Sample Complexity of Locally Differentially Private Quantum Hypothesis Testing

Hao-Chung Cheng[3,4,5], Christoph Hirche[1], Cambyse Rouzé[1,2]

[1]*Institute for Information Processing (tnt/L3S), Leibniz Universität Hannover, Germany*
[2]*Inria, Télécom Paris - LTCI, Institut Polytechnique de Paris, 91120 Palaiseau, France*
Zentrum Mathematik, Technische Universität München, 85748 Garching, Germany
[3]*Department of Electrical Engineering, National Taiwan University , Taipei 10617, Taiwan (R.O.C.)*
[4]*Physics Division, National Center for Theoretical Sciences, Taipei 10617, Taiwan (R.O.C.)*
[5]*Hon Hai (Foxconn) Quantum Computing Centre, New Taipei City 236, Taiwan (R.O.C.)*

***Abstract*—Quantum state discrimination is an important problem in many information processing tasks. In this work we are concerned with finding the best possible sample complexity when the states are preprocessed by a quantum channel that is required to be locally differentially private. We give achievability and converse bounds that nearly match the best known classical bounds. On the way, we prove several novel inequalities between quantum divergences that should be of independent interest.**

I. Introduction

Hypothesis testing is a fundamental primitive in information theory. The most basic setting is that of state discrimination where we are given a quantum state that is either in the state ρ or in the state σ. The goal is to identify which of these is the case and the corresponding error probability, assuming equal priors, is given by

$$p_e(\rho,\sigma) = \frac{1}{2}(1 - E_1(\rho\|\sigma)), \tag{I.1}$$

where $E_1(\rho\|\sigma) := \frac{1}{2}\|\rho - \sigma\|_1$ is the trace distance between the state ρ and σ. The chance of identifying the state correctly can be improved by considering the availability of n copies of the state which allows to measure them jointly. At this point several figures of merit can become useful. If one simply aims to minimize the probability of error in the asymptotic limit then the corresponding error exponent is famously given by the Chernoff exponent [3], [4].

A different approach is to instead minimize the number of samples needed to achieve a certain error probability. This is called the sample complexity of the discrimination problem and is defined as

$$\mathrm{SC}_\delta(\rho,\sigma) := \inf\{n \,|\, p_e(\rho^{\otimes n}, \sigma^{\otimes n}) \leq \delta\}. \tag{I.2}$$

For simplicity, we will from here on fix $\delta = 1/10$ and drop it from the notation. It is a folklore result that the sample complexity in this setting is given by [12], [7]

$$\mathrm{SC}(\rho,\sigma) = \Theta\left(\frac{1}{-\log F(\rho,\sigma)}\right), \tag{I.3}$$

where $F(\rho,\sigma)$ is the quantum fidelity. An intermediate problem is that of sequential hypothesis testing [23], [19]. In this work, we consider the sample complexity when the state in question is affected by a noisy channel before we receive it. Explicitly, we are considering the class of ϵ-locally differentially private quantum channels, denoted $\mathrm{LDP}(\epsilon)$, which is defined as [15]

$$\mathrm{LDP}_\epsilon = \{\mathcal{A} \,|\, E_{e^\epsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) = 0 \quad \forall \rho, \sigma\}. \tag{I.4}$$

That is, we are interested in the sample complexity

$$\mathrm{SC}_\epsilon(\rho,\sigma) := \inf_{\mathcal{A}\in\mathrm{LDP}_\epsilon} \mathrm{SC}(\mathcal{A}(\rho), \mathcal{A}(\sigma)), \tag{I.5}$$

namely the smallest possible sample size given that the states are subject to locally differetially private noise. In the classical case such problems have received a fair amount of attention recently [10], [17], [8], [1] and the classical equivalent of our particular problem was investigated in [2]. The main result of our work can be summarized by the following inequalities:

$$\begin{aligned}
&\max\left\{\frac{(e^\epsilon+1)\log 2.5}{2(e^\epsilon-1)H_{\frac{1}{2}}(\rho\|\sigma)}, \frac{16}{25}\frac{e^\epsilon}{(e^\epsilon-1)^2\, E_1(\rho\|\sigma)^2}\right\} \\
&\leq \mathrm{SC}_\epsilon(\rho,\sigma) \qquad\qquad\qquad\qquad\qquad\qquad \text{(I.6)} \\
&\leq \left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2 \frac{2\log 5}{E_1(\rho\|\sigma)^2}.
\end{aligned}$$

These mirror mostly the classical result in [2, Lemma

2], up to some differences in the ϵ-dependent constants that we will discuss later. In particular, we recover the interesting observation that the best known lower bound is given by a different divergence depending on the value of ϵ. Note that this ”phase-transition” is known to persist in the exact behaviour of binary classical probability distributions [20].

