Adversarial hypothesis testing and a quantum Stein's Lemma for restricted measurements [1]

Fernando G. S. L. Brandão, Aram W. Harrow, James R. Lee, and Yuval Peres

Abstract

Recall the classical hypothesis testing setting with two sets of probability distributions P and Q. One receives either n i.i.d. samples from a distribution $p \in P$ or from a distribution $q \in Q$ and wants to decide from which set the points were sampled. It is known that the optimal exponential rate at which errors decrease can be achieved by a simple maximum-likelihood ratio test which does not depend on p or q, but only on the sets P and Q.

We consider an adaptive generalization of this model where the choice of $p \in P$ and $q \in Q$ can change in each sample in some way that depends arbitrarily on the previous samples. In other words, in the k^{th} round, an adversary, having observed all the previous samples in rounds $1, \ldots, k-1$, chooses $p_k \in P$ and $q_k \in Q$, with the goal of confusing the hypothesis test. We prove that even in this case, the optimal exponential error rate can be achieved by a simple maximum-likelihood test that depends only on P and Q.

We then show that the adversarial model has applications in hypothesis testing for *quantum states* using restricted measurements. For example, it can be used to study the problem of distinguishing entangled states from the set of all separable states using only measurements that can be implemented with local operations and classical communication (LOCC). The basic idea is that in our setup, the deleterious effects of entanglement can be simulated by an adaptive classical adversary.

We prove a quantum Stein's Lemma in this setting: In many circumstances, the optimal hypothesis testing rate is equal to an appropriate notion of quantum relative entropy between two states. In particular, our arguments yield an alternate proof of Li and Winter's recent strengthening of strong subadditivity for von Neumann entropy.

Keywords: Hypothesis testing, quantum information theory, quantum Stein's Lemma, subadditivity, von Neumann entropy

I. Introduction

Asymmetric hypothesis testing is the problem of distinguishing between two sources where one wants to minimize the rate of false positives (type-1 error) subject to a constraint on the rate of false negatives (type-2 error). In the case of n i.i.d. samples from a classical or quantum source, a central result is the Chernoff-Stein Lemma [13], [31], [46] which states that for any constant bound on the type-2 error, the optimal type-1 error decreases at an exponential rate whose exponent is given by the classical (respectively, quantum) relative entropy. Similar results hold even when we generalize the problem so that the sources are described by an unknown parameter and one needs to design a test that works for any choice of the parameter [32].

Part of this work appeared in the Proceedings of the 5th conference on Innovations in Theoretical Computer Science (ITCS 2014).

Caltech
MIT
University of Washington

First main result: Adversarial hypothesis testing. In the first part of this paper (Section II), we generalize this problem further to allow the parameter to vary adaptively from sample to sample. Since we will allow the parameter to depend arbitrarily on previous samples, this can be thought of as *adversarial* hypothesis testing. That is, we wish to devise a test that can distinguish between samples from two different sets even against an adversary that can choose the distribution in each round based on which samples have previously been observed.

There are some simple cases where it is not hard to see that this additional power cannot help the adversary. For example, suppose we are given a coin with heads probability p and wish to distinguish between the cases where $p \in [0, 1/3]$ and where $p \in [2/3, 1]$. It is straightforward to show that this general problem is no harder than simply distinguishing a $1/3$-biased coin from a $2/3$-biased coin; equivalently, the adversary gains no advantage from the ability to be adaptive. On the other hand, distinguishing between the two settings $p \in \{1/3, 2/3\}$ and $p = 1/2$ is clearly impossible, as the adversary can simply choose with probability $1/2$ to flip the $1/3$-biased coin, and with probability $1/2$ to flip the $2/3$-biased coin. The resulting distribution of samples is indistinguishable from the one arising from $p = 1/2$. This stresses the role of *convexity* since even a non-adaptive adversary can simulate a convex combination of distributions by choosing randomly among them.

We will prove in Theorem 2 that this property is sufficient to characterize the optimal error rate for asymmetric hypothesis testing against an adaptive adversary. Specifically, if the two sources vary over convex sets of probability distributions, then the problem is no harder than in the i.i.d. case. Our Theorem 7 also establishes a version of this claim for symmetric hypothesis testing. These two results can be thought of as adversarial versions of the classic Chernoff-Stein Lemma and Chernoff's Theorem, respectively. Results in this direction were previously established for arbitrarily varying sources [19] which can be viewed as a special case of a non-adaptive adversary.

Quantum hypothesis testing, entanglement, and additivity. One of our main applications for our adversarial Chernoff-Stein Lemma is in quantum hypothesis testing, when the states to be distinguished need not be i.i.d. Indeed, a recurrent challenge in quantum information theory is that even apparently i.i.d. problems can involve complicated entangled states (meaning that they cannot be written as a convex combination of independent states). For example, the quantum capacity of an i.i.d channel requires maximizing over all n-component inputs, and in general it is known that achieving the capacity requires using states that are entangled across channel uses [16], [26]. This phenomenon in quantum information theory—where information-theoretic quantities for n copies of a system are not simply n times the one-copy quantity—is known generally as the "additivity" problem.

A similar additivity problem arises in quantum hypothesis testing when we wish to distinguish many copies of a fixed state against a family of states that include non-i.i.d. states. One important example is the *relative entropy of entanglement* E_R, which is a method of quantifying the entanglement in a state ρ as the minimum of its relative entropy with respect to any separable (i.e. non-entangled) state. Here, ρ is a multipartite state (e.g., shared between systems A, B, C) and separability refers to this partition. However, to establish the asymptotic hypothesis testing rate of ρ against separable states, we need to compare n copies of ρ against states that are separable with respect to our original partition, but not necessarily across the different copies. In our example, $\rho^{\otimes n}$ lives on systems $A_1, B_1, C_1, \ldots, A_n, B_n, C_n$ and we need to compare against states that are separable across the $A_1 \ldots A_n : B_1 \ldots, B_n : C_1 \ldots C_n$ partition, but possibly entangled within the $A_1, \ldots, A_n$ systems (and the $B_1, \ldots, B_n$ and $C_1, \ldots, C_n$ systems). Indeed, such entanglement across copies is known to be necessary to compute the relative entropy of entanglement, since examples exist [56] where $E_R(\rho \otimes \rho) < 2E_R(\rho)$.

Second main result: Restricted measurements. A further difficulty arises in the quantum setting when we consider restricted families of measurements, such as those arising from locality restrictions. Here, too,

the optimal measurement can be entangled across copies. Moreover, since the hypothesis testing problem involves maximizing distinguishability over allowable measurements and minimizing over states, it is possible for entanglement to either increase or decrease the rate.

One particularly relevant example for our work involves distinguishing many copies of a state ρ against a general separable state, using measurements from a class (such as 1-LOCC, defined below) which preserves the set of separable states. This distinguishability scenario was studied extensively in [48], [11], [40], [10]. Though it may initially seem to be an obscure question, it has found applications to understanding the quantum conditional mutual information [11], to channel coding [42], and to classical algorithms for separability testing [12] and the small-set expansion problem [2].

The main result of Section III provides quantum versions of the Chernoff-Stein Lemma and Chernoff's theorem for restricted measurements. The main idea is that the deleterious effects of entanglement in this setting are no worse than what could be achieved by an adaptive adversary. Thus quantum analogues follow as a corollary of our classical results. One application of these results is an alternate proof of the improved strong subadditivity inequality of Li and Winter [40].

Adaptive measurements. The main results in our paper show that certain variants of hypothesis testing are no more difficult than the original problem. Namely, in the classical case, we can allow an adversary to adaptively change the distribution without decreasing the hypothesis testing exponent, and in the quantum case, we can allow entangled states (under some conditions) while again achieving the same performance. A natural complementary question is whether hypothesis testing rates can be improved by allowing the distinguisher a broader family of tests. For example, classically one could consider the problem of distinguishing between two channels (stochastic maps) instead of between two probability distributions, and allowing the distinguisher to adaptively change the inputs to those channels. In the quantum setting, one might consider the problem of distinguishing $\rho^{\otimes n}$ from $\sigma^{\otimes n}$ using entangled and/or adaptive measurements.

This sort of adaptivity often does not help. When distinguishing two classical channels, there is essentially no advantage to using varying inputs [29]. On the other hand, in the quantum case, when given n copies of a state, entangled measurements across the n copies *can* improve the hypothesis-testing rate (see (55) and the surrounding discussion). However, if measurements are forced to be separable across the n copies, then adaptivity is again of no help [29] (see also [28, Section 3.5]). Thus the results in [29], [28] concern quite a different model (adaptivity of the tester and not of the adversary), and are thus incomparable to ours. Note that we also consider a different notion of separability, corresponding to cuts of the form $A_1 \dots A_n : B_1 \dots B_n$ instead of $A_1B_1 : A_2B_2 : \cdots : A_nB_n$.

II. Hypothesis testing against an adaptive adversary

A. Asymmetric hypothesis testing

Fix two distributions p and q over a finite domain Ω. Given i.i.d. samples $X_1, X_2, \dots, X_n$ from a distribution $r \in \{p, q\}$, the goal is to design a test which distinguishes the two possibilities based on the sample. The classical *Chernoff-Stein Lemma* characterizes the optimal exponential rate of error decay achievable in the one-sided error setting.

Consider any acceptance region $A_n \subseteq \Omega^n$ and the corresponding error probabilities $\alpha_n = p^n(A_n^c)$ and $\beta_n = q^n(A_n)$, where we use S^c to denote the complement of a set S. Then for $0 < \varepsilon < 1$, define

$$\beta_n^\varepsilon := \min_{\substack{A_n \subseteq \Omega^n \\ \alpha_n < \varepsilon}} \beta_n \,,$$

and denote the optimal error exponent

$$\mathcal{E}^\varepsilon(p, q) := \lim_{n\to\infty} \frac{-\log \beta_n^\varepsilon}{n} \,.$$

The following well-known lemma characterizes $\mathcal{E}^\varepsilon$ in terms of the relative entropy (see, e.g., Theorem 11.8.3 of [15]).

Lemma 1 (Chernoff-Stein Lemma)**.** *Consider any two distributions p and q over a finite domain Ω. Then $\mathcal{E}^\varepsilon(p,q) = D(p\,\|\,q)$ for any $\varepsilon \in (0,1)$.*

Here, $D(p\,\|\,q)$ is the *relative entropy,* given by

$$D(p\,\|\,q) := \sum_{x\in\Omega} p(x)\log\frac{p(x)}{q(x)}\,,$$

and we take $D(p\,\|\,q) := \infty$ when there is an $x \in \Omega$ such that $p(x) \neq 0$ but $q(x) = 0$.

The adaptive setting. Suppose now that $P, Q \subseteq \mathbb{R}^\Omega$ are closed, convex sets of probability distributions. An *adaptive P-strategy* $\hat{p}$ is a collection of functions $\{\hat{p}_k : \Omega^{k-1} \to P : k = 1, 2, \ldots\}$. Let $\mathcal{A}(P)$ denote the set of all adaptive P-strategies. For $x \in \Omega^n$, we denote

$$\hat{p}(x) := \prod_{k=1}^{n} \hat{p}_k(x_1, \ldots, x_{k-1})(x_k)\,.$$

As before, let $A_n \subseteq \Omega^n$ be an acceptance region, but now we define

$$\alpha_n := \sup_{\hat{p}\in\mathcal{A}(P)} \hat{p}(A_n^c)\,,$$

and

$$\beta_n^\varepsilon := \min_{\substack{A_n\subseteq\Omega^n \\ \alpha_n<\varepsilon}} \sup_{\hat{q}\in\mathcal{A}(Q)} \hat{q}(A_n)\,.$$

Foor $\varepsilon \in (0,1)$, we denote the *adversarial one-sided error exponent* by

$$\mathcal{E}^\varepsilon_{\mathrm{adv}}(P,Q) := \lim_{n\to\infty} \frac{-\log\beta_n^\varepsilon}{n}\,.$$

Observe that for single distributions $p, q \in \mathbb{R}^\Omega$, we have $\mathcal{E}^\varepsilon_{\mathrm{adv}}(\{p\},\{q\}) = \mathcal{E}^\varepsilon(p,q)$.

Theorem 2 (Adversarial Chernoff-Stein)**.** *Let Ω be a finite domain. For any closed, convex sets of probability distributions $P, Q \subseteq \mathbb{R}^\Omega$ and for any $\varepsilon \in (0,1)$, we have*

$$\mathcal{E}^\varepsilon_{\mathrm{adv}}(P,Q) = \min_{p\in P, q\in Q} D(p\,\|\,q). \tag{1}$$

Thus in the asymptotic regime, adversarial adaptive hypothesis testing is no harder than the i.i.d. setting. Indeed, when the distributions in P have full support, the hypothesis test used is a simple Neyman-Pearson test for p, q minimizing the RHS of (1). This result was previously known in the non-adaptive case, where it is sometimes referred to as *composite hypothesis testing* [38].

Proof of Theorem 2. We may assume that P and Q are compact; the general case can be reduced to this one by considering exhaustions of P and Q by compact convex sets. Let $p^* \in P$ and $q^* \in Q$ be minimizers of $D(p\,\|\,q)$ as p and q vary over P and Q, respectively. Since P and Q are compact and $D(p\,\|\,q)$ is lower semi-continuous, such p^*, q^* exist.

By considering non-adaptive strategies that simply play p^* and q^* in each coordinate, one sees that

$$\mathcal{E}^\varepsilon_{\mathrm{adv}}(P,Q) \leq \mathcal{E}^\varepsilon_{\mathrm{adv}}(\{p^*\},\{q^*\}) = \mathcal{E}^\varepsilon(p^*,q^*) = D(p^*\,\|\,q^*), \tag{2}$$

where the last equality is Lemma 1. Thus we need only prove that

$$\mathcal{E}^\varepsilon_{\mathrm{adv}}(P,Q) \geq D(p^*\,\|\,q^*). \tag{3}$$

Note that if $D(p^* \,\|\, q^*) = 0$, then (3) holds vacuously, thus we may assume that P, Q are disjoint.

We will establish that (3) holds under the assumption

$$\mathrm{supp}(p) = \mathrm{supp}(q) = \Omega \quad \forall p \in P, q \in Q. \tag{4}$$

For any distribution p over Ω, write $p_\theta := (1-\theta)p + \theta\frac{\mathbf{1}_\Omega}{|\Omega|}$, and denote $P_\theta := (1-\theta)P + \theta\frac{\mathbf{1}_\Omega}{|\Omega|}$ and $Q_\theta := (1-\theta)Q + \theta\frac{\mathbf{1}_\Omega}{|\Omega|}$. Since P and Q are disjoint compact convex sets, P_θ and Q_θ are disjoint compact convex sets for $\theta > 0$ sufficiently small. At the end of the argument, we will prove the following lemma.

Lemma 3. *For every pair of compact convex sets $P, Q \subseteq \mathbb{R}^\Omega$ and $\varepsilon \in (0,1)$, it holds that*

$$\mathcal{E}^\varepsilon_{\mathrm{adv}}(P, Q) \geq \lim_{\theta \to 0} \mathcal{E}^\varepsilon_{\mathrm{adv}}(P_\theta, Q_\theta).$$

Thus having established (3) under the assumption (4), we can conclude that it holds for general compact convex P and Q by writing

$$\mathcal{E}^\varepsilon_{\mathrm{adv}}(P, Q) \geq \lim_{\theta \to 0} \mathcal{E}^\varepsilon_{\mathrm{adv}}(P_\theta, Q_\theta) \geq \lim_{\theta \to 0} D(p^*_\theta \,\|\, q^*_\theta) \geq D(p^* \,\|\, q^*),$$

where $\{p^*_\theta, q^*_\theta\}$ is a sequence of minimizers for $\{D(p_\theta \,\|\, q_\theta) : p_\theta \in P_\theta, q_\theta \in Q_\theta\}$, and the latter inequality follows from lower semi-continuity of $(p, q) \mapsto D(p \,\|\, q)$.

