LOCC protocols with bounded width per round optimize convex functions

Debbie Leung[*]   Andreas Winter[†]   Nengkun Yu[*,‡]

20 April 2019

Abstract

We start with the task of discriminating finitely many multipartite quantum states using LOCC protocols, with the goal to optimize the probability of correctly identifying the state. We provide two different methods to show that finitely many measurement outcomes in every step are sufficient for approaching the optimal probability of discrimination. In the first method, each measurement of an optimal LOCC protocol, applied to a d_{loc}-dim local system, is replaced by one with at most $2d_{\text{loc}}^2$ outcomes, without changing the probability of success. In the second method, we decompose *any* LOCC protocol into a convex combination of a number of "slim protocols" in which each measurement applied to a d_{loc}-dim local system has at most d_{loc}^2 outcomes. To maximize any convex functions in LOCC (including the probability of state discrimination or fidelity of state transformation), an optimal protocol can be replaced by the best slim protocol in the convex decomposition without using shared randomness. For either method, the bound on the number of outcomes per measurement is independent of the global dimension, the number of parties, the depth of the protocol, how deep the measurement is located, and applies to LOCC protocols with infinite rounds, and the "measurement compression" can be done "top-down" – independent of later operations in the LOCC protocol. The second method can be generalized to implement LOCC instruments with finitely many outcomes: if the instrument has n coarse-grained final measurement outcomes, global input dimension D_0 and global output dimension D_i for $i = 1, \cdots, n$ conditioned on the i-th outcome, then one can obtain the instrument as a convex combination of no more than $R = \sum_{i=1}^{n} D_0^2 D_i^2 - D_0^2 + 1$ slim protocols; in other words, $\log_2 R$ bits of shared randomess suffice.

[*]Institute for Quantum Computing and Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada.

[†]ICREA & Física Teórica: Informació i Fenómens Quàntics, Departament de Física, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain

[‡]Centre for Quantum Software and Information, Faculty of Engineering and Information Technology, University of Technology Sydney NSW 2007, Australia.

1 Introduction

For a multi-partite quantum system, the class of operations that can be implemented by composing local operations on each individual part and classical communication between the parts is shorthanded LOCC. This class originates from the seminal work by Peres and Wootters [1] and its importance has been manifest in many subsequent results, such as [2, 3]. One motivation for the LOCC class is the operational difficulty of long range quantum communication. From a more fundamental perspective, LOCC is precisely the class of operations that can be implemented *without* shared entanglement; therefore LOCC provides a natural framework to study quantum nonlocality and entanglement. Understanding the power and the limitation of LOCC operations is one of the main goals of quantum information theory. In particular, we say that an information processing task exhibits nonlocality when it can be accomplished using global operations but not by LOCC operations. While the class of LOCC operations is well motivated, it does not have a succinct mathematical characterization, and the complexity grows rapidly with the number of rounds of communication.

In this paper, we first consider the quantum state discrimination problem, in which a list of quantum states is fixed in advance. A referee chooses a state from the list, prepares a copy, and distributes it to the discriminating party (or parties), whose goal is to identify which state has been prepared by the referee. In some situations the prepared state can be identified without error. Otherwise, one can relax the problem by assuming that the referee picks a state from the list according to some pre-determined distribution known to the parties, and their goal is to maximize the *success probability* (i.e., the probability of correctly identifying the state).

The quantum state discrimination problem provides a fruitful line of studies in our understanding of LOCC and nonlocality. If restricting the players to LOCC operations strictly decreases their probability of success, the problem exhibits nonlocality. A partial list of references on this problem can be found in [4]-[33]. Because there is no succinct description of all possible LOCC discrimination strategies, one widely used approach is to use a larger set of operations to study the limitation on the distinguishability power of LOCC, for instance, separable operations (SEP) or PPT-preserving operations [8, 9, 10, 23, 24, 27, 31, 32].

More recent studies have found useful structural and topological properties of LOCC [33]-[37]. In [34, 35, 36], the set of LOCC operations is shown to be not closed. Explicit entanglement transformation tasks are given in [34, 35, 36] that are *provably* not accomplished by any finite round LOCC protocol but that can be approximated with arbitrary precision when the number of communication rounds increases. Consequently, we cannot assume that an LOCC protocol has a finite number of communication rounds, called the *depth* of the protocol. Even after restricting to finite depth protocols for the task, it is not clear apriori whether more and more outcomes in some intermediate measurement in the protocol can approximate the ideal task better and better. The total width of the protocol refers to the total number of measurement outcomes in the protocol. A related concept is the width per measurement, which is the number of outcomes for each measurement. An LOCC protocol can be represented by a decision tree where vertices represent operations, and edges represent measurement outcomes, thus the names depth and total width of a protocol. (Note that the width per measurement is the degree of the root vertex, or the degree minus 1 for other vertices.) Reference [36] shows that the set of all LOCC protocols with a constant number of rounds of communication is *compact*, and provides an upper bound on the number of measurement outcomes.