The main challenge in deriving the above result lies in the need of several new entropic inequalities that we believe should also be of independent interest. To that end we use the recently established framework of f-divergences defined via integral representations from [16]. In particular, these divergences behave well when we consider their contraction coefficients which allows us to prove new bounds for several such coefficients under local differential privacy.

A. Notations

We denote by $\mathcal{S}_d$ the set of d-dimensional quantum states, and by $\mathcal{L}_d$ that of $d \times d$ arbitrary matrices. We will also adopt standard notations from quantum information theory: for a quantum system A, $|A|$ denotes the dimension of its associated Hilbert space, $\mathcal{S}_A$ the set of its quantum states, and $\mathcal{L}_A \equiv \mathcal{L}_{|A|}$. Given a self-adjoint matrix H, we denote by $\{H \geq 0\}$ the projection onto the sum of eigensubspaces corresponding to the non-negative eigenvalues of H, and write $H_+ := H\{H \geq 0\}$. Similarly, for two states $\rho, \sigma \in \mathcal{S}_A$, $\{\rho \geq \sigma\} := \{\rho - \sigma \geq 0\}$. We use natural logarithm $\log$ throughout this paper.

II. Sample complexity of hypothesis testing

Given any two states $\rho, \sigma \in \mathcal{S}_d$, the optimal sample complexity for the task of hypothesis testing between ρ and σ, namely the number of copies needed to distinguish ρ from σ, satisfies [7]

$$\mathrm{SC}(\rho,\sigma) = \Theta\left(\frac{-1}{\log F(\rho,\sigma)}\right) = \Theta\left(\frac{1}{d_B^2(\rho,\sigma)}\right), \quad \text{(II.1)}$$

where $F(\rho,\sigma) := \|\sqrt{\rho}\sqrt{\sigma}\|_1$ denotes the fidelity between the quantum states ρ and σ, and where $d_B(\rho,\sigma) := \sqrt{1 - F(\rho,\sigma)}$ is their so-called Bures distance. We recall that the fidelity is also given by the $\frac{1}{2}$-sandwiched Rényi divergence via $\widetilde{D}_{\frac{1}{2}}(\rho\|\sigma) = -2\log F(\rho,\sigma)$. In the following, we will also make use of the max-relative entropy $D_\infty(\rho\|\sigma) := \log \|\sigma^{-\frac{1}{2}}\rho\sigma^{-\frac{1}{2}}\|_\infty$, where we assume that σ is invertible for simplicity. Recently, a different family of Rényi divergences was introduced based on the following integral representation: for $\alpha > 0$ with $\alpha \neq 1$,

$$D_\alpha(\rho\|\sigma) = \frac{1}{\alpha - 1}\log\left(1 + (\alpha - 1)H_\alpha(\rho\|\sigma)\right), \quad \text{(II.2)}$$

with

$$\begin{aligned} &H_\alpha(\rho\|\sigma) \quad \text{(II.3)} \\ &= \alpha \int_1^\infty \Big(\gamma^{\alpha-2}E_\gamma(\rho\|\sigma) + \gamma^{-\alpha-1}E_\gamma(\sigma\|\rho)\Big) d\gamma\,, \end{aligned}$$

where $E_\gamma(\rho\|\sigma) := \mathrm{Tr}(\rho - \gamma\sigma)_+$ denotes the quantum Hockey-Stick divergence. We also recall that the limit $\alpha \to 1$ leads to the standard Umegaki relative entropy $D_1(\rho\|\sigma) \equiv D(\rho\|\sigma) := \mathrm{Tr}\left[\rho(\log\rho - \log\sigma)\right]$. Note that the Rényi divergences D_α defined above is different from the family of Petz Rényi divergences, and that they coincide when ρ and σ commute and correspond to two probability distributions P and Q. In this case,

$$\begin{aligned} D_\alpha(P\|Q) &= \frac{1}{\alpha-1}\log(\mathrm{Tr}(P^\alpha Q^{1-\alpha})) \\ &\Leftrightarrow H_\alpha(P\|Q) = \frac{1}{\alpha-1}\big(\mathrm{Tr}(P^\alpha Q^{1-\alpha}) - 1\big)\,. \end{aligned}$$

An important special case which we will consider here is the case $\alpha = 1/2$. Then $H_{\frac{1}{2}}(P,Q) = 2\big(1 - \mathrm{Tr}\, P^{\frac{1}{2}}Q^{\frac{1}{2}}\big) = \mathrm{Tr}\left[(\sqrt{P} - \sqrt{Q})^2\right]$ corresponds to (twice) the squared Hellinger distance between P and Q. It is a well-known fact that the sample complexity of hypothesis testing between two distributions satisfies

$$\mathrm{SC}(P,Q) = \Theta\left(\frac{1}{H_{\frac{1}{2}}(P,Q)}\right) \quad \text{(II.4)}$$