So let us now assume (4). For $n \in \mathbb{N}$ and $\delta > 0$, define an acceptance region

$$A_{n,\delta} = \left\{ x \in \Omega^n : \log \frac{p^*(x_1)p^*(x_2)\cdots p^*(x_n)}{q^*(x_1)q^*(x_2)\cdots q^*(x_n)} \geq n(D(p^* \,\|\, q^*) - \delta) \right\}.$$

Our first goal is to argue that for every $\delta > 0$, we have

$$\lim_{n \to \infty} \inf_{\hat{p} \in \mathcal{A}(P)} \hat{p}(A_{n,\delta}) = 1. \tag{5}$$

We will then show that for any adaptive Q-strategy $\hat{q}$, we have

$$\hat{q}(A_{n,\delta}) \leq e^{-n(D(p^* \,\|\, q^*) - \delta)}. \tag{6}$$

Once these are proved, letting $\delta \to 0$ yields the desired claim.

Toward proving (5), observe that, for every $\delta > 0$, $\lim_{n\to\infty}(p^*)^n(A_{n,\delta}) = 1$ by the law of large numbers. The following lemma will allow us to show that the same is true for $\hat{p} \in \mathcal{A}(P)$.

Lemma 4. *If (4) holds, then for any $p \in P$,*

$$\sum_{x \in \Omega} p(x) \log \frac{p^*(x)}{q^*(x)} \geq \sum_{x \in \Omega} p^*(x) \log \frac{p^*(x)}{q^*(x)}.$$

Proof. By Theorem 11.6.1 in [15], we have

$$D(p \,\|\, q^*) \geq D(p \,\|\, p^*) + D(p^* \,\|\, q^*).$$

Observing that $D(p \,\|\, q^*) - D(p \,\|\, p^*) = \sum_{x\in\Omega} p(x) \log \frac{p^*(x)}{q^*(x)}$, we see that this is precisely the desired inequality. □

Now, for $x \in \Omega$, define $L(x) := \log \frac{p^*(x)}{q^*(x)}$. Note that (4) implies

$$m = m(p^*, q^*) := \max\{|L(x)| : x \in \Omega\} < \infty. \tag{7}$$

Moreover, Lemma 4 yields

$$\mathbb{E}_p[L(x)] \geq \mathbb{E}_{p^*}[L(x)] = D(p^* \,\|\, q^*), \qquad \forall p \in P. \tag{8}$$

Let $\hat{p} \in \mathcal{A}(P)$ denote some adaptive P-strategy. Consider a sequence of random variables $\{X_k\}$ distributed according to $\hat{p}$ (i.e., X_k is sampled according to the measure $\hat{p}_k(X_1, X_2, \ldots, X_{k-1}) \in P$), and the corresponding martingale difference sequence

$$D_k := L(X_k) - \mathbb{E}[L(X_k) \mid X_1, \ldots, X_{k-1}] \,.$$

(Recall that the defining property of a martingale difference sequence is that $\mathbb{E}[|D_k|]$ is finite and $\mathbb{E}[D_k \mid X_1, \ldots, X_{k-1}] = 0$ for any $X_1, \ldots, X_{k-1}$.) Since the differences are uniformly bounded (cf. (7)), orthogonality of martingale difference sequences yields

$$\mathbb{E}\left(\sum_{k=1}^{n} D_k\right)^2 = \sum_{k=1}^{n} \mathbb{E}[D_k^2] \le 4m^2 n \,.$$

Chebyshev's inequality then implies that for any $\delta > 0$,

$$\mathbb{P}\left(\sum_{k=1}^{n} D_k \ge -\varepsilon n\right) \ge 1 - \frac{4m^2}{\delta^2}\frac{1}{n} \,. \tag{9}$$

On the other hand, (8) implies that for each k, one has $\mathbb{E}[L(X_k) \mid X_1, \ldots, X_{k-1}] \ge D(p^* \,\|\, q^*)$. Combining this with (9) yields

$$\hat{p}(A_{n,\delta}) = \mathbb{P}\left(\sum_{k=1}^{n} L(X_k) \ge n(D(p^* \,\|\, q^*) - \delta)\right) \ge \mathbb{P}\left(\sum_{k=1}^{n} D_k \ge -\delta n\right) \ge 1 - \frac{4m^2}{\delta^2}\frac{1}{n} \,. \tag{10}$$

Noting that the latter expression goes to 1 as $n \to \infty$ (uniformly in $\hat{p}$) confirms (5). We now turn to verifying (6).

Lemma 5. *For any $q \in Q$, we have*

$$\sum_{x \in \Omega} q(x) \frac{p^*(x)}{q^*(x)} \le 1 \,.$$

Proof. For $\lambda \in [0, 1]$, write $q_\lambda = \lambda q + (1 - \lambda) q^*$. Since q^* is the minimizer of $D(p^* \,\|\, q)$ for q in the convex set Q, we know that the derivative of $D(p^* \,\|\, q_\lambda)$ at $\lambda = 0$ is non-negative.

Calculate

$$\begin{aligned}
\frac{d}{d\lambda} D(p^* \,\|\, q_\lambda) &= \sum_{x \in \Omega} p^*(x) \frac{d}{d\lambda} \log \frac{p^*(x)}{q_\lambda(x)} \\
&= -\sum_{x \in \Omega} p^*(x) \frac{d}{d\lambda} \log \left(\frac{\lambda q(x) + (1 - \lambda) q^*(x)}{p^*(x)}\right) \\
&= -\sum_{x \in \Omega} p^*(x) \frac{q(x) - q^*(x)}{\lambda q(x) + (1 - \lambda) q^*(x)} \,.
\end{aligned}$$

Using the fact that the derivative is non-negative at $\lambda = 0$ yields

$$\sum_{x \in \Omega} \frac{p^*(x) q^*(x)}{q^*(x)} \ge \sum_{x \in \Omega} \frac{p^*(x) q(x)}{q^*(x)} \,,$$

but the left-hand side is equal to 1, yielding the desired result. □

With the preceding lemma in hand, we finish the proof of (6). Fix some adaptive Q-strategy $\hat{q}$. By Markov's inequality,

$$\hat{q}(A_{n,\delta}) \le e^{-n(D(p^* \,\|\, q^*) - \delta)} \, \mathbb{E}_{\hat{q}} \left[\frac{p^*(x_1) \cdots p^*(x_n)}{q^*(x_1) \cdots q^*(x_n)}\right] . \tag{11}$$

We now use the fact that, by Lemma 5, the sequence of likelihood ratios $\prod_{i=1}^n \frac{p^*(x_i)}{q^*(x_i)}$ is a supermartingale with respect to $\hat{q}$. (Recall that a sequence $X_1, X_2, \ldots$ is a supermartingale if $\mathbb{E}[X_n | X_1, \ldots, X_{n-1}] \le X_{n-1}$ for all choices of n and $X_1, \ldots, X_{n-1}$.) In particular,

$$\begin{aligned}
\mathbb{E}_{\hat{q}}\left[\frac{p^*(x_1)\cdots p^*(x_n)}{q^*(x_1)\cdots q^*(x_n)}\right] &= \mathbb{E}_{\hat{q}}\left[\frac{p^*(x_1)\cdots p^*(x_{n-1})}{q^*(x_1)\cdots q^*(x_{n-1})}\mathbb{E}_{\hat{q}_n(x_1,x_2,\ldots,x_{n-1})}\frac{p^*(x)}{q^*(x)}\right] \\
&\le \mathbb{E}_{\hat{q}}\left[\frac{p^*(x_1)\cdots p^*(x_{n-1})}{q^*(x_1)\cdots q^*(x_{n-1})}\right] \\
&\le \cdots \\
&\le 1, \qquad (12)
\end{aligned}$$

where in the second line we have applied Lemma 5 to the distribution $\hat{q}_n(x_1, x_2, \ldots, x_{n-1}) \in Q$, and then we have continued by induction. Combining this with (11) completes our verification of (6) and hence our proof of the theorem.

The proof of the theorem then follows from Lemma 3. □

Proof of Lemma 3. Let $(p_\theta^*, q_\theta^*) \in P_\theta \times Q_\theta$ be a pair minimizing $D(p_\theta \,\|\, q_\theta)$ over $(p_\theta, q_\theta) \in P_\theta \times Q_\theta$. Note that $D(p_\theta^* \,\|\, q_\theta^*) < \infty$ for $\theta > 0$.

For any $\delta > 0$, define the acceptance region

$$A_{n,\delta}^\theta := \left\{ x \in \Omega^n : \log \frac{p_\theta^*(x_1)p_\theta^*(x_2)\cdots p_\theta^*(x_n)}{q_\theta^*(x_1)q_\theta^*(x_2)\cdots q_\theta^*(x_n)} \ge n\left(D(p_\theta^* \,\|\, q_\theta^*) - \delta\right) \right\}.$$

Let $\hat{p}$ and $\hat{q}$ denote an adaptive P-strategy and Q-strategy, respectively, and define adaptive P_θ and Q_θ-strategies by

$$\begin{aligned}
(\hat{p}_\theta)_k(x_1, \ldots, x_{k-1}) &= (\hat{p}_k(x_1, \ldots, x_{k-1}))_\theta \\
(\hat{q}_\theta)_k(x_1, \ldots, x_{k-1}) &= (\hat{q}_k(x_1, \ldots, x_{k-1}))_\theta.
\end{aligned}$$

Then from the super martingale property (12),

$$\begin{aligned}
1 \ge \mathbb{E}_{\hat{q}_\theta}\left[\prod_{i=1}^n \frac{p_\theta^*(x_i)}{q_\theta^*(x_i)}\right] &= \mathbb{E}_{\hat{q}_\theta}\left[\mathbb{E}_{(\hat{q}_\theta)_n(x_1,\ldots,x_{n-1})}\left[\frac{p_\theta^*(x_n)}{q_\theta^*(x_n)}\right]\prod_{i=1}^{n-1}\frac{p_\theta^*(x_i)}{q_\theta^*(x_i)}\right] \\
&\ge (1-\theta)\mathbb{E}_{\hat{q}_\theta}\left[\mathbb{E}_{\hat{q}_n(x_1,\ldots,x_{n-1})}\left[\frac{p_\theta^*(x_n)}{q_\theta^*(x_n)}\right]\prod_{i=1}^{n-1}\frac{p_\theta^*(x_i)}{q_\theta^*(x_i)}\right] \\
&\ge \cdots \\
&\ge (1-\theta)^n \mathbb{E}_{\hat{q}}\left[\prod_{i=1}^n \frac{p_\theta^*(x_i)}{q_\theta^*(x_i)}\right] \\
&\ge (1-\theta)^n e^{n(D(p_\theta^* \,\|\, q_\theta^*)-\delta)}\hat{q}(A_{n,\delta}^\theta). \qquad (13)
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\hat{p}(A_{n,\delta}^\theta) &\ge (1-\theta)^n \hat{p}_\theta(A_{n,\delta}^\theta) \\
&\ge \hat{p}_\theta(A_{n,\delta}^\theta) - \theta n \\
&\ge 1 - \frac{4m(p_\theta^*, q_\theta^*)^2}{\delta^2}\frac{1}{n} - \theta n, && \text{with } m(\cdot,\cdot) \text{ defined in (7)} \\
&\ge 1 - \frac{4\log^2(1/\theta)}{\delta^2}\frac{1}{n} - \theta n, && \text{since } m(p_\theta^*, q_\theta^*) \le \log(1/\theta)
\end{aligned}$$

This bound approaches 1 as long as θ decreases at an appropriate rate with n, say by taking $\theta = 1/n^2$. As a result,

$$\lim_{n\to\infty} \inf_{\hat{p}\in\mathcal{A}(P)} \hat{p}\left(A_{n,\delta}^{1/n^2}\right) = 1.$$

Combining this with (13) shows that along the sequence of acceptance regions $\{A_{n,\delta}^{1/n^2}\}$, we have

$$\begin{aligned}\mathcal{E}_{\mathrm{adv}}^{\varepsilon}(P,Q) &\geq \lim_{n\to\infty}\left(D(p_{1/n^2}^* \,\|\, q_{1/n^2}^*) - \delta + \log(1-n^{-2})\right)\\ &\geq \lim_{n\to\infty}\left(\mathcal{E}_{\mathrm{adv}}^{\varepsilon}(P_{1/n^2}, Q_{1/n^2}) - \delta\right),\end{aligned}$$

where the second inequality follows from (2). Now taking $\delta \to 0$ completes the proof. □

B. Chernoff information and symmetric hypothesis testing

Suppose again that we have two distributions p and q over a finite domain Ω. We also have n i.i.d. samples $X_1, X_2, \ldots, X_n$ from a distribution $r \in \{p, q\}$, and a Bayesian hypothesis: The samples come from p with probability π_p and from q with probability π_q. Consider a test $T_n \subseteq \Omega^n$. If $(X_1, X_2, \ldots, X_n) \in T_n$, we declare that the sample came from p.

Our goal is to minimize the expected error

$$\delta_n(T_n) := \pi_p\, p^n(T_n^c) + \pi_q\, q^n(T_n)\,.$$

In this case, the best achievable error exponent is

$$\gamma(p,q) := \lim_{n\to\infty} -\frac{1}{n} \min_{T_n\subseteq\Omega^n} \log \delta_n(T_n)\,.$$

Observe that the constants π_p and π_q do not affect $\gamma(p,q)$.

For $\lambda \in (0,1)$, let us define

$$\Gamma^{\lambda}(p,q) := -\log \sum_{x\in\Omega, p(x)q(x)>0} p(x)^{\lambda} q(x)^{1-\lambda}\,,$$

and

$$\Gamma^*(p,q) := \sup_{\lambda\in(0,1)} \Gamma^{\lambda}(p,q)\,. \tag{14}$$

We have the following characterization due to Chernoff (see, e.g., Theorem 11.9.1 of [15]).

Theorem 6. *For any distributions p and q on Ω, one has*

$$\gamma(p,q) = \Gamma^*(p,q)\,.$$

Moreover, if $\mathrm{supp}(p) = \mathrm{supp}(q)$[1]*, then one has*

$$\gamma(p,q) = \Gamma^*(p,q) = D(r\,\|\,p) = D(r\,\|\,q)\,,$$

where r is the distribution given by

$$r(x) := \frac{p(x)^{\lambda(p,q)} q(x)^{1-\lambda(p,q)}}{\sum_{y\in\Omega} p(y)^{\lambda(p,q)} q(y)^{1-\lambda(p,q)}}\,,$$

and $\lambda(p,q)$ is the unique value of $\lambda \in (0,1)$ achieving the supremum in (14).

[1]The statement of Theorem 11.9.1 in [15] does not include the condition that $\mathrm{supp}(p) = \mathrm{supp}(q)$, but as was pointed out to us by an anonymous referee, there are examples where the theorem is false without this assumption.

We will prove a corresponding theorem in the adaptive setting. To this end consider again two closed, convex sets of distributions $P, Q \subseteq \mathbb{R}^\Omega$. Define the *adversarial two-sided error exponent*

$$\gamma_{\mathrm{adv}}(P,Q) := \lim_{n\to\infty} -\frac{1}{n} \min_{T_n \subseteq \Omega^n} \max_{\hat{p},\hat{q}} \log\left(\hat{p}(T_n^c) + \hat{q}(T_n)\right)$$

where the maximum is over all adaptive P-strategies $\hat{p}$ and adaptive Q-strategies $\hat{q}$.