In this paper, we focus on the width of LOCC protocols. In the more specialized context of multipartite quantum state discrimination, we show that LOCC protocols with finite width per measurement are sufficient to achieve the optimal probability of success in quantum state discrimination under LOCC with two different methods. The second method extends to optimizing any convex function, including the probability of state discrimination, and the fidelity of state transformation. Both results apply to LOCC (as defined in Section 2.2 of [36]). Informally, this class includes infinite round LOCC protocols that can be approximated better and better by adding more and more rounds of communications (without changing the earlier steps). Each protocol in this class can be represented by a tree that can be infinite. Each measurement is replaced by one with few outcomes in a "top-down" manner – starting from the root (where the protocol begins), we replace each measurement as we move down the tree (as the protocol progresses) in a way *independent* of how deep the protocol will be executed. The original task can be approximated better and better by going deeper in the resulting single finite-width infinite-depth protocol.

Our first method converts every measurement in the protocol (possibly with infinitely many outcomes) into one with no more than $2d_{\mathrm{loc}}^2$ outcomes where d_{loc} is the dimension of the local system being measured. If the protocol is finite with ℓ rounds of communication and d is the largest of the local dimensions, the total width of the protocol is upper bounded by $2^\ell d^{2\ell}$. Our second method converts every measurement into one with no more than d_{loc}^2 outcomes. If the protocol is finite with ℓ rounds of communication and d is the largest of the local dimensions, the total width of the protocol is upper bounded by $d^{2\ell}$.

Both methods are constructive, and rely on Caratheodory's theorem. They are simpler than the compression given by [36] for finite LOCC protocols, and our bounds are tighter (independent of the global dimension and the number of parties, independent of how deep the protocol has run, and has lower degree in the dimension). Most importantly, our compression is top-down.

The second method also implements any LOCC instrument with the aforementioned width per measurement by using additional shared randomness. If the protocol has n coarse-grained final measurement outcomes, global input dimension D_0 and global output dimension D_i for $i = 1, \cdots, n$ conditioned on the i-th outcome, then $\log_2 R$ bits of shared randomess suffice where $R = \sum_{i=1}^n D_0^2 D_i^2 - D_0^2 + 1$.

Towards the final stages of preparing this manuscript, we learnt of related results by Cohen [38], who shows that any LOCC quantum operation $\mathcal{E}$ with potentially unbounded width can be converted to one with finite width per round. The number of measurement outcomes per round is upper bounded by $\min(\kappa^2, \kappa^2 + d_{\mathrm{loc}}^2 - \chi)$ where κ is the global Kraus rank of $\mathcal{E}$ and $\chi = \dim\left(\mathrm{span}\{K_i^\dagger K_j\}_{i,j=1}^\kappa\right)$ where K_i's are the Kraus operators of $\mathcal{E}$. Cohen's method preserves the quantum operations. In comparison, our methods for optimizing concave functions need not preserve the quantum operations, but may have a tighter bound on the width per measurement in some regime. (For example, to discriminate many states shared by a large number of parties, each holding a small dimensional system, the Kraus rank scales as the number of states which is much larger than the local dimension.) Our second method can be extended to preserve the quantum operations by using finite amount of shared randomness. Our correspondence with Cohen had inspired improvements in aspects of our second method (including a discussion on the shared randomness, and the application of state transformation). It remains unclear how closely the approaches in these two papers are related, and whether the techniques can be combined to obtain better results.

In Section 2 we cover the mathematical background and define notations and concepts required for the discussion. The main results are presented in Section 3.

2 Preliminaries

The term *Hilbert space* here refers to any *finite* dimensional semidefinite inner product space over the complex numbers. Let $\mathcal{X}$ be an arbitrary Hilbert space. A *pure* quantum state of $\mathcal{X}$ is a *normalized* vector $|\Psi\rangle \in \mathcal{X}$. A quantum mechanical system is associated with a *Hilbert space* and we refer to both the system and the space with the same notation. A composite system is associated with the tensor product of the *Hilbert spaces* associated with the parts.

The space of linear operators mapping $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $\mathcal{L}(\mathcal{X}, \mathcal{Y})$, while $\mathcal{L}(\mathcal{X})$ is the shorthand for $\mathcal{L}(\mathcal{X}, \mathcal{X})$. We use $I_{\mathcal{X}}$ to denote the identity operator on $\mathcal{X}$, and often omit the system label $\mathcal{X}$. The adjoint (or Hermitian transpose) of $A \in \mathcal{L}(\mathcal{X}, \mathcal{X})$ is denoted by $A^\dagger$. The notation $A \geq 0$ means that A is positive semidefinite, and more generally $A \geq B$ means that $A - B$ is positive semidefinite. The positive square root of $A^\dagger A$ is denoted by $|A| = \sqrt{A^\dagger A}$.

A (general) quantum state is specified by its density operator $\rho \in \mathcal{L}(\mathcal{X})$, which is a positive semidefinite operator with trace one. The density operator of a pure state $|\psi\rangle$ is simply the projector $\psi := |\psi\rangle\langle\psi|$.

A quantum measurement $\mathcal{M}$ with input system $\mathcal{X}$ and output system $\mathcal{Y}$ is specified by a POVM $(A_1^\dagger A_1, A_2^\dagger A_2, \cdots)$ where each $A_i \in \mathcal{L}(\mathcal{X}, \mathcal{Y})$ and $\sum_i A_i^\dagger A_i = I$. If the initial state being measured is ρ,

$$\mathcal{M}(\rho) = \sum_i A_i \rho A_i^\dagger \otimes |i\rangle\langle i|$$

where i is the measurement outcome, and $A_i \rho A_i^\dagger$ is the corresponding unnormalized postmeasurement quantum state whose norm $\mathrm{tr} A_i \rho A_i^\dagger = A_i^\dagger A_i \rho$ gives the probability of obtaining outcome i. More generally, each A_i can take the input system $\mathcal{X}$ to an output system $\mathcal{Y}_i$, where the $\mathcal{Y}_i$'s may not have the same dimension.

