\Pi}_2(d_1)\}} \mathrm{P}_{\text{succ}}\Big(\hat{\Pi}^{(2)}(d_1)\Big|\text{out}_1 = \Pi_{d_1}^{(1)}\Big)$$

By symmetry, we know the expected success probability for the anti-trine is equivalent regardless of which outcome is obtained, and can immediately compute $\mathrm{P}_{\text{succ}}(\hat{\Pi}_{\text{AT}}) = 0.933$. From [45], when the outcome for the first subsystem is $\Pi(\theta) = \begin{pmatrix} \sin^2(\theta) & \cos(\theta)\sin(\theta) \\ \cos(\theta)\sin(\theta) & \cos^2(\theta) \end{pmatrix}$, the expected success probability given optimal choice of subsequent measurement is given by

$$\begin{aligned}\max_{\{\hat{\Pi}^{(2)}(d_1)\}} \left(\mathrm{P}_{\text{succ}}\Big(\hat{\Pi}^{(2)}(\theta)\Big|\text{out}_1 = \Pi(\theta)\Big)\right) &\leq \max_{\theta} \Big(\frac{1}{3} - \frac{1}{12}\cos(2\theta) - 0.288675\cos(\theta)\sin(\theta) \\ &\quad + \frac{1}{2}\sqrt{\frac{5}{12} - \frac{1}{6}\cos(2\theta) - 0.57735\cos(\theta)\sin(\theta)},\ 0.85\Big) \\ &= 0.933 \\ &= \mathrm{P}_{\text{succ}}\Big(\hat{\Pi}_{\text{AT}}, \{\hat{\Pi}_2^*(d_1)\}\Big).\end{aligned}$$

This proves that the best locally optimal strategy is the anti-trine with a success probability of $P_{\text{loc}}(\{\rho_j\}, \mathbf{q}) = 0.933$. Clearly, $\mathrm{P}_{\text{loc}}(\{\rho_j\}, \mathbf{q}) < \mathrm{P}_{\text{coll}}(\{\rho_j\}, \mathbf{q})$, from which it follows that a necessary condition for optimal locally adapative state discrimination of pure states is that $m = 2$.

B Proof of Theorem 2

Any adaptive protocol consists of a series of measurements, $\{\Pi_1, \Pi_2(d_1), ...\Pi_n(\mathbf{d}_{[n-1]})\}$, where all measurements after the first depend on previous measurement results. Then any individual measurement sequence can be written as a tensor product

$$\Pi_{\mathbf{d}_{[n]}} = \bigotimes_{k=1}^{n} \Pi_k(\mathbf{d}_{[k-1]}).$$

Let $\mathcal{S}_j$ be the set of all measurement sequences $\mathbf{d}_{[n]}$ such that the post-measurement decoding is $\hat{\rho} = \rho_j$. Then, we can define $\Pi'_j = \sum_{\mathbf{d}_{[n]} \in \mathcal{S}_j} \Pi_{\mathbf{d}_{[n]}}$ as the measurement element which leads to decoding $\hat{\rho} = \rho_j$, and the difference between the two success probabilities can be bounded with

$$\mathrm{P}_{\mathrm{succ}}\Big(\{\rho_j\}\Big) - \mathrm{P}_{\mathrm{succ}}\Big(\{\tilde{\rho}_j(\theta)\}\Big) = \sum q_j \mathrm{Tr}[\Pi'_j(\rho_j - \tilde{\rho}_j(\theta))]$$
$$\leq \max_j \left(\mathrm{Tr}[\Pi'_j(\rho_j - \tilde{\rho}_j(\theta))] \right).$$

Then, by Hölder's inequality, we see that

$$\left| \mathrm{Tr}\Big[\Pi'_j \bigotimes_k \rho_j^{(k)}\Big] - \mathrm{Tr}\Big[\Pi'_j \bigotimes_k \tilde{\rho}_j^{(k)}(\theta)\Big] \right| \leq \left\|\Pi'_j\right\|_\infty \left\| \bigotimes_k \rho_j^{(k)} - \bigotimes_k \tilde{\rho}_j^{(k)} \right\|_1$$
$$\leq \left\| \bigotimes_k \rho_j^{(k)} - \bigotimes_k \tilde{\rho}_j^{(k)} \right\|_1,$$

where the last inequality follows from noting that $\Pi'_j \leq \mathbb{I}$. Then we find that