Theorem 7 (Adversarial Chernoff's Theorem)**.** *For any finite domain Ω and closed, convex sets of distributions $P, Q \subseteq \mathbb{R}^\Omega$, we have*

$$\gamma_{\mathrm{adv}}(P,Q) = \min_{p\in P, q\in Q} \Gamma^*(p,q)\,. \tag{15}$$

Proof. We may assume that P and Q are compact; the general case can be reduced to this one by considering exhaustions of P and Q by compact convex sets. Assume P and Q are disjoint, since otherwise $\gamma_{\mathrm{adv}}(P,Q) = \min_{p\in P, q\in Q} \Gamma^*(p,q) = 0$. Let $p^* \in P, q^* \in Q$ be some pair that minimizes $\Gamma^*(p,q)$ over $p \in P, q \in Q$. First, we have

$$\gamma_{\mathrm{adv}}(P,Q) \le \gamma_{\mathrm{adv}}(\{p^*\},\{q^*\}) = \gamma(p^*,q^*) = \Gamma^*(p^*,q^*)\,,$$

where the latter equality is given by Theorem 6. Thus we are left to prove $\gamma_{\mathrm{adv}}(P,Q) \ge \Gamma^*(p^*,q^*)$.

Let us first assume that $\mathrm{supp}(p) = \mathrm{supp}(q) = \Omega$ for all $p \in P$ and $q \in Q$. After the argument, we will reduce the general case to this one. Consider $p \ne q$. Define $F_{p,q} : [0,1] \to \mathbb{R}$ by

$$F_{p,q}(\lambda) := \sum_{x\in\Omega} p(x)^\lambda q(x)^{1-\lambda}\,,$$

and calculate

$$F'_{p,q}(\lambda) = \sum_{x\in\Omega} p(x)^\lambda q(x)^{1-\lambda} \log\frac{p(x)}{q(x)}\,,$$
$$F''_{p,q}(\lambda) = \sum_{x\in\Omega} p(x)^\lambda q(x)^{1-\lambda} \left(\log\frac{p(x)}{q(x)}\right)^2.$$

Since $p \ne q$, $F''_{p,q}(\lambda) > 0$ for all $\lambda \in (0,1)$. Since additionally $\mathrm{supp}(p) = \mathrm{supp}(q)$,

$$F'_{p,q}(0) = D(p\,\|\,q) > 0$$
$$F'_{p,q}(1) = -D(q\,\|\,p) < 0\,.$$

We conclude that $F_{p,q}(\lambda)$ is minimized at a unique value $\lambda \in (0,1)$. Denote this value by $\lambda(p,q)$ and observe that $\Gamma^*(p,q) = \Gamma^{\lambda(p,q)}(p,q)$. Let $\lambda^* := \lambda(p^*,q^*)$.

Define now

$$T_n := \left\{ x \in \Omega^n : \prod_{i=1}^n p^*(x_i) \ge \prod_{i=1}^n q^*(x_i) \right\}.$$

Fix also an adaptive P-strategy $\hat{p}$ and an adaptive Q-strategy $\hat{q}$. We will show that

$$\Gamma^*(p^*,q^*) \le \lim_{n\to\infty} \frac{-\log(\hat{p}(T_n^c) + \hat{q}(T_n))}{n}\,. \tag{16}$$

We will need to employ the following easy variant of the "envelope theorem."

Lemma 8. *Consider a differentiable function $f : [0,1]^2 \to \mathbb{R}$. Define $V(t) = \inf_{\lambda\in[0,1]} f(\lambda,t)$ and suppose that for every $t \in [0,1]$, there is a unique $\lambda^*(t) \in (0,1)$ such that $V(t) = f(\lambda^*(t),t)$. If λ^* is*

differentiable at $t \in [0,1]$, then $V'(t) = f_2(\lambda^*(t), t)$ where f_2 is the partial derivative of f with respect to its second argument.

Proof. Let f_1 denote the partial derivative of f with respect to its first argument. Writing $V(t) = f(\lambda^*(t), t)$ and applying the chain rule yields

$$V'(t) = f_2(\lambda^*(t), t) + f_1(\lambda^*(t), t)\frac{d}{dt}\lambda^*(t).$$

The second term is zero because $f_1(\lambda^*(t), t) = 0$ by optimality of $\lambda^*(t)$. □

Remark 9. *Observe that if $f(\lambda, t)$ has $\frac{\partial^2}{\partial \lambda^2} f(\lambda, t) > 0$ for some $t \in [0,1]$, then $\lambda^*(t)$ is the unique solution of $\frac{\partial}{\partial \lambda} f(\lambda, t) = 0$ and is differentiable by the implicit function theorem. Note that the assumptions of Lemma 8 can be relaxed considerably; see, e.g., [43, Ch. 3].*

This allows us to prove the following.

Lemma 10. *For any distribution $q \in Q$, one has*

$$\sum_{x \in \Omega} q(x)\frac{p^*(x)^{\lambda^*}}{q^*(x)^{\lambda^*}} \leq \sum_{x \in \Omega} q^*(x)\frac{p^*(x)^{\lambda^*}}{q^*(x)^{\lambda^*}}\,.$$

Proof. For $t \in [0,1]$, define a distribution $q_t := tq + (1-t)q^* \in Q$. Moreover, define a function $f : [0,1]^2 \to \mathbb{R}$ by

$$f(\lambda, t) := F_{p^*, q_t}(\lambda)\,.$$

As we have already observed, for every fixed value of $t \in [0,1]$, it holds that $\lambda(p^*, q_t) \in (0,1)$ is the unique minimizer of $f(\lambda, t)$.

Let f_2 be the partial derivative of f in its second argument; then one computes:

$$f_2(\lambda, t) = \sum_{x \in \Omega}(q(x) - q^*(x))(1-\lambda)q_t(x)^{-\lambda}p^*(x)^{\lambda}\,.$$

If we let $V(t) = \min_{\lambda \in (0,1)} f(\lambda, t)$, then optimality of q^* implies $V'(0) \leq 0$. But now Lemma 8 (in conjunction with Remark 9) yields

$$\begin{aligned} 0 \geq V'(0) &= f_2(\lambda^*, 0) \\ &= \sum_{x \in \Omega}(q(x) - q^*(x))(1-\lambda^*)q^*(x)^{-\lambda^*}p^*(x)^{\lambda^*}\,. \end{aligned}$$

Rearranging yields the desired claim. □

The preceding lemma shows that the sequence $\prod_{i=1}^n \frac{p^*(x_i)^{\lambda^*}}{q^*(x_i)^{\lambda^*}}$ is a supermartingale with respect to $\hat{q}$. Thus we can write

$$\begin{aligned} \mathbb{E}_{\hat{q}}\left[\prod_{i=1}^n \frac{p^*(x_i)^{\lambda^*}}{q^*(x_i)^{\lambda^*}}\right] &= \mathbb{E}_{\hat{q}}\left[\prod_{i=1}^{n-1} \frac{p^*(x_i)^{\lambda^*}}{q^*(x_i)^{\lambda^*}}\mathbb{E}_{\hat{q}_n(x_1,\ldots,x_{n-1})}\frac{p^*(x_n)^{\lambda^*}}{q^*(x_n)^{\lambda^*}}\right] \\ &\leq e^{-\Gamma^*(p^*,q^*)}\mathbb{E}_{\hat{q}}\left[\prod_{i=1}^{n-1} \frac{p^*(x_i)^{\lambda^*}}{q^*(x_i)^{\lambda^*}}\right] \\ &\leq \cdots \\ &\leq e^{-n\Gamma^*(p^*,q^*)}\,, \end{aligned} \tag{17}$$

where in the second line we have used Lemma 10 along with the fact that $q = \hat{q}_n(x_1, \ldots, x_{n-1}) \in Q$, and then we have continued by induction.

By Markov's inequality, this implies $\hat{q}(T_n) \le e^{-n\Gamma^*(p^*,q^*)}$. By the symmetry of the preceding argument with respect to P and Q, the same bound of $\hat{p}(T_n^c) \le e^{-n\Gamma^*(p^*,q^*)}$ holds for $\hat{p}$. Combining these yields $\gamma_{\mathrm{adv}}(P,Q) \ge \Gamma^*(p^*,q^*)$, completing the proof.

General P and Q. Let us recall from Section II-A the notation p_θ for $p \in \Omega$, and the sets P_θ and Q_θ.

Lemma 11. *For any $\theta > 0$ sufficiently small, it holds that*

$$\gamma_{\mathrm{adv}}(P,Q) \ge \gamma_{\mathrm{adv}}(P_\theta, Q_\theta) - \theta\,.$$

Lemma 12. *It holds that*

$$\liminf_{\theta\to 0} \min_{(p,q)\in P\times Q} \Gamma^*(p_\theta, q_\theta) \ge \min_{(p,q)\in P\times Q} \Gamma^*(p,q)\,.$$

Let us first use them to complete the proof of our desired result for general P and Q using the result for P_θ and Q_θ. Employ Lemma 11 and then Lemma 12 to write:

$$\gamma_{\mathrm{adv}}(P,Q) \ge \liminf_{\varepsilon\to 0} \gamma_{\mathrm{adv}}(P_\theta, Q_\theta) = \liminf_{\theta\to 0} \min_{(p,q)\in P\times Q} \Gamma^*(p_\theta, q_\theta) \ge \min_{(p,q)\in P\times Q} \Gamma^*(p,q)\,.$$

This concludes the proof of Theorem 7, modulo the proofs of Lemmas 11 and 12. □

Proof of Lemma 11. Let $(p_\theta^*, q_\theta^*) \in P_\theta \times Q_\theta$ be a pair minimizing $\Gamma^*(p_\theta, q_\theta)$ over $(p_\theta, q_\theta) \in P_\theta \times Q_\theta$, and let λ^* denote their optimal exponent. Define the test

$$T_n(p_\theta^*, q_\theta^*) := \{x \in \Omega^n : p_\theta^*(x_1)\cdots p_\theta^*(x_n) \ge q_\theta^*(x_1)\cdots q_\theta^*(x_n)\}\,.$$

Let $\hat{q}$ denote an adaptive Q-strategy. We define an adaptive Q_θ-strategy $\hat{q}_\theta$ by

$$(\hat{q}_\theta)_k(x_1,\dots,x_{k-1}) = (\hat{q}_k(x_1,\dots,x_{k-1}))_\theta.$$

Then we have:

$$\exp\left(-n\Gamma^*(p_\theta^*, q_\theta^*)\right) \ge \mathbb{E}_{\hat{q}_\theta}\left[\prod_{i=1}^n \frac{p_\theta^*(x_i)^{\lambda^*}}{q_\theta^*(x_i)^{\lambda^*}}\right] = \mathbb{E}_{\hat{q}_\theta}\left[\mathbb{E}_{(\hat{q}_\theta)_n(x_1,\dots,x_{n-1})}\left[\frac{p_\theta^*(x_n)^{\lambda^*}}{q_\theta^*(x_n)^{\lambda^*}}\right]\prod_{i=1}^{n-1}\frac{p_\theta^*(x_i)^{\lambda^*}}{q_\theta^*(x_i)^{\lambda^*}}\right],$$

where the first inequality uses the supermartingale inequality (17). Moreover, for every $(x_1,\dots,x_{n-1}) \in \Omega^{n-1}$,

$$\mathbb{E}_{(\hat{q}_\theta)_n(x_1,\dots,x_{n-1})}\left[\frac{p_\theta^*(x_n)^{\lambda^*}}{q_\theta^*(x_n)^{\lambda^*}}\right] \ge (1-\theta)\,\mathbb{E}_{\hat{q}_n(x_1,\dots,x_{n-1})}\left[\frac{p_\theta^*(x_n)^{\lambda^*}}{q_\theta^*(x_n)^{\lambda^*}}\right],$$

thus continuing inductively yields

$$\begin{aligned}
\exp\left(-n\Gamma^*(p_\theta^*, q_\theta^*)\right) &\ge \mathbb{E}_{\hat{q}_\theta}\left[\prod_{i=1}^n \frac{p_\theta^*(x_i)^{\lambda^*}}{q_\theta^*(x_i)^{\lambda^*}}\right] \\
&\ge (1-\theta)^n\,\mathbb{E}_{\hat{q}}\left[\prod_{i=1}^n \frac{p_\theta^*(x_i)^{\lambda^*}}{q_\theta^*(x_i)^{\lambda^*}}\right] \\
&\ge (1-\theta)^n \hat{q}\left(\{(x_1,\dots,x_n) \in \Omega^n : p_\theta^*(x_1)\cdots p_\theta^*(x_n) \ge q_\theta^*(x_1)\cdots q_\theta^*(x_n)\}\right) \\
&= (1-\theta)^n\,\hat{q}\left(T_n(p_\theta^*, q_\theta^*)\right).
\end{aligned}$$

Doing the symmetric analysis with an adaptive P-strategy yields

$$\Gamma^*(p_\theta^*, q_\theta^*) \le \theta - \frac{1}{n}\max_{\hat{p},\hat{q}}\left[\log\left(\frac{\hat{q}(T_n(p_\theta^*, q_\theta^*)) + \hat{p}(T_n^c(p_\theta^*, q_\theta^*))}{2}\right)\right],$$

and then taking the limit as $n \to \infty$ gives

$$\gamma_{\mathrm{adv}}(P_\theta, Q_\theta) = \Gamma^*(q^*_\theta, p^*_\theta) \leq \gamma_{\mathrm{adv}}(P, Q) + \theta\,. \qquad \square$$

Proof of Lemma 12. Let $\{(p^n, q^n) \in P \times Q : n = 1, 2, \ldots\}$ denote a sequence of distributions, and consider a sequence $\{\theta_n\}$ with $\theta_n \to 0$ as $n \to \infty$. Since $P \times Q$ is compact, we may pass to a subsequence where (p^n, q^n) converges. Let $(\bar{p}, \bar{q}) \in P \times Q$ be the limit. Note that $(\bar{p}, \bar{q})$ is also a limit of the sequence $\{(p^n_{\theta_n}, q^n_{\theta_n})\}$.

Observe now that $\Gamma^*(p, q)$ is a supremum of continuous functions, and thus $(p, q) \mapsto \Gamma^*(p, q)$ is lower semi-continuous. This implies that

$$\lim_{n\to\infty} \Gamma^*(p^n_{\theta_n}, q^n_{\theta_n}) \geq \Gamma^*(\bar{p}, \bar{q})\,,$$

completing the proof. $\square$

III. Distinguishing quantum states with restricted measurements

A central problem in quantum information is to distinguish between a pair of quantum states ρ and σ. As usual, there is a tradeoff between errors of type 1 and 2, i.e., mistaking ρ for σ and vice versa. The quantum Neyman-Pearson lemma states that the optimal tradeoff curve between errors of type 1 and 2 is achieved by choosing

$$\mathcal{M} = \{\theta\rho - \sigma \geq 0\},$$

for some $\theta \geq 0$, where $\{X \geq 0\}$ denotes the projector onto the eigenvectors of X with nonnegative eigenvalue. The estimation strategy is then to perform the measurement $\{\mathcal{M}, I - \mathcal{M}\}$ and guess ρ upon obtaining the outcome corresponding to POVM element $\mathcal{M}$ or σ upon obtaining the outcome corresponding to $I - \mathcal{M}$.