The most general quantum operation $\mathcal{E}$ with input system $\mathcal{X}$ and output system $\mathcal{Y}$ acts as $\mathcal{E}(\rho) = \sum_i A_i \rho A_i^\dagger$, where $\sum_i A_i^\dagger A_i = I$. An instrument with input system $\mathcal{X}$ acts as $\mathcal{I}(\rho) = \sum_i \mathcal{I}_i(\rho) \otimes |i\rangle\langle i|$ where each $\mathcal{I}_i$ is a completely positive map, and $\sum_i \mathcal{I}_i$ is trace preserving. A measurement is a fine-grained instrument in which all CP maps has Kraus rank 1.

Consider an ensemble of quantum states

$$S = \{p_1\rho_1, \cdots, p_n\rho_n\} \subset \mathcal{L}(\mathcal{X})$$

where ρ_k are normalized states and $p_k \geq 0$, $\sum_k p_k \leq 1$. Then $\sum_k p_k$ is called the probability of the ensemble S. If $\sum_k p_k = 1$, the ensemble is called *normalized*.

Throughout this paper, we focus on multipartite quantum systems of the form,

$$\mathcal{X} = \mathcal{X}_1 \otimes \mathcal{X}_2 \otimes \cdots \otimes \mathcal{X}_m.$$

LOCC, or local operations and classical communication, on this system $\mathcal{X}$, is the set of operations such that each party is restricted to performing quantum operations on their individual local systems and they may communicate classical information (w.l.o.g., measurement outcomes) to the

other parties. Any LOCC operation can be decomposed into rounds; at each round, one party applies a quantum operation on his/her local system and broadcasts a classical message to all other parties. An LOCC protocol can have infinitely many rounds of communication. See [36] for detail.

For any LOCC protocol $\mathcal{P}$ on $\mathcal{X}$ (potentially infinitely wide and with infinitely many rounds, but one that can be approximated by adding rounds and can be represented by a tree), denote its ℓ-round prefix by $\mathcal{P}_\ell$. For any ensemble $S = \{p_1\rho_1, \cdots, p_n\rho_n\} \subset \mathcal{X}$, the probability of successful discrimination by $\mathcal{P}$ can be defined as follows,

$$P(\mathcal{P}, S) = \lim_{\ell \to \infty} P(\mathcal{P}_\ell, S),$$

where $P(\mathcal{P}_\ell, S)$ denotes the probability of successful discrimination of S by $\mathcal{P}_\ell$. As $\mathcal{P}_\ell$ can potentially have infinite width, $P(\mathcal{P}_\ell, S)$ is similarly defined as a limit.

In our second method, we use the following notation and terminology derived from [36]. Any protocol in LOCC can be represented as a possibly infinite tree. The protocol starts at the root and moves through the tree along edges, always further away from the root. Each vertex v is associated with an instrument applied to a local system $\mathcal{H}_{\text{loc}}$ held by one party. We can write this instrument as

$$\mathcal{L}_v(\rho) = \sum_{w:\text{child of } v} \mathcal{L}_{(w,v)}(\rho) \otimes |(w,v)\rangle\langle(w,v)|,$$

where each outgoing edge (w, v) is associated with a CP map $\mathcal{L}_{(w,v)}$ acting on $\mathcal{H}_{\text{loc}}$, such that $\sum_{w:\text{child of } v} \mathcal{L}_{(w,v)}$ is trace-preserving. Each vertex v at depth ℓ can be reached by a unique path from the root r, $(r, v_1), (v_1, v_2), \cdots, (v_{\ell-1}, v)$ and the vertex is associated with a "cumulative" CP map

$$\mathcal{N}_v := \mathcal{L}_{(v, v_{\ell-1})} \circ \cdots \circ \mathcal{L}_{(v_2, v_1)} \circ \mathcal{L}_{(v_1, r)}.$$

The LOCC protocol implements an LOCC instrument $\mathcal{L}$ which can be specified as follows. Consider any function $f: L \to O$ from the set of leaves L of the tree, to a set of outcomes O. The instrument $\mathcal{L}$ is implemented by running the protocol from the root until arriving at a leaf v and then outputting $f(v)$. For each $o \in O$, let $\mathcal{I}_o = \sum_{v \in f^{-1}(o)} \mathcal{N}_v$. The instrument implemented by the LOCC protocol is given by $\mathcal{I}(\rho) = \sum_o \mathcal{I}_o(\rho) \otimes |o\rangle\langle o|$. For a finite tree, $\sum_o \mathcal{I}_o$ is trace preserving. For an infinite tree we have to make an extra assumption that for every input the protocol terminates with probability 1.

We can fine-grain an LOCC protocol by breaking up CP maps associated with edges into Kraus-rank-1 CP maps, enlarging the tree, and modifying the coarse-graining function f accordingly, without changing the instrument implemented by the protocol. We call such a protocol *fine-grained*.