$$\begin{aligned}
\left\| \bigotimes_{k=1}^n \rho_k - \bigotimes_{k=1}^n \tilde{\rho}_k \right\|_1 &= \left\| \bigotimes_{k=1}^n \rho_k - \rho_1 \otimes \bigotimes_{k=2}^n \tilde{\rho}_k + \rho_1 \otimes \bigotimes_{k=2}^n \tilde{\rho}_k - \bigotimes_{k=1}^n \tilde{\rho}_k \right\|_1 \\
&\leq \left\| \bigotimes_{k=1}^n \rho_k - \rho_1 \otimes \bigotimes_{k=2}^n \tilde{\rho}_k \right\| + \left\| \rho_1 \otimes \bigotimes_{k=2}^n \tilde{\rho}_k - \bigotimes_{k=1}^n \tilde{\rho}_k \right\|_1 \\
&\leq \sum_{\ell=0}^{n-1} \left\| \bigotimes_{k=1}^{\ell+1} \rho_k \otimes \bigotimes_{k=\ell+2}^n \tilde{\rho}_k - \bigotimes_{k=1}^{\ell} \rho_k \otimes \bigotimes_{k=\ell+1}^n \tilde{\rho}_k \right\|_1 \\
&= \sum_{\ell=0}^{n-1} \left\| \bigotimes_{k=1}^{\ell} \rho_k \otimes \Big(\rho_{\ell+1} - \tilde{\rho}_{\ell+1}\Big) \otimes \bigotimes_{k=\ell+2}^n \tilde{\rho}_k \right\|_1.
\end{aligned}$$

Let us consider the ℓ^{th} term. We denote the eigenvalues of ρ_k as $\{\lambda_1^{(k)}, \lambda_2^{(k)}\}$ and likewise the eigenvalues of $\tilde{\rho}_k$ as $\{\tilde{\lambda}_1^{(k)}, \tilde{\lambda}_2^{(k)}\}$. Finally, we denote the eigenvalues of $\rho_{\ell+1} - \tilde{\rho}_{\ell+1}(\theta)$ as $\{\sigma_1(\theta), \sigma_2(\theta)\}$. Then, we have

$$\begin{aligned}
\left\| \bigotimes_{k=1}^{\ell} \rho_k \otimes \Big(\rho_{\ell+1} - \tilde{\rho}_{\ell+1}\Big) \otimes \bigotimes_{k=\ell+2}^n \tilde{\rho}_k \right\|_1 &= \prod_{k=1}^{\ell} \Big(|\lambda_1^{(k)}| + |\lambda_2^{(k)}|\Big) \times \Big(|\sigma_1(\theta)| + |\sigma_2(\theta)|\Big) \\
&\quad \times \prod_{k=\ell+2}^{n} \Big(|\tilde{\lambda}_1^{(k)}| + |\tilde{\lambda}_2^{(k)}|\Big) \\
&= |\sigma_1(\theta)| + |\sigma_2(\theta)|
\end{aligned}$$

since $|\lambda_1^{(k)}| + |\lambda_2^{(k)}| = 1$ for all k. Now, we bound the eigenvalues of $\rho_\ell - \tilde{\rho}_\ell$ with

$$\begin{aligned}
\left\|\rho_\ell - \tilde{\rho}_\ell\right\|_1 &= \left\|\rho_\ell - U(\theta)\rho_\ell U(\theta)^\dagger\right\|_1 \\
&= \left\| \Big[\rho_\ell, U(\theta)\Big] U(\theta)^\dagger \right\|_1 \\
&= \left\| \Big[\rho_\ell - \mathbb{I}, U(\theta) - \mathbb{I}\Big] \right\|_1,
\end{aligned}$$

where we have used that identity commutes with all operators and the unitary invariance of the trace norm. Since $0 \leq \rho_\ell \leq \mathbb{I}$ all ℓ we have that

$$\left\|\rho_\ell - \tilde{\rho}_\ell\right\|_1 \leq \left\|U(\theta) - \mathbb{I}\right\|_1.$$

Since $U(\theta)$ is a rotation, we can easily calculate its eigenvalues to $e^{\pm i\theta}$. Hence, we conclude that

$$\begin{aligned}\left\|\rho_\ell - \tilde{\rho}_\ell\right\|_1 &\leq \left|e^{i\theta} - 1\right| + \left|e^{-i\theta} - 1\right| \\ &= 4\sin(|\theta|/2).\end{aligned}$$

From this, we have

$$\left\|\bigotimes_{k=1}^{n}\rho_k - \bigotimes_{k=1}^{n}\tilde{\rho}_k\right\|_1 \leq \sum_{\ell=0}^{n-1} 4\sin(|\theta|/2) = 4n\sin(|\theta|/2)$$

and the statement follows.