Remark on terminology: We briefly introduce some notation here, and additional background and definitions for the reader unfamiliar with quantum information theory can be found in Appendix A. The finite domain Ω from Section II is replaced with $V = \mathbb{C}^d$ with the standard Euclidean inner product, and we denote the set of density operators on V by $\mathcal{D}(V)$. Let $\mathcal{L}(V)$ denote linear operators on V and let $E(V) = \{\mathcal{M} \in L(V) : 0 \leq \mathcal{M} \leq I\}$ be the space of POVM elements. A *measurement* $\mathcal{M} = (\mathcal{M}_1, \mathcal{M}_2, \ldots)$ is a collection of POVM elements that sum to I, and $\mathcal{M}(\rho) = (\mathrm{tr}(\mathcal{M}_1\rho), \mathrm{tr}\,\mathcal{M}_2\rho), \ldots)$ refers to the probability distribution of measurement outcomes resulting from applying $\mathcal{M}$ to ρ. For our purposes we will consider both two-outcome measurements and measurements with finitely many nonzero POVM elements. Call these sets $E_2(V)$ and $E_{\mathbb{N}}(V)$ respectively. For $E_2(V)$, the measurement $\{\mathcal{M}, I - \mathcal{M}\}$ is of course determined by the first POVM element $\mathcal{M}$ and so where it is not ambiguous we will use $\mathcal{M}$ to refer to the measurement. Further background on quantum states and measurements can be found in the appendix.

One well-known case of state distinguishability is when ρ and σ have prior probabilities p and $1 - p$, respectively, and we wish to minimize the total probability of error. In this case the optimal measurement $\mathcal{M}$ is given by $\mathcal{M} = \{p\rho - (1 - p)\sigma \geq 0\}$, and the probability of error is $\frac{1 - \|p\rho - (1-p)\sigma\|_1}{2}$, where $\|\cdot\|_1$ denotes the Schatten 1-norm. (Here $\mathcal{M}$ corresponds to guessing "ρ" and $I - \mathcal{M}$ to guessing "σ".) The familiar *trace distance* $\frac{1}{2}\|\rho - \sigma\|_1$ corresponds to the case $p = 1/2$.

We modify this basic problem of state distinguishability in three (simultaneous) ways:

1) We consider only measurements $\mathcal{M}$ from some restricted class $M \subseteq E_2(V)$.
2) We allow ρ, σ to be drawn adversarially from some sets R, S, respectively. (This means that an adversary chooses ρ, σ in each round with knowledge of all previous measurement outcomes.)

3) We consider the asymptotic limit in which M, R, S are replaced by families $\mathbf{M} = (M^1, M^2, \ldots), \mathbf{R} = (R^1, R^2, \ldots), \mathbf{S} = (S^1, S^2, \ldots)$ with M^n, R^n, S^n describing measurements and states on $V^{\otimes n}$. Our goal is then, for each n, to find a measurement $\mathcal{M} \in M^n$ that will effectively distinguish any state $\rho \in R^n$ from any state $\sigma \in S^n$.

These changes render the problem a good deal more abstract, and introduce a large number of new parameters. Thus, it may be helpful to keep in mind a prototypical example that was one of the motivations for this work. For some fixed bipartite state ρ over $A \otimes B$, let R^n be the singleton set $\{\rho^{\otimes n}\}$, and let $S^n := \mathrm{Sep}(A^{\otimes n} : B^{\otimes n})$. This corresponds to studying the asymptotic distinguishability of many copies of ρ from a separable state on the same number of systems. For this special case, we introduce the notation

$$\boldsymbol{\rho} := (\{\rho\}, \{\rho^{\otimes 2}\}, \ldots) \tag{18}$$

$$\mathbf{Sep}(A : B) := (\mathrm{Sep}(A : B), \mathrm{Sep}(A^{\otimes 2} : B^{\otimes 2}), \ldots). \tag{19}$$

Where the context is understood, we will often omit the reference to A, B and simply write Sep or **Sep**. Finally, we will consider a restricted class of measurements $\mathbf{M}$, such as the class of 1-LOCC measurements (as discussed in [48], [11], [40], [10]).

A. Background on restricted quantum measurements

We begin by introducing notation, describing known results on restricted-measurement distinguishability, and presenting a few small new results to help clean up the landscape. In Section III-B, we describe our restricted-measurement version of the quantum Stein's Lemma, and in Section III-C we give an application to quantum conditional mutual information.

1) Quantum Stein's Lemma: If ρ, σ are density matrices on a space V, then the *relative entropy of ρ with respect to σ* is

$$D(\rho \,\|\, \sigma) := \mathrm{tr}\left(\rho(\log \rho - \log \sigma)\right). \tag{20}$$

If $\ker(\sigma) \not\subseteq \ker(\rho)$, we take $D(\rho \,\|\, \sigma) := \infty$.

Following the classical case, we define an *acceptance operator* $\mathcal{M}^n \in E(V^{\otimes n})$ (analogous to the acceptance region T_n), with corresponding error probabilities $\alpha_n = \mathrm{tr}\left((I - \mathcal{M}^n)\rho^{\otimes n}\right)$ and $\beta_n := \mathrm{tr}\left(\mathcal{M}^n \sigma^{\otimes n}\right)$. Again we can define $\beta_n^\varepsilon := \min\{\beta_n : \alpha_n < \varepsilon\}$ and

$$E(\rho, \sigma) := \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{-\log \beta_n^\varepsilon}{n} \tag{21}$$

Hiai and Petz [31] proved the following quantum analogue of Lemma 1:

$$D(\rho \,\|\, \sigma) = E(\rho, \sigma). \tag{22}$$

See also [6], [39] for elegant and elementary proofs. The "strong converse" of (22) was proved by Ogawa and Nagaoka [46], and can be thought of as showing that (22) holds when the limit of $\varepsilon \to 0$ in (21) is replaced by any fixed $\varepsilon \in (0, 1)$.

2) Asymptotic composite hypothesis testing: An important generalization of hypothesis testing is when ρ and σ are chosen from sets $R, S \subseteq \mathcal{D}(V)$, respectively, and we need to design our test with knowledge only of R and S. This problem is known as *composite hypothesis testing* and is closely related to the classical Sanov's theorem.

One case of particular interest to quantum information is when $\rho \in \mathcal{D}(A \otimes B)$ and S is the set of separable states on $A \otimes B$, i.e., $S = \mathrm{Sep}(A : B)$. The quantity $D(\rho \,\|\, \mathrm{Sep}) := D(\rho \,\|\, \mathrm{Sep}(A : B))$ is known as the *relative entropy of entanglement* [55] and has been widely studied as an entanglement measure (see, e.g., Table I in [11]); note that it is usually written as $E_R(\rho)$.

One challenge in working with the relative entropy of entanglement is that $D(\rho^{\otimes n} \,\|\, \mathrm{Sep})$ will not in general be equal to $n \cdot D(\rho \,\|\, \mathrm{Sep})$, reflecting the fact that $\mathrm{Sep}(A^{\otimes n} : B^{\otimes n})$ is larger than the convex hull of $\{\sigma_1 \otimes \cdots \otimes \sigma_n : \sigma_1, \ldots, \sigma_n \in \mathrm{Sep}(A : B)\}$. Intuitively, $\mathrm{Sep}(A^{\otimes n} : B^{\otimes n})$ can be thought of as the set of states on the $2n$ systems $A_1 \ldots A_n B_1 \ldots B_n$ which are separable across the $A_1 \ldots A_n : B_1 \ldots B_n$ cut, but may be entangled arbitrarily among the A systems and among the B systems. This is an example of the quantum-information phenomenon known as the *additivity* problem (see, e.g., [58], [52]).

Definition 1. *Let* $\mathbf{R} = (R^1, R^2, \ldots)$, $\mathbf{S} = (S^1, S^2, \ldots)$, *with* $R^n, S^n \subseteq \mathcal{D}(V^{\otimes n})$. *Then the* asymptotic relative entropy *of* $\mathbf{R}$ *with respect to* $\mathbf{S}$ *is*

$$D(\mathbf{R} \,\|\, \mathbf{S}) := \lim_{n\to\infty} \inf_{\substack{\rho \in R^n \\ \sigma \in S^n}} \frac{D(\rho \,\|\, \sigma)}{n}. \tag{23}$$

We further define

$$\alpha_n(\mathcal{M}) := \sup_{\rho \in R^n} \mathrm{tr}\left((I - \mathcal{M})\rho\right) \tag{24}$$

$$\beta_n(\mathcal{M}) := \sup_{\sigma \in S^n} \mathrm{tr}\left(\mathcal{M}\sigma\right) \tag{25}$$

$$\beta_n^\varepsilon := \inf\{\beta_n(\mathcal{M}) : \alpha_n(\mathcal{M}) < \varepsilon\} \tag{26}$$

$$E(\mathbf{R}, \mathbf{S}) := \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{-\log \beta_n^\varepsilon}{n} \tag{27}$$

In Eqs. (24) and (25), we have $\mathcal{M} \in E(V)$ and in (26) there is an implicit dependence on R^n, S^n. Note that the limits of Eq. (23) (resp. Eq. (27)) may not exist, in which case we leave $D(\mathbf{R} \,\|\, \mathbf{S})$ (resp. $E(\mathbf{R}, \mathbf{S})$) undefined. See [7] for a discussion of replacing the $\lim$ with $\liminf$ or $\limsup$.

An important special case of Eq. (23) is the *regularized relative entropy of entanglement* [54], which is defined to be $\lim_{n\to\infty} \frac{1}{n} D(\rho^{\otimes n} \,\|\, \mathrm{Sep})$, and is normally denoted $E_R^\infty(\rho)$. In our notation this quantity is given by

$$D(\boldsymbol{\rho} \,\|\, \mathbf{Sep}). \tag{28}$$

An important result about composite quantum hypothesis testing is that error exponent $\min_{\rho \in R^1} D(\rho \| \sigma)$ can be achieved by a test that depends only on R_1 and σ [5], [27]. In terms of Definition 1, this can be expressed as

$$D(\mathbf{R} \,\|\, \mathbf{S}) = E(\mathbf{R}, \mathbf{S}), \tag{29}$$

whenever $\mathbf{R}, \mathbf{S}$ are of the form $R^n = \{\rho^{\otimes n} : \rho \in R^1\}$ and $S^n = \{\sigma^{\otimes n}\}$, for some set R^1 and some state σ. We call results of the form (29) "quantum Stein's Lemmas," because, like the classical Chernoff-Stein Lemma, they give an equality between a relative entropy and an error exponent for hypothesis testing.

A quantum Stein's Lemma has also been proven in the case when $\mathbf{R} = \boldsymbol{\rho}$ for a fixed state ρ and $\mathbf{S}$ is a family of sets. In this case, (29) is proved in [8] in the case where $\mathbf{S}$ is a *self-consistent* family of states, defined as follows.

Definition 2 ([8])**.** $\mathbf{S} = (S^1, S^2, \ldots)$ *is a self-consistent family of states if*

1) *Each* S^n *is convex and closed.*
2) *There exists a full-rank state* σ *such that each* S^n *contains* $\sigma^{\otimes n}$.
3) *For each* $\sigma \in S^n$, $\mathrm{tr}_n \sigma \in S^{n-1}$.
4) *If* $\sigma_n \in S^n, \sigma_m \in S^m$ *then* $\sigma_n \otimes \sigma_m \in S^{n+m}$.
5) S^n *is closed under permutation.*

Some important cases of self-consistent families of states are $\mathbf{Sep}$ (defined in Section III-A1), PPT (defined in Appendix A, although it will not be used in this paper) and $\boldsymbol{\sigma}$ for any full-rank state σ.

3) Hypothesis testing with restricted measurements: We now introduce the problem of quantum hypothesis testing with restricted measurements. The full set of [two-outcome] measurements on $V^{\otimes n}$ (i.e. $E_2(V^{\otimes n})$) consists of all $\{\mathcal{M}, I - \mathcal{M}\}$ where $0 \leq \mathcal{M} \leq I$. However, it is often useful to consider smaller classes of measurements, such as those that two parties can perform with local operations and classical communication (LOCC). When considering restricted classes of measurements, our objective might be to minimize the probability of error (subject to the usual tradeoff between type I and type II errors), or it might be to maximize the classical relative entropy of the output distributions. In the former case we will use measurements in $E_2(V)$ and in the latter we will use measurements in $E_{\mathbb{N}}(V)$.

Definition 3. *Let* $\mathbf{R} = (R^1, R^2, \ldots)$, $\mathbf{S} = (S^1, S^2, \ldots)$, *with* $R^n, S^n \subseteq \mathcal{D}(V^{\otimes n})$, *and* $\mathbf{M} = (M^1, M^2, \ldots)$, *with* $M^n \subseteq E_{\mathbb{N}}(V^{\otimes n})$. *Then the* asymptotic relative entropy *of* $\mathbf{R}$ *with respect to* $\mathbf{S}$ *under measurements* $\mathbf{M}$ *is*

$$D_{\mathbf{M}}(\mathbf{R} \,\|\, \mathbf{S}) := \lim_{n\to\infty} D_{M^n}(R^n \,\|\, S^n) \tag{30a}$$

$$D_{M^n}(R^n \,\|\, S^n) := \sup_{\mathcal{M}\in M^n} \inf_{\substack{\rho\in R^n \\ \sigma\in S^n}} \frac{D(\mathcal{M}(\rho) \,\|\, \mathcal{M}(\sigma))}{n}. \tag{30b}$$

For $\mathcal{M} \in E(V^{\otimes n})$, *we further define*

$$\alpha_n(\mathcal{M}) := \sup_{\rho\in R^n} \operatorname{tr}((I - \mathcal{M})\rho) \tag{31}$$

$$\beta_n(\mathcal{M}) := \sup_{\sigma\in S^n} \operatorname{tr}(\mathcal{M}\sigma). \tag{32}$$

Now we restrict M^n *to two-outcome measurements and use* $\mathcal{M}$ *as a shorthand for* $\{\mathcal{M}, I - \mathcal{M}\}$ *to define*

$$\beta_n^\varepsilon(\mathbf{M}) := \inf_{\mathcal{M}\in M^n \cap E_2(V^{\otimes n})} \{\beta_n(\mathcal{M}) : \alpha(\mathcal{M}) < \varepsilon\} \tag{33}$$

$$E_{\mathbf{M}}(\mathbf{R}, \mathbf{S}) := \lim_{\varepsilon\to 0} \lim_{n\to\infty} \frac{-\log \beta_n^\varepsilon}{n} \tag{34}$$

As before, the quantities (30) *and* (34) *are left undefined when the corresponding limit does not exist.*

Following our notation for families of states, we use boldface (e.g. $\mathbf{M}$) to denote families of measurements. In particular, we define $\mathrm{SEP}(A : B)$ to denote separable measurements on $A : B$ (i.e. M where every POVM element has the form $\sum_i X_i \otimes Y_i$ with $X_i, Y_i \geq 0$) and denote the corresponding family by

$$\mathbf{SEP}(A : B) = (\mathrm{SEP}(A : B), \mathrm{SEP}(A^{\otimes 2} : B^{\otimes 2}), \ldots).$$

Again we will often write SEP or **SEP** where the systems A, B are clear from context. Note that $\mathrm{Sep}(A : B)$ and $\mathrm{SEP}(A : B)$ both refer to sets of matrices that can be written as $\sum_i X_i \otimes Y_i$ with $X_i, Y_i \geq 0$; the difference is that Sep refers to density matrices (i.e. matrices with trace one) and SEP to measurements made up from POVM elements (i.e. matrices with operator norm ≤ 1).

Another important class of measurements is ALL^n, which is simply the set of all valid quantum measurements on n systems: i.e. $\mathsf{ALL}^n = E_{\mathbb{N}}(V^{\otimes n})$. The corresponding family is denoted ALL. Some useful structural facts about D_{ALL^n} are proved in [4].