3 Main Result

3.1 The first method and resulting bounds

Theorem 1. *Suppose an ensemble of multipartite quantum states $S = \{p_1\rho_1, \cdots, p_n\rho_n\} \subset \mathcal{L}(\otimes_{j=1}^m \mathcal{X}_j)$ with $p_k \geq 0$, $\sum_k p_k = 1$ can be distinguished by some LOCC protocol $\mathcal{P}$ with success probability t. Then*

5

there exists an LOCC protocol $\mathcal{P}'$ achieving the same success probability t but in which each measurement requires at most $2d_{\text{loc}}^2$ outcomes, where d_{loc} is the dimension of the local system measured. If $\mathcal{P}$ has finitely many rounds of communication ℓ, the total width can be bounded by $2^\ell d^{2\ell}$, where d is the maximum local dimension.

We first discuss informally the intuition behind the constructive proof. We obtain the bound by recursively "compress" an arbitrary measurement in the protocol while preserving the success probability, depth of the protocol, and the induced post-measurement ensembles. The compression for a measurement is done in several steps:

1. Show that the measurement can be performed in two stages (as a composition of two measurements).

2. Show that the first stage measurement can be modified to "equalize" the success probability on the induced post-measurement ensemble for each measurement outcome. This step preserves the success probability of the protocol.

3. A convexity argument shows that all but a finite number of measurement outcomes can be dropped for the first stage while preserving the probability of success.

4. The modifications in steps 2-3 are compatible with the two stage implementation of the original measurement. So, the second stage measurement is applied for each of the finitely many outcomes in stage 1. This preserves the depth and success probability of the protocol. Furthermore, all the subsequent steps in the original protocol are unaffected.

The following lemma will be needed for steps 1, 2, and 4 above.

Lemma 2. *For any pair of matrices X, Y of the same width, there exists a matrix C of the same size as X and a matrix D of the same size as Y such that*

$$C\sqrt{X^\dagger X + Y^\dagger Y} = X,$$
$$D\sqrt{X^\dagger X + Y^\dagger Y} = Y,$$
$$C^\dagger C + D^\dagger D = I.$$

Proof of Lemma 2. If $X^\dagger X + Y^\dagger Y$ is nonsingular, then the choices

$$C = X(X^\dagger X + Y^\dagger Y)^{-1/2},$$
$$D = Y(X^\dagger X + Y^\dagger Y)^{-1/2}.$$

imply $C^\dagger C + D^\dagger D = I$. Otherwise, replace $(X^\dagger X + Y^\dagger Y)^{-1/2}$ by its restriction on the support of $X^\dagger X + Y^\dagger Y$ in the above expressions of C and D, and add to the expression of C a projector onto the null space of $X^\dagger X + Y^\dagger Y$. □

Proof of Theorem 1. Without loss of generality, the LOCC protocol $\mathcal{P}$ has the following form. In the first round, one of the parties (w.l.o.g., the first party) applies a measurement $\mathcal{M}$ with POVM $(A_1^\dagger A_1, A_2^\dagger A_2, \cdots)$, possibly with infinitely many outcomes. Then, the party broadcasts the measurement outcome. In the second round, another party applies another local measurement

that can depend on the first outcome, and broadcasts the second outcome. This goes on, either for some finitely many rounds, say, ℓ, or indefinitely.

We construct $\mathcal{P}'$ from $\mathcal{P}$ as follows.

For simplicity, we focus on the compression method on the first measurement $\mathcal{M}$. Every possible measurement outcome i induces a post-measurement ensemble $S_i = \{p_1 A_i \rho_1 A_i^\dagger, \cdots, p_n A_i \rho_n A_i^\dagger\}$. Denote the probability of the ensemble S_i by q_i. We focus on the set of i's for which $q_i > 0$. Conditioned on the outcome i, S_i / q_i is a normalized ensemble, with some probability of successful discrimination t_i (see Section 2). The t_i's are related to the total success probability by

$$t = \sum_i q_i t_i. \tag{1}$$

We first show that $\mathcal{M}$ can be performed in two stages. and that the first stage can be modified to some $\mathcal{M}'$ to equalize the success probability for each outcome. Assume without loss of generality,

$$t_1 \geq t_2 \geq t_3 \geq \cdots \tag{2}$$

If $t_i = t$ for all i, we are done. So, suppose there exists some k such that $t > t_k$, which also implies $t_1 > t$. There exists $0 < s$ such that

$$\frac{q_1 t_1 + s q_k t_k}{q_1 + s q_k} = t. \tag{3}$$

To see this, note that $t = (1 - \lambda) t_1 + \lambda t_k$ for some $\lambda \in (0, 1)$. Then, it suffices for $\frac{s q_k}{q_1 + s q_k} = \lambda$, which holds if

$$s = \frac{\lambda q_1}{(1 - \lambda) q_k}. \tag{4}$$

We now consider the two cases $s \leq 1$ and $s > 1$ separately.