(see [5] for the lower bound, and [6] for the upper bound). To recover this classical asymptotic relation from (II.1), we need to relate $H_{\frac{1}{2}}$ to the usual quantities:

Lemma II.1. *For any two states $\rho, \sigma \in \mathcal{S}_d$,*

$$1 - F(\rho,\sigma) \leq \frac{H_{\frac{1}{2}}(\rho\|\sigma)}{2} \leq E_1(\rho\|\sigma) \leq \sqrt{1 - F(\rho,\sigma)^2}.$$

Proof. The first inequality follows from $\widetilde{D}_{\frac{1}{2}}(\rho\|\sigma) \leq D_{\frac{1}{2}}(\rho\|\sigma)$ and the second from $E_\gamma \leq E_1$. This is essentially also the argument in the proof of [16, Corollary 5.6]. The third inequality is the usual Fuchs-van-de-Graaf inequality [12]. □

Hence, we can argue from these bounds that also in the quantum setting we can at least partially express SC in terms of $H_{\frac{1}{2}}(\rho\|\sigma)$:

Proposition II.2. *For any two states ρ and σ with $H_{\frac{1}{2}}(\rho\|\sigma) \leq 1$, we have*

$$\frac{\log(2.5)}{2H_{\frac{1}{2}}(\rho\|\sigma)} \leq \mathrm{SC}(\rho,\sigma) \leq \frac{\log 5}{1-F(\rho\|\sigma)}\,. \tag{II.5}$$

The proof can be found in Appendix A-A.

III. Locally differentially private hypothesis testing

This paper aims at finding tight bounds on the sample complexity of optimal hypothesis tests subject to local privacy guarantees [11], [18]. Loosely speaking, a random mechanism (e.g. an algorithm or communication protocol) is said to be locally differentially private (LDP) if its output does not vary significantly with arbitrary perturbation of the input. LDP was initially introduced in the classical setting for the scenario where a database is compiled from numerous clients, each insisting on individual privacy assurances. In this scenario, each client employs an algorithm $\mathcal{A}$ to obfuscate their input to the database. The primary objective is not to create similarities between neighboring states, but rather to conceal the broader information being transmitted.

In the present article, we focus on the task of LDP hypothesis testing. There, identical copies of a state $\omega \in \{\rho,\sigma\}$ are given and we need to discriminate between the hypotheses $\omega=\rho$ and $\omega=\sigma$. However, as opposed to the nonprivate setting, we are restricted in the set of positive operator-valued measures (POVM) which we can use in order to complete the task:

Definition III.1 (ϵ-locally differentially private channel [15])**.** *Given $\epsilon \geq 0$, a quantum channel $\mathcal{A}:\mathcal{L}_A \to \mathcal{L}_B$ is called ϵ-locally differentially private if for all states $\rho,\sigma \in \mathcal{S}_A$,*

$$E_{e^\epsilon}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) = 0\,.$$

We denote by $\mathrm{LDP}_\epsilon(A,B)$ the set of ϵ-locally differentially private quantum channels from A to B.

Then, given two states $\rho,\sigma \in \mathcal{S}_A$ and $\epsilon \geq 0$, their optimal sample complexity for ϵ-LDP hypothesis testing with outputs in B is defined as

$$\mathrm{SC}_{B,\epsilon}(\rho,\sigma) := \inf_{\mathcal{A}\in \mathrm{LDP}_\epsilon(A,B)} \mathrm{SC}(\mathcal{A}(\rho),\mathcal{A}(\sigma))\,. \tag{III.1}$$

In the following, we often will not specify the output B to the locally differentially private algorithms $\mathcal{A}$ as it won't play an important role in our derivations. Thus, we will more simply denote $\mathrm{SC}_\epsilon \equiv \mathrm{SC}_{B,\epsilon}$. From (II.5), we directly get that

$$\mathrm{SC}_\epsilon(\rho,\sigma) = \Theta\left(\frac{-1}{\sup_{\mathcal{A}\in\mathrm{LDP}_\epsilon(A,B)} \log F(\mathcal{A}(\rho)\|\mathcal{A}(\sigma))}\right). \tag{III.2}$$

The above expression is somewhat unsatisfactory due to the presence of an optimization over all LDP mechanisms. Ideally, we would like to derive tight upper and lower bounds for the SC_ϵ which do not depend on such optimization.