One further definition we will need (following [48], but with different notation) is the idea of a *compatible pair*.

Definition 4. *If* $\mathbf{M}$ *is a collection of measurements and* $\mathbf{S}$ *is a collection of states, we say that* $(\mathbf{M}, \mathbf{S})$ *are a compatible pair if (a)* $\mathbf{S}$ *is closed under permutations of the systems and under convex combinations, and (b) applying a measurement in* $\mathbf{M}$ *to a state in* $\mathbf{S}$ *and conditioning on any outcome leaves a residual state*

that is still in $\mathbf{S}$*. More concretely for positive integers* n, k*, for* $\rho_{n+k} \in S^{n+k}$*, for* $\mathcal{M}^k = (\mathcal{M}^k_j)_{j=1,2,\ldots} \in M^k$*, and for* j *a positive integer, define*

$$\tilde{\omega}_n = \mathrm{tr}_{n+1,\ldots,n+k}[\rho_{n+k}(I_n \otimes \mathcal{M}^j_k)],$$

and (assuming that $\mathrm{tr}\,\tilde{\omega}_n \neq 0$*) we define*

$$\omega_n = \frac{\tilde{\omega}_n}{\mathrm{tr}\,\tilde{\omega}_n}.$$

(Here the permutation symmetry of $\mathbf{S}$ *means that we can assume for convenience that* $\mathcal{M}^k$ *acts on the last* k *systems.) If* $(\mathbf{M}, \mathbf{S})$ *is a compatible pair then for any choice of* n, k, j, ρ_{n+k}, M_k*, either* $\mathrm{tr}\,\tilde{\omega}_n = 0$ *or* $\omega_n \in S^n$*.*

The main example of compatible pair which motivates our work is $(\mathbf{SEP}, \mathbf{Sep})$. We could also consider $(\mathsf{LOCC}, \mathbf{Sep})$, or $(\mathbf{M}, \mathbf{Sep})$ where $\mathbf{M}$ is any other subset of $\mathbf{SEP}$. Compatible pairs also arise from resource theories, in which there is typically a family of free quantum operations and free quantum states, with the property that the free operations preserve the set of free quantum states. In some cases, these can be defined by starting with the set of operations (e.g. LOCC operations which yield the set of separable states) or the set of states (e.g. thermal states of some fixed Hamiltonians). We will be interested in a slightly different setting in which quantum operations are replaced by measurements. Besides $(\mathbf{M}, \mathbf{Sep})$ with $\mathbf{M} \subset \mathbf{Sep}$ other examples of compatible pairs are:

- *Symmetry constraints.* For each n, fix a group G_n of unitaries acting on $V^{\otimes n}$. These should be compatible in the sense that $G_n \otimes I \subseteq G_{n+1}$ and $\pi(G_n) = G_n$ for any permutation π of the n systems. If S^n is the set of all states that commute with G_n and M^n is any subset of the measurements that commute with G_n, then $\mathbf{S} = (S^n)_{n\geq 1}$ and $\mathbf{M} = (M^n)_{n\geq 1}$ are compatible pairs. This has been studied in the context of the resource theory of asymmetry [24].
- In quantum optics we can take $\mathbf{S}$ to be the convex hull of Gaussian quantum states and $\mathbf{M}$ the measurements that can be implemented with Gaussian quantum operations [37].
- Let $\mathbf{S}$ be the set of stabilizer states and $\mathbf{M}$ the set of Pauli measurements. The famous Gottesman-Knill theorem [23] includes the fact that $\mathbf{S}$ is closed under measurements from $\mathbf{M}$.

For each of these compatible pairs, if we consider $\mathbf{S}$ to be set of free states then the relative entropy $D(\boldsymbol{\rho} \,\|\, \mathbf{S})$ can be viewed as a cost of the state ρ, with a meaning made more precise in [33], [9].

We will need some more mild regularity conditions on the classes of measurements we consider.

Definition 5. $\mathbf{M} = (M^1, M^2, \ldots)$ *is a* self-consistent *family of measurements if*

- *For any* k, l *and any* $\mathcal{M}^k \in M^k, \mathcal{M}^l \in M^l$*, we have* $\mathcal{M}^k \otimes \mathcal{M}^l \in M^{k+l}$ *and* $\mathcal{M}^k \otimes I_l \in M^{k+l}$*.*
- M^n *is closed under permutations of the* n *systems.*
- M^n *is closed under finite labelled mixtures. In other words, if* $\{\mathcal{M}^{(i)}\}_i$ *are a collection of measurements in* M^n *where* $\mathcal{M}^{(i)}$ *has POVM elements* $\{\mathcal{M}^{(i)}_j\}_j$ *and* $\{p_i\}_i$ *is a probability distribution then the measurement with POVM elements* $\{p_i \mathcal{M}^{(i)}_j\}_{i,j}$ *is in* M^n*.*

This last condition on measurements needs a little more explanation. First, the measurement outcomes are labelled by pairs of integers, so we need to relax our definition of $E_{\mathbb{N}}(V)$ and allow measurements indexed by any finite set. Second, observe that the property of closure under finite labelled mixtures is implied by the following natural two conditions: (1) that M^n is convex, and (2) that M^n is closed under relabeling of outcomes, i.e. if $(\mathcal{M}_1, \mathcal{M}_2, \ldots) \in M^n$ and $\pi : \mathbb{N} \mapsto \mathbb{N}$ is an injective map then $(\mathcal{M}_{\pi(1)}, \mathcal{M}_{\pi(2)}, \ldots) \in M^n$. These in turn (along with the other self-consistency properties) are satisfied by all the examples of families of measurements mentioned in this paper.

Our main results (in Sections III-B and III-D) involve compatible pairs with self-consistent families of measurements, and we also discuss previously known results about compatible pairs in Section III-A5.

4) Relations between distinguishability measures: Finally, we state some known and new results that relate the different versions of $D, E, D_{\mathbf{M}}, E_{\mathbf{M}}$. The following statement is a consequence of the minimax theorem.

Lemma 13. *Let R, $S \subseteq D(V)$ be closed and convex, while $M \subset E_{\mathbb{N}}(V)$ is closed under finite labelled mixtures (as defined in Definition 5). Then*

$$\sup_{\mathcal{M}\in M} \min_{\substack{\rho\in R\\ \sigma\in S}} D(\mathcal{M}(\rho) \parallel \mathcal{M}(\sigma)) = \min_{\substack{\rho\in R\\ \sigma\in S}} \sup_{\mathcal{M}\in M} D(\mathcal{M}(\rho) \parallel \mathcal{M}(\sigma)) \tag{35}$$

Note that the LHS is trivially $\leq$ the RHS, and that the RHS is the form of restricted-measurement distinguishability introduced by Piani [48].

Our Lemma will rely on a minimax theorem that is similar to the minimax theorems of Kneser, Fan and Sion from the 1950s [51] but which needs to handle the possibility that the relative entropy can be infinite.

Lemma 14 (Thm 5.2 of [20])**.** *Let X be a compact and convex subset of a Hausdorff topological vector space and let Y be a convex subset of a linear space. Let $f : X \times Y \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous on X for fixed $y \in Y$, convex in x and concave in y. Then*

$$\sup_{y\in Y} \inf_{x\in X} f(x,y) = \inf_{x\in X} \sup_{y\in Y} f(x,y). \tag{36}$$

Proof of Lemma 13. We will take our set X to be $R \times S$ with an element x representing a pair of density matrices (ρ, σ). Let $\mathcal{P}(M)$ denote the set of probability distributions over M with countable support and define $Y = \mathcal{P}(M)$. We can now define

$$f((\rho,\sigma),\mu) := \mathbb{E}_{\mathcal{M}\sim\mu} D(\mathcal{M}(\rho) \parallel \mathcal{M}(\sigma)). \tag{37}$$

Clearly f is affine, and hence concave, in μ. For fixed $\mathcal{M}$ (and thus fixed μ), the relative entropy is known to be convex and lower semicontinuous [49], [18]. Thus we meet the conditions of Lemma 14. Note also that the lower semicontinuity of f and the compactness of $R \times S$ guarantees that the $\min$ is achieved. Lemma 14 then implies that

$$\min_{\substack{\rho\in R\\ \sigma\in S}} \sup_{\mu\in\mathcal{P}(M)} f((\rho,\sigma),\mu) \leq \sup_{\mu\in\mathcal{P}(M)} \min_{\substack{\rho\in R\\ \sigma\in S}} f((\rho,\sigma),\mu) \tag{38}$$

(In fact it establishes an equality but we write $\leq$ to emphasize the direction that we are trying to prove.)

Eq. (38) is close to what we want but has $\mathcal{P}(M)$ in place of M. Since $\mathcal{P}(M)$ includes distributions which assign probability 1 to a particular measurement, we have

$$\min_{\substack{\rho\in R\\ \sigma\in S}} \sup_{\mathcal{M}\in M} D(\mathcal{M}(\rho) \parallel \mathcal{M}(\sigma)) \leq \min_{\substack{\rho\in R\\ \sigma\in S}} \sup_{\mu\in\mathcal{P}(M)} f((\rho,\sigma),\mu). \tag{39}$$

Upper bounding the $\sup$ over $\mathcal{P}(M)$ in terms of a $\sup$ over $\mathcal{M}$ is less trivial, and will need to use the fact that M is closed under countable labeled mixtures. Fix ρ, σ, μ and suppose that μ assigns probability

p_i to $\mathcal{M}^{(i)}$ for $i = 1, 2, \ldots$. Let $\{\mathcal{M}_j^{(i)}\}_{j=1,2,\ldots}$ be the POVM elements of $\mathcal{M}^{(i)}$. Then we will define the measurement $\mathcal{M}$ with POVM elements $\{p_i \mathcal{M}_j^{(i)}\}_{i,j}$, and by our hypothesis, $\mathcal{M} \in M$. Then

$$f((\rho,\sigma),\mu) = \sum_i p_i D(\mathcal{M}^{(i)}(\rho) \,\|\, \mathcal{M}^{(i)}(\sigma)) \tag{40a}$$

$$= \sum_{i,j} p_i \operatorname{tr}[\mathcal{M}_j^{(i)}\rho](\log \operatorname{tr}[\mathcal{M}_j^{(i)}\rho] - \log \operatorname{tr}[\mathcal{M}_j^{(i)}\sigma]) \tag{40b}$$

$$= \sum_{i,j} \operatorname{tr}[p_i \mathcal{M}_j^{(i)}\rho](\log \operatorname{tr}[p_i \mathcal{M}_j^{(i)}\rho] - \log \operatorname{tr}[p_i \mathcal{M}_j^{(i)}\sigma]) \tag{40c}$$

$$= D(\mathcal{M}(\rho) \,\|\, \mathcal{M}(\sigma)) \tag{40d}$$

We can take the minimum over ρ, σ to obtain

$$\min_{\substack{\rho\in R\\ \sigma\in S}} f((\rho,\sigma),\mu) \leq \min_{\substack{\rho\in R\\ \sigma\in S}} D(\mathcal{M}(\rho) \,\|\, \mathcal{M}(\sigma)), \tag{41}$$

where $\mathcal{M}$ depends on μ. Next we can take the sup over μ to obtain

$$\sup_{\mu\in\mathcal{P}(M)} \min_{\substack{\rho\in R\\ \sigma\in S}} f((\rho,\sigma),\mu) \leq \sup_{\mathcal{M}\in M} \min_{\substack{\rho\in R\\ \sigma\in S}} D(\mathcal{M}(\rho) \,\|\, \mathcal{M}(\sigma)). \tag{42}$$

Finally combining the inequalities (39), (38) and (42) implies the proof of the lemma.

□

We remark that some versions of the minimax theorem (i.e. Thm 4.2 of [51]) require only a weaker form of concavity in which for any $p \in [0,1]$ and any $x \in X, y_1, y_2 \in Y$, there exist $y_0 \in Y$ such that $f(x,y_0) \geq pf(x,y_1) + (1-p)f(x,y_2)$. In other words, y_0 does not have to be $py_1 + (1-p)y_2$ but could be an arbitrary point and indeed Y does not even have to be a linear space. This would perfectly fit our approach of taking labelled mixtures of measurements. However, since our theorem needs to handle the possibility that $D(\cdot \,\|\, \cdot) = \infty$, we cannot directly use Thm 4.2 of [51].

Known facts: The following relations between the quantities have been derived previously.

$$E(\boldsymbol{\rho},\boldsymbol{\sigma}) = D(\rho \,\|\, \sigma) \qquad \text{quantum Stein's Lemma [31]} \tag{43}$$

$$D(\{\rho\} \,\|\, S^1) \geq D(\boldsymbol{\rho} \,\|\, \mathbf{S}) \qquad \text{for } \mathbf{S} \text{ satisfying property (4) of Definition 2} \tag{44}$$

$$D(\mathbf{R} \,\|\, \mathbf{S}) \geq D_{\mathbf{M}}(\mathbf{R} \,\|\, \mathbf{S}) \qquad \text{from monotonicity of relative entropy} \tag{45}$$

$$E(\boldsymbol{\rho},\mathbf{S}) = D(\boldsymbol{\rho} \,\|\, \mathbf{S}) \qquad \text{for } \mathbf{S} \text{ a self-consistent family (Definition 2)[8]} \tag{46}$$

We can, in fact, relate D_{ALL}, D, E for any ρ and any closed convex $\mathbf{S}$ using

$$D_{\mathsf{ALL}}(\boldsymbol{\rho} \,\|\, \mathbf{S}) \overset{(68)}{\geq} E(\boldsymbol{\rho},\mathbf{S}) \overset{(46)}{=} D(\boldsymbol{\rho} \,\|\, \mathbf{S}) \overset{(45)}{\geq} D_{\mathsf{ALL}}(\boldsymbol{\rho} \,\|\, \mathbf{S}) \tag{47}$$

The main goal of the second half of this paper is to extend these results as far as possible to $D_{\mathbf{M}}$ and $E_{\mathbf{M}}$.

5) Superadditivity: When we consider families of states and measurements, it is not *a priori* clear whether the distinguishability per system should increase or decrease with the number of systems. We say that a quantity $f(\rho)$ is *subadditive* if $f(\rho_{XY}) \leq f(\rho_X) + f(\rho_Y)$ (e.g., entropy) and *superadditive* if $f(\rho_{XY}) \geq f(\rho_X) + f(\rho_Y)$ (e.g., most entanglement measures). A function f is weakly subadditive $f(\rho^{\otimes n}) \leq nf(\rho)$ and is weakly superadditive if $f(\rho^{\otimes n}) \geq nf(\rho)$). If a function is both (weakly) subadditive and superadditive then we say it is (weakly) *additive.*

One of the main results known so far about relative entropy with restricted measurements is due to Piani [48], who used these measures to prove a superadditivity inequality:

$$D(\rho_{XY} \,\|\, S^2) \geq D_{\mathbf{M}}(\rho_X \,\|\, S^1) + D(\rho_Y \,\|\, S^1) \qquad \text{for compatible } (\mathbf{M}, S) \text{ [48]} \tag{48}$$

$$D(\boldsymbol{\rho} \,\|\, \mathbf{S}) \geq D_{\mathbf{M}}(\rho \,\|\, S^1) \qquad \text{as a corollary of (48) [48]} \tag{49}$$

In fact, Piani's result can easily be improved to show that $D_{\mathbf{M}}(\mathbf{R} \,\|\, \mathbf{S})$ is superadditive whenever $(\mathbf{M}, \mathbf{R})$ and $(\mathbf{M}, \mathbf{S})$ are compatible pairs, or in fact when $\mathbf{R}$ satisfies a milder condition.