If $s \leq 1$, let $B = \sqrt{A_1^\dagger A_1 + s A_k^\dagger A_k}$. Consider the induced ensemble $BS = \{p_1 B \rho_1 B^\dagger, \cdots, p_n B \rho_n B^\dagger\}$. The probability of the ensemble BS is equal to $q_1 + s q_k$. We now show that $BS / (q_1 + s q_k)$ has success probability $(q_1 t_1 + s q_k t_k) / (q_1 + s q_k)$, which equals to t. To see this:

> Consider a modification to $\mathcal{M}$ by replacing A_1 and A_k by B and $\sqrt{1 - s} A_k$ respectively. Call the resulting measurement $\tilde{\mathcal{M}}$. If $BS / (q_1 + s q_k)$ has probability of success greater than t, replacing $\mathcal{M}$ by $\tilde{\mathcal{M}}$ in $\mathcal{P}$ outperforms $\mathcal{P}$, contradicting its optimality. Conversely, consider the application to the ensemble BS a binary measurement $\mathcal{N}$ defined by the POVM $(C^\dagger C, D^\dagger D)$ where C, D are obtained as in Lemma 2 with $X = A_1$ and $Y = \sqrt{s} A_k$. The lemma guarantees that $\mathcal{N}$ is a valid measurement on the postmeasurement space of $\tilde{\mathcal{M}}$. Using Lemma 2 to combine the effects due to $\tilde{\mathcal{M}}$ and $\mathcal{N}$, one can see that the outcome of $\mathcal{N}$ corresponding to $C^\dagger C$ induces the postmeasurement ensemble S_1 while the outcome corresponding to $D^\dagger D$ induces the postmeasurement ensemble $s S_k$. Therefore, $BS / (q_1 + s q_k)$ has success probability at least $(q_1 t_1 + s q_k t_k) / (q_1 + s q_k)$ which is equal to t (see (3)).

Observe that modifying $\mathcal{M}$ to $\tilde{\mathcal{M}}$ replaces t_1 by t, q_1 by $q_1 + s q_k$, q_k by $(1 - s) q_k$, while t_k is left unchanged. Also, applying $\mathcal{N}$ after $\tilde{\mathcal{M}}$ gives the original $\mathcal{M}$.

If $s > 1$, (4) can be rewritten as $\frac{1}{s} = \frac{(1 - \lambda) q_k}{\lambda q_1}$. A similar argument holds (and we do not repeat it here), with A_1 and A_k interchanged. In this case, we replace A_k by $B' = \sqrt{\frac{1}{s} A_1^\dagger A_1 + A_k^\dagger A_k}$ and A_1 by $\sqrt{1 - \frac{1}{s}} A_1$ to obtain $\tilde{\mathcal{M}}$, and t_k is replaced by t.

7

In either case, modifying $\mathcal{M}$ into $\tilde{\mathcal{M}}$ *strictly* increases the probability to have an induced postmeasurement ensemble that has probability of success equal to t. We repeat this modification until all postmeasurement ensembles have probability of success equal to t (a property we need later when we reduce the number of outcomes). The resulting measurement is the desired first stage measurement $\mathcal{M}'$, say, with POVM $(B_1^\dagger B_1, B_2^\dagger B_2, \cdots)$. Also, from the above discussion, for each outcome of $\mathcal{M}'$, there is a subsequent second stage binary measurement that completes $\mathcal{M}$.

In the next step, we replace $\mathcal{M}'$ by $\mathcal{M}''$ which has only d_1^2 measurement outcomes, where d_1 is the dimension of the system measured (and held by the first party). For this we use Carathéodory's Theorem (which has a constructive proof):

Lemma 3 (Carathéodory's Theorem [39]). *Let H be a subset of $\mathbb{R}^n$ and $\mathrm{conv}(H)$ its convex hull. Then any $x \in \mathrm{conv}(H)$ can be expressed as a convex combination of at most $n+1$ elements of H.*

To rewrite the sum $\sum_i B_i^\dagger B_i = I$, note that $\sum_i u_i \frac{B_i^\dagger B_i}{\mathrm{tr} B_i^\dagger B_i} = \frac{I}{d_1}$, where $u_i = \frac{\mathrm{tr} B_i^\dagger B_i}{d_1}$ form a distribution. So, we can apply Carathéodory's Theorem with $H = \{\frac{B_i^\dagger B_i}{\mathrm{tr} B_i^\dagger B_i}\}_i$ which is a subset of all trace 1 $d_1 \times d_1$ hermitian matrices with $n = d_1^2 - 1$, and obtain I as a sum of at most d_1^2 operators, each is a positive multiple of some $B_i^\dagger B_i$. This new sum defines a new first stage measurement $\mathcal{M}''$, which is similar to of $\mathcal{M}'$, but now only d_1^2 outcomes are possible. For each outcome of $\mathcal{M}''$, the induced postmeasurement ensemble is the same as in $\mathcal{M}'$ and has success probability t.

Finally, for each outcome of $\mathcal{M}''$, we apply the binary measurement that brings the postmeasurement ensemble back to that of $\mathcal{M}$. The total number of outcomes is at most $2d_1^2$. This completes the compression of the first measurement $\mathcal{M}$.

After the first round of communication, conditioned on each outcome, the parties now hold a new, normalized, ensemble, and they try their best to discriminate it (with $\ell - 1$ rounds of communication if $\mathcal{P}$ has ℓ rounds). A similar compression can now be applied to the next measurement. Repeating the process, each measurement in the protocol has no more than $2d_{\mathrm{loc}}^2$ outcomes. If $\mathcal{P}$ has ℓ rounds, the total number of outcomes is at most $2^\ell d^{2\ell}$, where $d = \max\{d_1, d_2, \cdots, d_m\}$ is the maximum local dimension. $\qquad\square$

Note that without the constraint of being in an LOCC protocol, a measurement on a d-dimensional system can be compressed to d^2 outcomes. This bound $2d^2$ shows that to optimize state discrimination in LOCC, about twice as many outcomes (or one additional bit of communication) are sufficient. This is independent on the number of rounds (and finite or not), how deep the parties have executed the protocol, the number of parties or the total dimension of the system, and not on the number of states in S. In comparison, in [36] each measurement in round ℓ out of r has at most $nD^{4(r-\ell+1)}$ outcomes where n is the number of outcomes, after coarse-graining, at the end of the protocol (which is $|S|$ for state discrimination) and D is the global dimension. Most importantly, this bound diverges when r diverges.

If we apply Carathéodory's Theorem (Lemma 3) to the original POVM $\{A_i^\dagger A_i\}$ to reduce the number of measurement outcomes, the probability of discrimination need not be preserved. We introduce the first stage modification to equalize the probability of correct discrimination for each outcome, and need to add a second stage measurement, thereby getting an additional factor of 2 in the bound $2d_1^2$. The next method improves the bound to d_1^2. It is based on the limited number of outcomes for extremal measurement and the limited size of the support of extremal distributions, both of which are corollaries of Carathéodory's Theorem.

8

3.2 The second method and improved bounds

The second method implements any LOCC protocol with finite width per measurement and shared randomness.

Theorem 4. *Let $\mathcal{P}$ be a fine-grained LOCC protocol (see the end of Section 2) implementing an instrument $\mathcal{I}$. Then, $\mathcal{P}$ can be written as a convex combination $\mathcal{P} = \sum_i \lambda_i \mathcal{P}^{(i)}$, where:*

1. *each $\mathcal{P}^{(i)}$ is an LOCC protocol implementing some instrument $\mathcal{L}^{(i)}$;*

2. *each $\mathcal{P}^{(i)}$ has the same tree structure as $\mathcal{P}$;*

3. *each edge CP map $\mathcal{L}_e^{(i)}$ is proportional to the corresponding edge CP map $\mathcal{L}_e$ of $\mathcal{P}$;*

4. *$\mathcal{L}_e = \sum_i \lambda_i \mathcal{L}_e^{(i)}$;*

5. *for each i and each vertex v associated with a local operation on a d_{loc}-dim system, at most d_{loc}^2 outgoing edges of v have nonzero edge CP maps.*

To prove the above, we first describe and prove a corollary of Carathéodory's Theorem.

Corollary 5 (Improved Carathéodory's Theorem). *Let $H = \{v_i\} \subset \mathbb{R}^n$, $v \in \mathbb{R}^n$. Consider the set of probability distributions p on H with barycentre v, i.e.*

$$P(H;v) := \left\{ p \text{ p.d. s.t. } v = \sum_i p_i v_i \right\}.$$

This set is closed and convex. Its extreme points have support cardinality at most $n+1$.

This corollary allows us to write any original probability distribution with barycentre v, which by definition is an element of $P(H;v)$, as a convex combination of such extremal distributions, each of which has support at most $n+1$.

Proof of Corollary 5. The convexity and closedness are clear. We prove the statement concerning the support cardinality of the extremal points of $P(H;v)$ via its contrapositive. Consider any given $q \in P(H;v)$ with support S of size $|S| \geq n+2$. This gives an expression of $v = \sum_{i \in S} q_i v_i$ in which all $q_i > 0$. But this just says that v is in the convex hull of $\{v_i : i \in S\}$, so, we can use Lemma 3 to express v as a convex combination of at most $n+1$ elements of S, $v = \sum_{i \in S} r_i v_i$ with some $r_i = 0$. Consider the relation

$$v = \sum_{i \in S} q_i v_i = \sum_{i \in S} r_i v_i.$$

Since for all $i \in S$, $q_i > 0$, there exists a $t > 0$ such that for all $i \in S$, $q_i - t r_i \geq 0$. This gives $q = tr + (1-t)r'$ for some other probability distribution r', which by linearity is also an element of $P(H;v)$. But $q \neq r$ since q has support strictly larger than that of r, therefore, q is not an extreme point of $P(H;v)$. Taking the contrapositive, extreme points of $P(H;v)$ have support cardinality at most $n+1$. $\qquad\square$

A special case of the above corollary upper bounds the number of outcomes in extremal measurements (by choosing the v_i's to be density matrices and v to be the maximally mixed state).

Lemma 6 (Corollary 2.48 in [40]). *For any extremal measurement on a Hilbert space $\mathcal{X}$, there are at most* $\dim(\mathcal{X})^2$ *nonzero POVM elements.* $\qquad\square$

This lemma was used in [41]. Other sources for it include [42] and [43, Corollary 1].

Proof of Theorem 4. As before, it suffices to consider the first measurement on $\mathcal{X}_1$ made by the first party. For any such measurement,

$$\mathcal{M}(\rho) = \sum_i A_i \rho A_i^\dagger \otimes |i\rangle\langle i|,$$

we can always consider the canonical form,

$$\mathcal{M}(\rho) = \sum_i \sqrt{A_i^\dagger A_i} \rho \sqrt{A_i^\dagger A_i} \otimes |i\rangle\langle i|,$$

because there exists isometry U_i such that $A_i = U_i \sqrt{A_i^\dagger A_i}$ and so the two measurements differ only by a conditional isometry, which can be delayed to the next action round of this party. Thus, we only need to consider the POVM of each local measurement.

We can decompose this POVM as a convex combination of POVMs of extremal measurements. By Lemma 6 above, each extremal measurement has no more than d_1^2 outcomes.