Lemma 15. *Let* $(\mathbf{M}, \mathbf{S})$ *be a compatible pair with* $\mathbf{M}$ *a self-consistent family. Let* $\mathbf{R}$ *be a family of states that is closed under partial trace, i.e. satisfying* $\operatorname{tr}_j R^n \subseteq R^{n-1}$ *for each* $1 \leq j \leq n$*. Then for all* $\rho_{XY} \in \mathcal{D}(V^{\otimes k} \otimes V^{\otimes l})$*, if we identify* X *with* $V^{\otimes k}$ *and* Y *with* $V^{\otimes l}$*, we have*

$$D_{M^{k+l}}(\rho_{XY} \,\|\, S^{k+l}) \geq D_{M^k}(\rho_X \,\|\, S^k) + D_{M^l}(\rho_Y \,\|\, S^l). \tag{50}$$

Moreover,

$$D_{\mathbf{M}}(\mathbf{R} \,\|\, \mathbf{S}) = \lim_{n\to\infty} \frac{1}{n} D_{M_n}(R^n \,\|\, S^n) = \sup_n \frac{1}{n} D_{M_n}(R^n \,\|\, S^n). \tag{51}$$

Proof. The argument is a direct adaptation of the proof of Theorem 1 in [48].

Let $\mathcal{M}^X \in M^k, \mathcal{M}^Y \in M^l$ be arbitrary. Define an orthonormal basis $|1\rangle, |2\rangle, \ldots$ corresponding to the outcomes $1, 2, \ldots$ of $\mathcal{M}^X$. Define $p_i(\rho_X) = \operatorname{tr}(\mathcal{M}_i^X \rho_X) = (\mathcal{M}^X(\rho))_i$ and $\rho_Y^i = \operatorname{tr}_X[(\mathcal{M}_i^X \otimes I_Y)\rho_{XY}]/p_i(\rho_X)$. Choose $\sigma_{XY} \in S^{k+l}$ and define $p_i(\sigma_X)$ and σ_Y^i analogously. Then

$$\begin{aligned}
&D\left((\mathcal{M}^X \otimes \mathcal{M}^Y)(\rho_{XY}) \,\|\, (\mathcal{M}^X \otimes \mathcal{M}^Y)(\sigma_{XY})\right) \\
&= D\left(\sum_{i\geq 1} p_i(\rho_X) |i\rangle\langle i| \otimes \mathcal{M}^Y(\rho_Y^i) \,\|\, \sum_{i\geq 1} p_i(\sigma_X) |i\rangle\langle i| \otimes \mathcal{M}^Y(\sigma_Y^i)\right) && \text{(52a)} \\
&= D\left(\mathcal{M}^X(\rho_X) \,\|\, \mathcal{M}^X(\sigma_X)\right) + \sum_{i\geq 1} p_i(\rho_X) D\left(\mathcal{M}^Y(\rho_Y^i) \,\|\, \mathcal{M}^Y(\sigma_Y^i)\right) && \text{(52b)} \\
&\geq D\left(\mathcal{M}^X(\rho_X) \,\|\, \mathcal{M}^X(\sigma_X)\right) + D\left(\sum_{i\geq 1} p_i(\rho_X) \mathcal{M}^Y(\rho_Y^i) \,\|\, \sum_{i\geq 1} p_i(\rho_X) \mathcal{M}^Y(\sigma_Y^i)\right) && \text{(52c)} \\
&= D\left(\mathcal{M}^X(\rho_X) \,\|\, \mathcal{M}^X(\sigma_X)\right) + D\left(\mathcal{M}^Y(\rho_Y) \,\|\, \mathcal{M}^Y\left(\sum_{i\geq 1} p_i(\rho_X) \sigma_Y^i\right)\right), && \text{(52d)} \\
&\geq \inf_{\tilde{\sigma}_X \in S^k} D\left(\mathcal{M}^X(\rho_X) \,\|\, \mathcal{M}^X(\tilde{\sigma}_X)\right) + \inf_{\tilde{\sigma}_Y \in S^l} D\left(\mathcal{M}^Y(\rho_Y) \,\|\, \mathcal{M}^Y(\tilde{\sigma}_Y)\right) && \text{(52e)}
\end{aligned}$$

where (52a) follows from Proposition 1 of [48], (52b) from direct calculation, (52c) from joint convexity of relative entropy, and (52d) from linearity of the measurement. (In Piani's proof in [48] the analogues of the third and fourth lines were Lemma 1 and Property 2 of Proposition 1 respectively.)

We now take the infimum over σ_{XY} and then the supremum over $\mathcal{M}^X, \mathcal{M}^Y$, yielding (50). If we instead take the infimum over σ_{XY} and ρ_{XY} before taking the supremum over $\mathcal{M}^X, \mathcal{M}^Y$, then we find

that $D_{M^{k+l}}(R^{k+l} \,\|\, S^{k+l}) \geq D_{M^k}(R^k \,\|\, S^k) + D_{M^l}(R^l \,\|\, S^l)$. In other words, if $f(n) = D_{M_n}(R^n \,\|\, S^n)$ then f is superadditive (i.e. $f(k+l) \geq f(k) + f(l)$). This implies (51).

□

The preceding lemma says that $D_{\mathbf{M}}(\rho \,\|\, \mathbf{S})$ is a superadditive function of ρ for compatible pairs $(\mathbf{M}, \mathbf{S})$. The compatibility requirement here is essential. The pair $(\mathsf{ALL}, \mathbf{Sep})$ is not compatible, and $D(\cdot \,\|\, \mathbf{Sep})$ is known to be strictly subadditive (i.e. not superadditive) in some cases [56]. This does not directly yield an example of strict subadditivity for $D_{\mathsf{ALL}}(\cdot \,\|\, \mathbf{Sep})$ but can be modified to do so. The example in [56] is the antisymmetric Werner state $\rho = \frac{I - \mathrm{SWAP}}{d(d-1)} \in \mathcal{D}(\mathbb{C}^d \otimes \mathbb{C}^d)$. In [56], it is proved that

$$D(\rho \| \operatorname{Sep}(\mathbb{C}^d : \mathbb{C}^d)) = 1 \qquad \text{and} \qquad D(\rho \otimes \rho \| \operatorname{Sep}(\mathbb{C}^{d^2} : \mathbb{C}^{d^2})) = 1 + O(1/d), \tag{53}$$

showing that $D(\cdot \,\|\, \mathbf{Sep})$ can be strictly subadditive. Observe that if we measure ρ with the two outcome measurement $\{\frac{I \pm \mathrm{SWAP}}{2}\}$ and label the outcomes +/- then we will always obtain the outcome - while for any $\sigma \in \mathrm{Sep}$ we have $\Pr[-] \leq 1/2$. Thus $D_{\mathsf{ALL}}(\rho \| \operatorname{Sep}(\mathbb{C}^d : \mathbb{C}^d)) \geq 1$ (and in fact equality holds). On the other hand, monotonicity of relative entropy implies that

$$D_{\mathsf{ALL}}(\rho \otimes \rho \| \operatorname{Sep}(\mathbb{C}^{d^2} : \mathbb{C}^{d^2})) \leq D(\rho \otimes \rho \| \operatorname{Sep}(\mathbb{C}^{d^2} : \mathbb{C}^{d^2})) = 1 + O(1/d). \tag{54}$$

Thus we have an example where $D_{\mathsf{ALL}}(\cdot \| \mathbf{Sep})$ is strictly subadditive.

On the other hand, $D_{\mathbf{M}}(\cdot \,\|\, \mathbf{S})$ can be strictly superadditive (i.e., not subadditive). Let us consider the simple situation in which $R^n = \{\rho^{\otimes n}\}$ and $S^n = \{\sigma^{\otimes n}\}$. It is a consequence of the quantum Stein's Lemma (22) (see also [27]) that

$$D(\rho \,\|\, \sigma) = \lim_{n \to \infty} \frac{1}{n} D_{\mathsf{ALL}}(\rho^{\otimes n} \,\|\, \sigma^{\otimes n}).$$

Thus, any example in which

$$\max_{M \in \mathsf{ALL}} D(M(\rho) \,\|\, M(\sigma)) < D(\rho \,\|\, \sigma) \tag{55}$$

will yield an example in which $D_{\mathbf{M}}(\cdot \,\|\, \mathbf{S})$ is strictly superadditive. In fact, Proposition 5 of [4] (building upon Lemma 1 of [47]) states that (55) holds whenever $D(\rho \,\|\, \sigma)$ is finite and $\rho\sigma \neq \sigma\rho$. Thus superadditivity is a generic property of $D_{\mathbf{M}}(\cdot \,\|\, \cdot)$.

B. A quantum Stein's Lemma for restricted measurements

Theorem 16 (Quantum Stein's Lemma for restricted measurements)**.** *For any compatible pairs* $(\mathbf{M}, \mathbf{R})$ *and* $(\mathbf{M}, \mathbf{S})$ *with* $\mathbf{M}$ *a self-consistent family and* $\mathbf{R}, \mathbf{S}$ *closed,*

$$D_{\mathbf{M}}(\mathbf{R} \,\|\, \mathbf{S}) = E_{\mathbf{M}}(\mathbf{R}, \mathbf{S})\,. \tag{56}$$

Proof. For any positive integer k, suppose $0 \leq E_k < \frac{1}{k} D_{M^k}(R^k \,\|\, S^k)$. (If $D_{M^k}(R^k \,\|\, S^k) = \infty$ then this means that E_k is an arbitrary nonnegative number.) The supremum over measurements in the definition of $D_{M^k}(R^k \,\|\, S^k)$ means that there exists $\mathcal{M}^k \in M^k$ such that

$$\frac{1}{k} D(\mathcal{M}^k(R^k) \,\|\, \mathcal{M}^k(S^k)) > E_k.$$

Define $P := \mathcal{M}^k(R^k)$ and $Q := \mathcal{M}^k(S^k)$. Then

$$\frac{1}{k} D(p \,\|\, q) > E_k \qquad \forall p \in P, q \in Q. \tag{57}$$

Given a state $\rho \in \mathcal{D}(V^{\otimes nk})$, we apply $\mathcal{M}^k$ to each block of k systems, obtaining outcomes $x_1, \ldots, x_n$. Then since $(\mathbf{M}, \mathbf{R})$ and $(\mathbf{M}, \mathbf{S})$ are compatible pairs, the distribution of each x_i, conditioned on any

possible value of $x_1, \ldots, x_{i-1}$, is an element of P (if $\rho \in R^{nk}$) or Q (if $\rho \in S^{nk}$). Thus, according to Theorem 2, there is a sequence of acceptance regions that achieves the rate E_k. Thus for any $\varepsilon \in (0,1)$,

$$\liminf_{n\to\infty} -\frac{1}{nk} \log \beta^\varepsilon_{nk} \geq E_k \tag{58}$$

Given a state in R^{nk+l} or S^{nk+l} for $l < k$ we can discard l systems and obtain a state in in R^{nk} or S^{nk}, using the fact that $\mathbf{R}, \mathbf{S}$ are closed under partial trace. Thus we can drop the k-dependence from the LHS of (58) to obtain

$$\liminf_{n\to\infty} -\frac{1}{n} \log \beta^\varepsilon_{n} \geq E_k. \tag{59}$$

Since this holds for any k, we can take the $\limsup$ over k to find

$$\liminf_{n\to\infty} -\frac{1}{n} \log \beta^\varepsilon_{n} \geq \limsup_{k\to\infty} \frac{1}{k} D_{M^k}(R^k \,\|\, S^k) = D_{\mathbf{M}}(\mathbf{R} \,\|\, \mathbf{S}). \tag{60}$$

This last equality is due to Lemma 15.

The reverse inequality can be obtained by the following standard argument which we adapt from [31]. See also footnote 11 of [11] where roughly the same result was stated and attributed to [31], [46]. We include a proof here for completeness and because previous work did not technically show the same results.

For a positive integer n and $\varepsilon > 0$, let $\mathcal{M} := (\mathcal{M}, I - \mathcal{M}) \in M_n$ be a measurement such that $\mathrm{tr}[\mathcal{M}\rho] > 1 - \varepsilon$. Then for any $\rho \in R_n, \sigma \in S_n$,

$$\begin{aligned}
\sup_{\mathcal{M}' \in M_n} D(\mathcal{M}'(\rho) \,\|\, \mathcal{M}'(\sigma)) \geq & D(\mathcal{M}(\rho) \| \mathcal{M}(\sigma)) && (61)\\
= & \,\mathrm{tr}[\mathcal{M}\rho] \log \mathrm{tr}[\mathcal{M}\rho] - \mathrm{tr}[\mathcal{M}\rho] \log \mathrm{tr}[\mathcal{M}\sigma] && (62)\\
& + \mathrm{tr}[(I-\mathcal{M})\rho] \log \mathrm{tr}[(I-\mathcal{M})\rho] - \mathrm{tr}[(I-\mathcal{M})\rho] \log \mathrm{tr}[(I-\mathcal{M})\sigma] && (63)\\
\geq & - h_2(\mathrm{tr}[\mathcal{M}\rho]) - \mathrm{tr}[\mathcal{M}\rho] \log \mathrm{tr}[\mathcal{M}\sigma] && (64)\\
\geq & - 1 - (1-\varepsilon) \log \mathrm{tr}[\mathcal{M}\sigma] && (65)
\end{aligned}$$

Here we define $h_2(p) = -\log(p) - \log(1-p)$ and take log to be base 2. Rearranging yields

$$-\log \mathrm{tr}[\mathcal{M}\sigma] \leq \frac{1 + \sup_{\mathcal{M}' \in M_n} D(\mathcal{M}'(\rho) \,\|\, \mathcal{M}'(\sigma))}{1-\varepsilon}. \tag{66}$$

To relate this to $\beta^\varepsilon_n(\mathbf{M})$ we take the $\inf$ over $\rho \in R_n, \sigma \in S_n$ and then the $\sup$ over $\mathcal{M} \in M^n \cap E_2(V^{\otimes n})$ satisfying $\alpha_n(\mathcal{M}) < \varepsilon$. This implies that

$$-\beta^\varepsilon_n(\mathbf{M}) \leq \frac{1 + \inf_{\rho \in R_n, \sigma \in S_n} \sup_{\mathcal{M}' \in M_n} D(\mathcal{M}'(\rho) \,\|\, \mathcal{M}'(\sigma))}{1-\varepsilon}. \tag{67}$$

We can now use Lemma 13 to exchange the $\inf$ and $\sup$. Finally we can divide by n and take the $\limsup$ in n to obtain

$$\limsup_{n\to\infty} -\frac{1}{n} \beta^\varepsilon_n \leq \frac{1}{1-\varepsilon} \limsup_{n\to\infty} \frac{1}{n} D_{M_n}(R_n \,\|\, S_n) \tag{68}$$

Combining (60) and (68) and taking $\varepsilon \to 0$ we finally establish that

$$E_{\mathbf{M}}(\mathbf{R}, \mathbf{S}) = D_{\mathbf{M}}(\mathbf{R} \,\|\, \mathbf{S}). \tag{69}$$

□

This is analogous to the result in [8], which established $E(\boldsymbol{\rho}, \mathbf{S}) = D_{\mathsf{ALL}}(\boldsymbol{\rho} \,\|\, \mathbf{S})$ for self-consistent sets of states $\mathbf{S}$, but incomparable because in general $(\mathsf{ALL}, \mathbf{S})$ will not be a compatible pair.

While this shows that the optimal hypothesis testing rate for this restricted-measurement setting does indeed reduce to a relative entropy, it may be difficult to compute $D_{\mathbf{M}}$ because of the regularization (i.e. $\lim_{n\to\infty}$) and optimization over measurements in (30). However, in some special cases, it is known how to carry out this optimization; e.g. [30] computes the relative entropy of a pure entangled state with respect to the maximally mixed state under various restricted classes of measurements.