The same reasoning can be applied to subsequent measurements. For each vertex, the maps associated with the outgoing edges may take the state to spaces of different dimensions. To perform the induction through the tree, we need to make the additional observation that, for a fine-grained protocol, the range of each edge map has dimension no bigger than the input dimension. $\qquad\square$

Theorem 7. *Suppose an ensemble of multipartite quantum states $S = \{p_1\rho_1, \cdots, p_n\rho_n\} \subset \mathcal{L}(\otimes_{j=1}^m \mathcal{X}_j)$ with $p_k \geq 0$, $\sum_k p_k = 1$ can be distinguished by some LOCC protocol $\mathcal{P}$ with success probability t. Then there exists an LOCC protocol $\mathcal{P}'$ achieving the same success probability t but in which each measurement requires at most d_{loc}^2 outcomes, where d_{loc} is the dimension of the local system measured. If $\mathcal{P}$ has $\ell < \infty$ many rounds of communication, the total width can be bounded by $d^{2\ell}$, where d is the maximum local dimension.*

Proof. Because the probability of success is linear in the decomposition in Theorem 4, the best slim protocol has probability of success at least t (and at most t by the optimality of the original protocol). So, we can replace the original protocol by this slim protocol, in which each measurement on a system with local dimension d_{loc} has no more than d_{loc}^2 outcomes. $\qquad\square$

From the above proof, it is evident that Theorem 7 applies to the maximization of any function that is linear, or more generally convex, in the LOCC instrument.

To implement $\mathcal{P}$ via slim protocols as given by the decomposition in Theorem 4, the parties need to share randomness. The next theorem bounds the required amount of shared randomness when the implemented instrument has finite input and output dimensions and finitely many classical outcomes.

Theorem 8. *Let $\mathcal{P}$ be a fine-grained LOCC protocol (see the end of Section 2) implementing an instrument $\mathcal{I}$ with n coarse-grained outcomes. Let D_0 be the total input dimension, and D_i be the total output dimension of the CP map conditioned on the outcome being i for $i = 1, \cdots, n$. Then, $\mathcal{I} = \sum_{i=1}^{R} \mu_i \mathcal{I}^{(i)}$, with each $\mathcal{I}^{(i)}$ an instrument implemented by a slim protocol $\mathcal{P}^{(i)}$ satisfying all the conditions in Theorem 4, and $R \leq \sum_{i=1}^{n} D_0^2 D_i^2 - D_0^2 + 1$.*

Proof. From Theorem 4, $\mathcal{I} = \sum_i \lambda_i \mathcal{I}^{(i)}$ where the λ_i's form a probability distribution. The affine space of instruments with n coarse-grained outcomes and with the stated input and output dimensions has dimension $\sum_{i=1}^{n} D_0^2 D_i^2 - D_0^2$ (since the CP map corresponding to the i-th outcome is represented by a hermitian Choi matrix specified by $D_0^2 D_i^2$ real parameters, and the trace-preserving constraint removes D_0^2 real degrees of freedom. Applying Carathéodory's Theorem (Lemma 3), we can rewrite $\mathcal{I} = \sum_i \mu_i \mathcal{I}^{(i)}$ where at most $R = \sum_{i=1}^{n} D_0^2 D_i^2 - D_0^2 + 1$ of the μ_i's are nonzero. So, $\log_2 R$ shared bits of randomness are sufficient. □

We note that for an LOCC protocol $\mathcal{P}$ represented by an infinite tree, Theorem 8 provides an exact implementation of the corresponding LOCC instrument $\mathcal{I}$ as a finite mixture of slim LOCC protocols, each of which can be represented by a potentially infinite tree and each defines a bona fide instrument, with probability 1. The ℓ-round prefix of this compressed infinite protocol converges to $\mathcal{P}$.

4 Acknowledgements

References

[1] A. Peres and W. K. Wootters. Optimal detection of quantum information. *Phys. Rev. Lett.*, 66(9):1119-1122 (1991).

[2] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70, 1895-1899 (1993).

[3] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54, 3824 (1996).

[4] S. Massar and S. Popescu. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. A*, 74, 1259 (1995).

[5] 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. *Phys. Rev. A*, 59:1070-1091 (1999).

[6] C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal. Unextendible Product Bases and Bound Entanglement. *Phys. Rev. Lett.* 82, 5385 (1999).

[7] J. Walgate, A. J. Short, L. Hardy and V. Vedral. Local Distinguishability of Multipartite Orthogonal Quantum States. *Phys. Rev. Lett.* 85, 4972 (2000).

[8] B. M. Terhal, D. P. DiVincenzo, and D. W. Leung. Hiding bits in Bell states. *Phys. Rev. Lett.*, 86(25), 5807-5810 (2001).

[9] D. P. DiVincenzo, D. W. Leung, and B. M. Terhal. Quantum data hiding. *IEEE Trans. Inf. Theory*, 48(3), 580-598 (2002).

[10] T. Eggeling and R. F. Werner. Hiding classical data in multipartite quantum states. *Phys. Rev. Lett.* 89, 097905 (2002).

[11] S. Ghosh, G. Kar and A. Roy, A. Sen(De) and U. Sen Distinguishability of Bell States. *Phys. Rev. Lett.* 87, 277902 (2001).

[12] M. Horodecki, A. Sen(De), U. Sen, and K. Horodecki. Local Indistinguishability: More Nonlocality with Less Entanglement. *Phys. Rev. Lett.* 90, 047902 (2003).

[13] S. De Rinaldis. Distinguishability of complete and unextendible product bases. *Phys. Rev. A* 70, 022309 (2004).

[14] A. Chefles. Condition for unambiguous state discrimination using local operations and classical communication. *Phys. Rev. A* 69, 050307 (2004).

[15] S. Ghosh, G. Kar, A. Roy and D. Sarkar Distinguishability of maximally entangled states. *Phys. Rev. A* 70, 022304 (2004).

[16] H. Fan. Distinguishability and indistinguishability by LOCC. *Phys. Rev. Lett.* 92, 177905 (2004).

[17] M. Nathanson. Distinguishing Bipartite Orthogonal States using LOCC: Best and Worst Cases. *J. Math. Phys.* 46, 062103 (2005).

[18] J. Watrous. Bipartite Subspaces Having No Bases Distinguishable by Local Operations and Classical Communication. *Phys. Rev. Lett.* 95, 080505 (2005).

[19] M. Hayashi, D. Markham, M. Murao, M. Owari and S. Virmani. Bounds on Multipartite Entangled Orthogonal State Discrimination Using Local Operations and Classical Communication. *Phys. Rev. Lett.* 96, 040501 (2006).

[20] S. M. Cohen. Local distinguishability with preservation of entanglement *Phys. Rev. A* 75, 052313 (2007).

[21] R. Duan, Y. Feng, Z. Ji, and M. Ying. Distinguishing Arbitrary Multipartite Basis Unambiguously Using Local Operations and Classical Communication. *Phys. Rev. Lett.* 98, 230502 (2007).

[22] R. Duan, Y. Feng, Y. Xin and M. Ying. Distinguishability of quantum states by separable operations. *IEEE Trans. Inf. Theory* 55, 1320 (2009).

[23] M. Matthews and A. Winter. On the Chernoff distance for asymptotic LOCC discrimination of bipartite quantum states. *Commun. Math. Phys.* 285(1), 161 (2008).

[24] M. Matthews, S. Wehner, and A. Winter. Distinguishability of Quantum States Under Restricted Families of Measurements with an Application to Quantum Data Hiding. *Commun. Math. Phys.* 291(3), 813 (2009).

[25] M. Kleinmann, H. Kampermann, and D. Bruß. Asymptotically perfect discrimination in the local-operation-and-classical-communication paradigm. *Phys. Rev. A*, 84, 042326 (2011).

[26] S. Bandyopadhyay. More Nonlocality with Less Purity. *Phys. Rev. Lett.*, 106, 210402 (2011).

[27] N. Yu, R. Duan and M. Ying. Four Locally Indistinguishable Ququad-Ququad Orthogonal Maximally Entangled States. *Phys. Rev. Lett.*, 109, 020506 (2012).

[28] S. Bandyopadhyay, S. Ghosh and G. Kar. LOCC distinguishability of unilaterally transformable quantum states. *New J. Phys.* 13 123013 (2011).

[29] N. Yu, R. Duan and M. Ying. Any $2 \otimes n$ subspace is locally distinguishable. *Phys. Rev. A* 84, 012304 (2011).

[30] A. Cosentino. Positive-partial-transpose-indistinguishable states via semidefinite programming. *Phys. Rev. A* 87, 012321 (2013)

[31] N. Yu, R. Duan and M. Ying. Distinguishability of Quantum States by Positive Operator-Valued Measures with Positive Partial Transpose. *IEEE Trans. Inf. Theory*, 60(4):2069-2079 (2014).

[32] S. Bandyopadhyay, A. Cosentino, N. Johnston, V. Russo, J. Watrous, N. Yu. Limitations on separable measurements by convex optimization. *IEEE Trans. Inf. Theory*, 61(6), 3593-3604 (2014).

[33] S. M. Cohen. Structure of local quantum operations and classical communication: Finite versus infinite rounds. *Phys. Rev. A*, 91, 042106 (2015).

[34] E. Chitambar. Local Quantum Transformations Requiring Infinite Rounds of Classical Communication. *Phys. Rev. Lett.*, 107, 190502 (2011).

[35] E. Chitambar, W. Cui, and H. Lo. Increasing entanglement monotones by separable operations. *Phys. Rev. Lett.*, 108, 240504 (2012).

[36] E. Chitambar, D. W. Leung, L. Mančinska, M. Ozols, and A. Winter. Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask). *Commun. Math. Phys.*, 328(1), 303-326 (2014).

[37] S. M. Cohen. General approach to quantum channel impossibility by local operations and classical communication. *Phys. Rev. Lett.*, 118, 020501 (2017).

[38] S. M. Cohen. Strong bounds on required resources for quantum channels by local operations and classical communication. *Private communication and upcoming arXiv submission*.

[39] R. Rockafellar. *Convex Analysis*. 2nd printing. Princeton Mathematical Series. Princeton University Press, 1996.

[40] J. Watrous. *The Theory of Quantum Information*. Cambridge University Press, 2018.

[41] E. B. Davies. *Information and Quantum Measurement*. *IEEE Trans. Inf. Theory*, 24(5), 596 (1978).

[42] K. Parthasarathy. Extremal decision rules in quantum hypothesis testing. *Inf. Dim. Analysis, Quantum Prob. Rel. Topics*, 2(4), 557-568 (1999).

[43] G. M. D'Ariano, P. Lo Presti, P. Perinotti. Classical randomness in quantum measurements. *J. Phys. A: Math. Gen.*, 38, 5979-5991 (2005).