C. *Stronger Subadditivity of Quantum Entropy*

We now present an application of Theorem 16 to a strengthening of the celebrated strong subadditivity inequality of Lieb and Ruskai for the quantum entropy [41], which can be written as

$$I(A : B \mid C)_\rho \geq 0 \tag{70}$$

where

$$\begin{aligned} I(A : B \mid C)_\rho &:= H(AC)_\rho + H(BC)_\rho - H(ABC)_\rho - H(C)_\rho \\ &:= H(\rho_{AC}) + H(\rho_{BC}) - H(\rho_{ABC}) - H(\rho_C) \end{aligned}$$

denotes the conditional mutual information of a state ρ_{ABC}. In what follows we will often omit the subscript ρ when the state is understood. See Appendix A for additional discussion.

In [11], the following lower bound was shown for any state ρ_{ABC}:

$$I(A : B \mid C) \geq D_{\mathsf{ALL}}(\boldsymbol{\rho}_{ABC} \,\|\, \mathbf{Sep}(A : BC)) - D_{\mathsf{ALL}}(\boldsymbol{\rho}_{AC} \,\|\, \mathbf{Sep}(A : C)) \tag{71}$$

Moreover the following inequality was shown

$$D_{\mathsf{ALL}}(\boldsymbol{\rho}_{ABC} \,\|\, \mathbf{Sep}(A : BC)) - D_{\mathsf{ALL}}(\boldsymbol{\rho}_{AC} \,\|\, \mathbf{Sep}(A : C)) \geq E_{\text{1-LOCC}}(\boldsymbol{\rho}, \mathbf{Sep}(A : B)), \tag{72}$$

with 1-LOCC the class of all measurements that can be implemented by quantum local operations and classical communication from Bob to Alice (see Appendix A for the precise definition). This implies that the conditional mutual information is lower bounded by $E_{\text{1-LOCC}}(\boldsymbol{\rho}, \mathbf{Sep}(A : B))$. (Ref. [11] actually stated a weaker result in terms of the 1-LOCC (trace) distance, but their proof essentially contains (72) as an intermediate step. In reading [11], [40] beware that they use the symbols D and E with meanings reversed from our conventions.)

In [40] the following apparent strengthening of (72) was obtained:

$$D_{\mathsf{ALL}}(\boldsymbol{\rho}_{ABC} \,\|\, \mathbf{Sep}(A : BC)) \geq D_{\mathsf{ALL}}(\boldsymbol{\rho}_{AC} \,\|\, \mathbf{Sep}(A : C)) + D_{\text{1-LOCC}}(\boldsymbol{\rho}_{AB} \,\|\, \mathbf{Sep}(A : B)), \tag{73}$$

which implies

$$I(A : B \mid C) \geq D_{\text{1-LOCC}}(\boldsymbol{\rho}_{AB} \,\|\, \mathbf{Sep}(A : B)). \tag{74}$$

At the time of [40] it was known only that $D_{\text{1-LOCC}} \geq E_{\text{1-LOCC}}$ (see discussion in the proof of Theorem 16) and so (73) was believed to be stronger than (72). Theorem 16 shows that (73) is equivalent to (72) and so it can be used in conjunction with [11] to give an alternative proof of (74). This possibility was already discussed in [11]; see the discussion surrounding Eq. (43) of that paper.

D. Symmetric hypothesis testing with restricted measurements

Our main result on symmetric hypothesis testing against an adaptive adversary (Theorem 7) makes it natural to conjecture a corresponding result for symmetric quantum hypothesis testing. For quantum states ρ, σ, define

$$\Gamma^*(\rho,\sigma) := \max_{0\leq\lambda\leq 1} \Gamma^\lambda(\rho,\sigma) := \max_{0\leq\lambda\leq 1} -\log \operatorname{tr}(\rho^\lambda \sigma^{1-\lambda}) \tag{75}$$

$$\Gamma^*_{\mathbf{M}}(\mathbf{R},\mathbf{S}) := \lim_{n\to\infty} \sup_{\mathcal{M}\in M^n} \inf_{\substack{\rho\in R^n \\ \sigma\in S^n}} \frac{\Gamma^*(\mathcal{M}(\rho) \,\|\, \mathcal{M}(\sigma))}{n} \tag{76}$$

$$\gamma_{\mathbf{M}}(\mathbf{R},\mathbf{S}) := \lim_{n\to\infty} \sup_{\mathcal{M}\in M^n} \inf_{\substack{\rho\in R^n \\ \sigma\in S^n}} -\frac{1}{n}\log \operatorname{tr}(\mathcal{M}\sigma + (I-\mathcal{M})\rho) \tag{77}$$

A quantum analogue of Chernoff's Theorem was proven in [45], [1] and in our notation can be expressed as

$$\gamma_{\mathsf{ALL}}(\boldsymbol{\rho},\boldsymbol{\sigma}) = \Gamma^*(\rho,\sigma).$$

With restricted measurements, we might ask whether an analogue of Theorem 16 holds.

Conjecture 17. *If* $(\mathbf{M},\mathbf{R})$ *and* $(\mathbf{M},\mathbf{S})$ *are compatible pairs, then*

$$\gamma_{\mathbf{M}}(\mathbf{R},\mathbf{S}) = \Gamma^*_{\mathbf{M}}(\mathbf{R},\mathbf{S}).$$

A plausible route to proving the conjecture is to use the strategy of the proof of Theorem 16, replacing the adversarial Chernoff-Stein Lemma with the adversarial Chernoff's Theorem (Theorem 7)). However, there are several limits and sup/inf steps and we have not verified that these compose in the required ways.

E. Open questions

Having established a quantum Stein's Lemma for restricted measurements, we would like to know if a strong converse can also be proven, or more generally if we can calculate the error exponent for the type-2 error when the type-1 error is required to be $< \varepsilon$ for some fixed $\varepsilon \in (0,1)$. The difficulty is that $D_{\mathbf{M}}(\cdot \,\|\, \mathbf{S}) > D_{M^1}(\cdot \,\|\, S^1)$ in general, and we would need to control the rate of convergence as a function of n in the lim used to define $D_{\mathbf{M}}(\cdot \,\|\, \mathbf{S})$.

Like many information-theoretic quantities, $D(\boldsymbol{\rho} \,\|\, \mathbf{Sep})$ and $D_{\mathbf{M}}(\boldsymbol{\rho} \,\|\, \mathbf{Sep})$ (for various natural choices of $\mathbf{M}$) are operationally interesting, but are hard in practice to compute. We would like to know the complexity of estimating them (which is a variant of the usual question about the hardness of testing separability, cf. [25], [12]) and whether good relaxations exist (cf. [3]).

Finally, a major application of restricted-measurement distinguishability is to the related questions of k-extendable states[2], tripartite states with low conditional mutual information (i.e. "approximate Markov states", cf. [34]), and the quality of approximations achieved by the sum-of-squares hierarchy (cf. [2]). A few of the more prominent open questions here are:

- If $I(A:B \mid E)_\rho$ is small then it was recently discovered [22], [53] that an "approximate recovery" map $T : E \to E \otimes B$ exists such that $(\mathrm{id} \otimes T)\rho_{AE} \approx \rho_{ABE}$ in the sense of (among other measures) the measured relative entropy, i.e.

$$D_{\mathsf{ALL}}(\rho_{ABE} \,\|\, (\mathrm{id}\otimes T)\rho_{AE}) \leq I(A:B \mid E)_\rho. \tag{78}$$

[2] A bipartite state ρ_{AB} is said to be k-extendable if there exists a state $\tilde{\rho}_{AB_1\ldots B_k}$ such that $\tilde{\rho}_{AB_i} = \rho_{AB}$ for each i. The idea of k-extendability was introduced in [50], [17], where it was proved that for any fixed dimension of A and/or B, the set of k-extendable states approaches the set of separable states. However, the rate of convergence is an open question.

Ref. [21] found that we cannot replace the $D_{\mathsf{ALL}}(\cdot \,\|\, \cdot)$ on the LHS with the ordinary quantum relative entropy $D(\cdot \,\|\, \cdot)$. Their result leaves open the question of what relationship beween $\min_T D(\rho_{ABE} \,\|\, (\mathrm{id} \otimes T)\rho_{AE})$ and $I(A : B \mid E)$ is possible. Can we simply multiply $I(A : B \mid E)$ by some constant, or can these quantities differ by an amount that grows with dimensions? We do not even know whether the ratio between these quantities could be arbitrarily large in fixed dimensions.

- How large can $D_{\mathbf{M}}(\rho \,\|\, \mathrm{Sep})$ be when ρ is k-extendable and $\mathbf{M}$ is the class of separable measurements? Sharp bounds are known [12] when $\mathbf{M} = 1\text{-LOCC}$, and if they could be extended to separable measurements it would have implications for quantum Merlin-Arthur games with multiple Merlins [25] as well as for classical optimization algorithms.
- The ability of semidefinite programming hierarchies to estimate small-set expansion can be understood in terms of a restricted-measurement distinguishability problem [2]. A major open question is whether small-set expansion on graphs of size n can be well-approximated by $O(\log n)$ levels of these hierarchies, which would imply a quasipolynomial-time algorithm for the problem. Can tools from quantum information shed further light here?

APPENDIX

This appendix contains a very brief review of the quantum formalism and notation used in this paper. For a much more detailed introduction to quantum information theory, see [57], or for an overview of the field of quantum computing and quantum information more generally see [44], [36].

Density matrices. The quantum analogue of a probability distribution over $[d] = \{1, \ldots, d\}$ is called a *density matrix*, or simply a *state*. Density matrices must be positive semi-definite and have trace one. These conditions are analogous to the requirement that probabilities must be nonnegative and normalized; indeed diagonal density matrices correspond exactly to probability distributions. If A is a finite-dimensional Hilbert space, then define $\mathcal{D}(A)$ to be the set of density matrices on A, meaning the set of operators on A that are positive semi-definite and have trace one. Let $\mathcal{L}(A, B)$ denote the set of bounded linear operators from A to B, and let $\mathcal{L}(A) := \mathcal{L}(A, A)$.

Tensor product. To describe composite quantum systems, we use the tensor product. The tensor product of a vector $x \in \mathbb{C}^{d_1}$ and a vector $y \in \mathbb{C}^{d_2}$ is denoted $x \otimes y$ and has entries that run over all $x_{i_1} y_{i_2}$ for $i_1 \in [d_1], i_2 \in [d_2]$. Similarly, if X and Y are matrices, then their tensor product $X \otimes Y$ has matrix elements $(X \otimes Y)_{(i_1,i_2),(j_1,j_2)} = X_{i_1,j_1} Y_{i_2,j_2}$. For vector spaces A, B, we let $A \otimes B$ denote the span of $\{a \otimes b : a \in A, b \in B\}$. Note that $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2} \cong \mathbb{C}^{d_1 d_2}$. Finally, in each case we use the tensor power notation $X^{\otimes n}$ to stand for

$$\overbrace{X \otimes X \otimes \cdots \otimes X}^{n \text{ times}}.$$

Product and separable states. The tensor product is used to combine quantum states in the same way that independent classical probability distributions are combined to form a joint distribution. Indeed, if p, q are probability distributions of independent random variables, then $p \otimes q$ denotes the joint distribution. Similarly, if ρ and σ are density matrices, then $\rho \otimes \sigma$ denotes the state of a system that is in a so-called *product state*. The convex hull of the set of product states is called the set of *separable states*. We write $\mathrm{Sep}(A : B)$ to indicate the split along which we demand that the states be separable, e.g.

$$\mathrm{Sep}(A : B) = \mathrm{conv}\{\alpha \otimes \beta : \alpha \in \mathcal{D}(A), \beta \in \mathcal{D}(\beta)\}. \tag{79}$$

Although the set $\mathrm{Sep}(A : B)$ is convex, it is not easy to work with. For example, computational hardness results are known for the weak membership problem. Instead, it is sometimes more convenient to consider

the relaxation PPT, which denotes the set of states with Positive Partial Transpose. The partial transpose operator Γ (meant to resemble the right half of the T that usually denotes transpose) acts linearly on $\mathcal{L}(A \otimes B)$ by mapping $X \otimes Y$ to $X \otimes Y^T$; equivalently we can write it as $\mathrm{id}_A \otimes T_B$, where id_A is the identity operator on $\mathcal{L}(A)$ and T_B is the transpose operator on $\mathcal{L}(B)$. We define $\mathrm{PPT}(A : B) = \{\rho \in \mathcal{D}(A \otimes B) : \rho^\Gamma \in \mathcal{D}(A : B)\}$. This set is easier to work with because it has a semidefinite-programming characterization. Moreover, it is straightforward to show that $\mathrm{Sep}(A : B) \subset \mathrm{PPT}(A : B)$. However, in general this inclusion is strict, and as the dimensions of A, B grow large, PPT can be an arbitrarily bad approximation for Sep [3].

Partial trace. Another concept from probability theory that we will need to generalize is the idea of a marginal distribution. Say we have a density matrix $\rho_{AB} \in \mathcal{D}(A \otimes B)$. The subscript emphasizes the systems which ρ describes, which are analogous to the random variables corresponding to a probability distribution. To obtain the state on only the A system, we apply the *partial trace* operator $\mathrm{tr}_B := \mathrm{id}_A \otimes \mathrm{tr}_B$ to ρ_{AB}. The action of the partial trace is often denoted by writing only the subscripts, as in

$$\rho_A := \mathrm{tr}_B \rho_{AB} \qquad \text{and} \qquad \rho_B := \mathrm{tr}_A \rho_{AB}. \tag{80}$$

(This notation generalizes; e.g. if $\rho \in \mathcal{D}(A \otimes B \otimes C)$, then $\rho_B = \mathrm{tr}_{AC} \rho_{ABC} = \mathrm{tr}_A \mathrm{tr}_C \rho_{ABC}$, etc.) Concretely, $(\rho_A)_{i,i'} = \sum_j (\rho_{AB})_{(i,j),(i',j)}$ and $(\rho_B)_{j,j'} = \sum_i (\rho_{AB})_{(i,j),(i,j')}$. We see that if ρ is diagonal then this coincides with the idea of a marginal distribution from classical probability theory.

Measurements. Although technically all of physics is described by quantum mechanics, it is often convenient to make a distinction between quantum information, which is often carried in very small systems such as single atoms or single photons, and classical information, which is carried in macroscopic systems, such as a bit in a classical RAM. The bridge from quantum state to probability distribution is given by a *measurement* (also sometimes called a POVM, for Positive-Operator-Valued Measure), which formally is a collection of matrices (POVM elements) $\mathcal{M} = (\mathcal{M}_1, \ldots, \mathcal{M}_k)$ satisfying $\mathcal{M}_i \geq 0$ for each i (meaning each $\mathcal{M}_i$ is positive semi-definite) and $\mathcal{M}_1 + \cdots + \mathcal{M}_k = I$. Performing the measurement $\mathcal{M}$ on state ρ yields outcome i with probability $\mathrm{tr}[\rho \mathcal{M}_i]$. Thus we can interpret $\mathcal{M}$ as a linear map from $\mathcal{L}(V)$ to $\mathbb{R}^k$, with the psd and normalization conditions serving to guarantee that $\mathcal{M}$ maps $\mathcal{D}(V)$ to valid probability distributions.

Measurements on multipartite states. For our purposes, we will consider a quantum state to be destroyed after it is measured. However, if we have a quantum state on multiple systems, such as $A \otimes B$, and we measure only system A, then we will still have a quantum state on system B. In this case, the probability of obtaining outcome i is $\mathbb{P}[i] = \mathrm{tr}[\mathcal{M}_i \rho_A]$ and the residual state in this case is

$$\frac{\mathrm{tr}_A[(\mathcal{M}_i \otimes I)\rho_{AB}]}{\mathbb{P}[i]}. \tag{81}$$

Since $\sum_i \mathcal{M}_i = I$, we can verify that if we average over all measurement outcomes, then system B is left in the state ρ_B, independent of the choice of measurement. This is an important feature of quantum mechanics; despite the possibility of entanglement, there is no way for Alice (who controls system A) to signal to Bob (who controls system B) through her choice of measurement.

Restricted classes of measurements. Consider a bipartite system $A \otimes B$, with systems A, B held by Alice and Bob respectively. Performing a general measurement on $A \otimes B$ may require that Alice and Bob exchange quantum messages, so it is often more practical for them to consider only measurements that they can perform using Local Operations and Classical Communication (LOCC). Although such restricted measurements were initially introduced to model these practical restrictions, they have since arisen in settings such as [11], [40] for completely different reasons. The class LOCC is difficult to work with and

is cumbersome to even properly define—see [14] for a discussion—so we will often work with various restrictions or relaxations of it. A restriction which is interesting in its own right is the class 1-LOCC, which corresponds to Alice performing a measurement locally and sending the outcome to Bob. We say that $\mathcal{M} \in \text{1-LOCC}$ if $\mathcal{M} = \{\mathcal{M}_{i,j}\}$ with $\mathcal{M}_{i,j} = X_i \otimes Y_{i,j}$, each $X_i, Y_{i,j} \geq 0$, $\sum_i X_i = I$ and for each i, $\sum_j Y_{i,j} = I$. On the other hand, a useful relaxation is the set SEP, for which each $\mathcal{M}_i$ should have the form $\mathcal{M}_i = \sum_j X_{i,j} \otimes Y_{i,j}$ with each $X_{i,j}, Y_{i,j} \geq 0$. An even further relaxation is PPT for which we demand only that each $\mathcal{M}_i^\Gamma \geq 0$ (apart from the usual conditions that $\sum_i \mathcal{M}_i = I$ and each $\mathcal{M}_i \geq 0$). Finally we use ALL to denote the set of all measurements. Summarizing, we have

$$\text{1-LOCC} \subset \text{LOCC} \subset \text{SEP} \subset \text{PPT} \subset \text{ALL}.$$

In each case, we consider measurements with any finite number of outcomes, so these classes are technically not compact.

Entanglement swapping. An important concept in our work (building on [48]) is that of compatible pairs of families of measurements and states. We say that a POVM element $\mathcal{M}_i$ is compatible with a family of states $\mathbf{S}$ if for each n and each $\rho \in S^n$, applying $\mathcal{M}_i$ to the first system leaves a residual state (defined by (81)) that is in S^{n-1}. A family of measurements $\mathbf{M}$ is compatible with $\mathbf{S}$ if each POVM element of each measurement in $\mathbf{M}$ is compatible with $\mathbf{S}$. If $\mathbf{S} = \text{Sep}$, then 1-LOCC, LOCC, SEP are all compatible with $\mathbf{S}$. If $\mathbf{S} = \text{PPT}$ then the set of compatible measurements includes PPT. However, it is easy to construct examples of incompatible pairs. Let $|1\rangle, \ldots, |d\rangle$ be an orthonormal basis of column vectors for $\mathbb{C}^d$ and define $|\Psi\rangle = \frac{1}{d} \sum_{i,j \in [d]} |i\rangle \otimes |j\rangle \otimes |i\rangle \otimes |j\rangle$. Observe that Ψ has entanglement between systems 1:3 and systems 2:4, but is product across the 13:24 cut. Now consider a measurement acting on systems 12. One can calculate that

$$\text{tr}_{12}[(\mathcal{M}_i \otimes I) |\Psi\rangle \langle\Psi|] = \frac{\mathcal{M}_i^T}{d}. \tag{82}$$

Thus, if $\mathcal{M}_i^T$ is proportional to an entangled state, then the measurement can create entanglement on the previous unentangled states 3,4 that were not measured. This phenomenon—in which we start with $A_1 : A_2$ and $B_1 : B_2$ entanglement, measure $A_1 B_1$ and end with $A_2 : B_2$ entanglement—is called entanglement swapping [35] and is one of the main new difficulties encountered in attempting to perform hypothesis testing with respect to classes such as Sep.

Entropy. The classical (Shannon) entropy of a distribution p is given by $H(p) = -\sum_i p_i \log(p_i)$. The quantum analogue is called the von Neumann entropy, and is given by $H(\rho) = -\text{tr}[\rho \log \rho]$. Observe that $H(\rho)$ is the Shannon entropy of the eigenvalues of ρ, and coincides with the Shannon entropy when we consider probability distributions to be diagonal density matrices. If ρ_{ABC} is a multipartite state, then we let $H(A)_\rho := H(\rho_A), H(AB)_\rho = H(\rho_{AB})$, etc. When ρ is understood, we may write simply $H(A), H(AB), \ldots$. Analogous to the classical mutual information, conditional entropy, etc. we can define

$$H(A \mid B) := H(AB) - H(B) \tag{83}$$

$$I(A : B) := H(A) + H(B) - H(AB) \tag{84}$$

$$I(A : B \mid C) := H(AC) + H(BC) - H(ABC) - H(C), \tag{85}$$

in each case with an implicit dependence on some state ρ. Finally, the quantum relative entropy is $D(\rho \,\|\, \sigma) := \text{tr}[\rho(\log \rho - \log \sigma)]$. Many of these quantities behave similarly to their classical analogues, but a number of new subtleties emerge; see Chapter 11 of [57] or Chapter 11 of [44] for more information.

ACKNOWLEDGMENTS

We are grateful to Keiji Matsumoto for helpful conversations about hypothesis testing, and AWH and FGSLB also thank the Mittag-Leffler Institute for their hospitality while some of this work was done. FB was funded by EPSRC. AWH was funded by NSF grant CCF-1111382, CCF-1452616, and ARO contract W911NF-12-1-0486. JRL was supported by NSF grants CCF-1217256 and CCF-0905626.

REFERENCES

[1] K. Audenaert, J. Calsamiglia, L. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, and F. Verstraete. Discriminating states: The quantum Chernoff bound. *Phys. Rev. Lett.*, 98, 2007, arXiv:quant-ph/0610027.

[2] B. Barak, F. G. Brandão, A. W. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th symposium on Theory of Computing*, STOC '12, pages 307–326, 2012, arXiv:1205.4484.

[3] S. Beigi and P. W. Shor. Approximating the set of separable states using the positive partial transpose test. *J. Math. Phys.*, 51(4):042202, 2010, arXiv:0902.1806.

[4] M. Berta, O. Fawzi, and M. Tomamichel. On variational expressions for quantum relative entropies. *Letters in Mathematical Physics*, 107(12):2239–2265, 2017, arXiv:1512.02615.

[5] I. Bjelaković, J.-D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze, and A. Szkoła. A quantum version of Sanov's theorem. *Commun. Math. Phys.*, 260(3):659–671, 2005, arXiv:quant-ph/0412157.

[6] I. Bjelaković and R. Siegmund-Schultze. Quantum Stein's lemma revisited, inequalities for quantum entropies, and a concavity theorem of Lieb, 2012, arXiv:quant-ph/0307170.

[7] G. Bowen and N. Datta. Beyond iid in quantum information theory. In *Information Theory, 2006 IEEE International Symposium on*, pages 451–455. IEEE, 2006, arXiv:quant-ph/0604013.

[8] F. G. Brandão and M. B. Plenio. A generalization of quantum Stein's lemma. *Commun. Math. Phys.*, 295:791, 2010, arXiv:0904.0281.

[9] F. G. S. L. Brandão and G. Gour. Reversible framework for quantum resource theories. *Phys. Rev. Lett.*, 115:070503, Aug 2015, arXiv:1502.03149.

[10] F. G. S. L. Brandão and A. W. Harrow. Quantum de Finetti theorems under local measurements with applications, 2012, arXiv:1210.6367.

[11] F. G. S. L. Brandão, M. Christandl, and J. Yard. Faithful squashed entanglement. *Commun. Math. Phys.*, 306(3):805–830, 2011, arXiv:1010.1750.

[12] F. G. S. L. Brandão, M. Christandl, and J. Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proc. 43rd Annual ACM Symp. Theory of Computing*, pages 343–351, NY, USA, 2011. ACM New York, arXiv:1011.2751.

[13] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 1952.

[14] E. Chitambar, D. Leung, L. Mancinska, M. Ozols, and A. Winter. Everything you always wanted to know about LOCC (but were afraid to ask), 2012, arXiv:1210.4583.

[15] T. M. Cover and J. A. Thomas. *Elements of information theory*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second edition, 2006.

[16] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin. Quantum channel capacity of very noisy channels. *Phys. Rev. A*, 57:830, 1998, arXiv:quant-ph/9706061.

[17] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri. Complete family of separability criteria. *Phys. Rev. A*, 69:022308, Feb 2004, arXiv:quant-ph/0308032.

[18] M. J. Donald. On the relative entropy. *Communications in Mathematical Physics*, 105(1):13–34, 1986.

[19] F. Fangwei and S. Shiyi. Hypothesis testing for arbitrarily varying source. *Acta Mathematica Sinica*, 12(1):33–39, Mar 1996.

[20] B. Farkas and S. G. Révész. Potential theoretic approach to rendezvous numbers. *Monatshefte für Mathematik*, 148(4):309–331, Aug 2006, arXiv:math/0503423.

[21] H. Fawzi and O. Fawzi. Efficient optimization of the quantum relative entropy. *Journal of Physics A: Mathematical and Theoretical*, 51(15):154003, mar 2018, arXiv:1705.06671.

[22] O. Fawzi and R. Renner. Quantum conditional mutual information and approximate Markov chains. *Communications in Mathematical Physics*, 340(2):575–611, 2015, arXiv:1410.0664.

[23] D. Gottesman. *The Heisenberg Representation of Quantum Computers*. International Press, Cambridge, MA, 1999, arXiv:quant-ph/9807006.

[24] G. Gour and R. W. Spekkens. The resource theory of quantum reference frames: manipulations and monotones. *New Journal of Physics*, 10(3):033023, 2008, arXiv:0711.0043.

[25] A. W. Harrow and A. Montanaro. An efficient test for product states, with applications to quantum Merlin-Arthur games. In *Proc. 51st Symp. on FOCS*, pages 633–642, 2010, arXiv:1001.0017.

[26] M. B. Hastings. A counterexample to additivity of minimum output entropy. *Nature Physics*, 5, 2009, arXiv:0809.3972.

[27] M. Hayashi. Optimal sequence of quantum measurements in the sense of Stein's lemma in quantum hypothesis testing. *J. Phys. A*, 35(50):10759–10773, 2002, arXiv:quant-ph/0208020.

[28] M. Hayashi. *Quantum information: an introduction*. Springer-Verlag, 2006.

[29] M. Hayashi. Discrimination of two channels by adaptive methods and its application to quantum system. *IEEE Trans. Inf. Theory*, 55(8):3807–3820, 2009, arXiv:0804.0686.

[30] M. Hayashi and M. Owari. Tight asymptotic bounds on local hypothesis testing between a pure bipartite state and the white noise state. *IEEE Transactions on Information Theory*, 63(6):4008–4036, 2017, arXiv:1409.3897.

[31] F. Hiai and D. Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Commun. Math. Phys.*, 143(1):99–114, 1991.

[32] W. Hoeffding. Asymptotically optimal tests for multinomial distributions. *Ann. Math. Statist.*, 36(2):369–401, 04 1965.

[33] M. Horodecki, J. Oppenheim, and R. Horodecki. Are the laws of entanglement theory thermodynamical? *Phys. Rev. Lett.*, 89:240403, Nov 2002, arXiv:quant-ph/0207177.

[34] B. Ibinson, N. Linden, and A. Winter. Robustness of quantum Markov chains. *Commun. Math. Phys.*, 277(2):289–304, 2008, arXiv:quant-ph/0611057.

[35] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert. "Event-ready-detectors" Bell experiment via entanglement swapping. *Phys. Rev. Lett.*, 71:4287–4290, Dec 1993.

[36] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. AMS, 2002.

[37] L. Lami, B. Regula, X. Wang, R. Nichols, A. Winter, and G. Adesso. Gaussian quantum resource theories, 2018, arXiv:1801.05450.

[38] E. Levitan and N. Merhav. A competitive Neyman-Pearson approach to universal hypothesis testing with applications. *Information Theory, IEEE Transactions on*, 48(8):2215–2229, 2002.

[39] K. Li. Second-order asymptotics for quantum hypothesis testing. *The Annals of Statistics*, 42(1):171–189, 2014, arXiv:1208.1400.

[40] K. Li and A. Winter. Relative entropy and squashed entanglement. *Communications in Mathematical Physics*, 326(1):63–80, 2014, arXiv:1210.3181.

[41] E. Lieb and M. Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *J. Math. Phys.*, 14(12):1938, 1973.

[42] W. Matthews and S. Wehner. Finite blocklength converse bounds for quantum channels. *IEEE Transactions on Information Theory*, 60(11):7317–7329, 2014, arXiv:1210.4722.

[43] P. Milgrom. *Putting Auction Theory to Work*. Cambridge University Press, 2004.

[44] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, New York, 2000.

[45] M. Nussbaum and A. Szkoła. The chernoff lower bound for symmetric quantum hypothesis testing. *The Annals of Statistics*, 37(2):1040–1057, 2009, arXiv:quant-ph/0607216.

[46] T. Ogawa and H. Nagaoka. Strong converse and Stein's lemma in quantum hypothesis testing. *Information Theory, IEEE Transactions on*, 46(7):2428–2433, 2000, arXiv:quant-ph/9906090.

[47] D. Petz. Monotonicity of quantum relative entropy revisited. *Rev. Math. Phys.*, 15(01):79–91, 2003, arXiv:quant-ph/0209053.

[48] M. Piani. Relative entropy of entanglement and restricted measurements. *Phys. Rev. Lett.*, 103:160504, Oct 2009, arXiv:0904.2705.

[49] E. Posner. Random coding strategies for minimum entropy. *IEEE Transactions on Information Theory*, 21(4):388–391, 1975.

[50] G. A. Raggio and R. F. Werner. Quantum statistical mechanics of general mean field systems. *Helv. Phys. Acta*, 62:980–1003, 1989.

[51] M. Sion. On general minimax theorems. *Pacific J. Math.*, 8:171–176, 1958.

[52] G. Smith. Quantum channel capacities. In *Information Theory Workshop (ITW), 2010 IEEE*, pages 1–5, 2010, arXiv:1007.2855.

[53] D. Sutter. *Approximate quantum Markov chains*. PhD thesis, ETH Zurich, 2018-01, arXiv:1802.05477.

[54] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Phys. Rev. A*, 57:1619–1633, Mar 1998, arXiv:quant-ph/9707035.

[55] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight. Quantifying entanglement. *Phys. Rev. Lett.*, 78:2275–2279, Mar 1997, arXiv:quant-ph/9702027.

[56] K. G. H. Vollbrecht and R. F. Werner. Entanglement measures under symmetry. *Phys. Rev. A*, 64:062307, Nov 2001, arXiv:quant-ph/0010095.

[57] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013, arXiv:1106.1445.

[58] M. M. Wolf, T. S. Cubitt, and D. Perez-Garcia. Are problems in quantum information theory (un)decidable?, 2011, arXiv:1111.5425.