In the classical setting, fundamental limits of statistical problems under LDP have been successfully characterized using information-theoretic concepts. Arguably one of the most fundamental notions in (quantum) information theory revolves around data processing. Under the influence of a quantum channel, numerous relevant quantities exhibit monotonic behavior. This characteristic allows us to attribute operational significance to these quantities concerning distinguishability, consequently facilitating their utility in assessing physical properties. For instance, the data processing inequality states that, for any two states $\rho,\sigma\in\mathcal{S}_A$ and any quantum channel $\mathcal{N}:\mathcal{L}_A\to\mathcal{L}_B$, $D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq D(\rho\|\sigma)$. Intuitively, and in view of the operational interpretation of the relative entropy as a measure of distinguishability between ρ and σ, it becomes clear that applying a quantum channel to the state never simplifies the discrimination task, thus leading to a reduction in the relative entropy. The above contraction is so fundamental to information theory that it is often taken as a requirement for any metric on quantum states to be called an information measure. Data processing can further be quantified through the use of so-called contraction coefficients, defined as

$$\eta(\mathcal{N}) := \sup_{\substack{\rho,\sigma\in\mathcal{S}_A\\ \rho\neq\sigma}} \frac{D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))}{D(\rho\|\sigma)}. \tag{III.3}$$

See also [13], [14], [16] for additional properties and discussions. The study of classical statistical problems under local privacy through the use of contraction coefficients was initiated in [8], [9], where it was shown that $\mathrm{SC}_\epsilon(P,Q) = \Theta(\epsilon^{-2}\|P-Q\|_{\mathrm{TV}}^{-2})$, where $\|P-Q\|_{\mathrm{TV}} := \frac{1}{2}\sum_{x\in\mathcal{X}}|P(x)-Q(x)|$ denotes the total variation between distributions P and Q over the alphabet $\mathcal{X}$. More recently, the following lower and upper bounds were derived in [2, Lemma 2]:

$$\left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2 \frac{\log(2.5)}{8H_{\frac{1}{2}}(P,Q)} \leq \mathrm{SC}_\epsilon(P,Q) \tag{III.4}$$

$$\frac{2}{25e^{-\epsilon}(e^\epsilon-1)^2\left\|P-Q\right\|_{\mathrm{TV}}^2} \leq \mathrm{SC}_\epsilon(P,Q) \tag{III.5}$$

$$\mathrm{SC}_\epsilon(P,Q) \leq \left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2 \frac{2\log(5)}{\|P-Q\|_{\mathrm{TV}}^2}\,. \tag{III.6}$$

In the next sections, we aim at extending the above upper and lower bounds in to the framework of quantum states. All omitted proofs can be found in the appendix.

A. Achieving LDP optimal sample complexity

Our first main result is the following achievability bound for LDP hypothesis testing:

Theorem III.2 (Achievability of LDP hypothesis testing)**.** *For any two states $\rho, \sigma \in \mathcal{S}_A$,*

$$\mathrm{SC}_\epsilon(\rho,\sigma) \leq \left(\frac{e^\epsilon+1}{e^\epsilon-1}\right)^2 \frac{2\log 5}{E_1(\rho\|\sigma)^2}\,. \tag{III.7}$$

B. Optimality of LDP sample complexity

Next, we aim at finding a lower bound for SC_ϵ. In fact, we derive two different ones.

a) Converse, Part I: For our first lower bound, we make use of contraction coefficients for the trace distance and the relative entropy: given a quantum channel $\mathcal{N}: \mathcal{L}_A \to \mathcal{L}_B$,

$$\begin{aligned}\eta_{\mathrm{Tr}}(\mathcal{N}) &:= \sup_{\substack{\rho,\sigma\in\mathcal{S}_A\\ \rho\neq\sigma}} \frac{\|\mathcal{N}(\rho-\sigma)\|_1}{\|\rho-\sigma\|_1}\\ &= \sup_{\Psi\perp\Phi} E_1(\mathcal{N}(\Psi)\|\mathcal{N}(\Phi))\,,\end{aligned}$$

where the second equality was shown in [21] with an optimization over orthogonal pure states. We start by proving a couple of Lemmas that generalize their classical analogues given in [2]. The first one, whose proof we defer to Appendix A-C, uses the tools from [16, Section 5.1].

Lemma III.3. *For any two states $\rho, \sigma \in \mathcal{S}_A$,*

$$\begin{aligned}&H_{\frac{1}{2}}(\rho\|\sigma)\\ &\leq \left(\frac{(e^{\frac{1}{2}D_\infty(\rho\|\sigma)}-1)^2}{e^{D_\infty(\rho\|\sigma)}-1} + \frac{(e^{\frac{1}{2}D_\infty(\sigma\|\rho)}-1)^2}{e^{D_\infty(\sigma\|\rho)}-1}\right) E_1(\rho\|\sigma)\end{aligned}$$

This essentially follows the proof of [16, Proposition 5.2] but gives, by always choosing the tightest known bound, a slightly better result than [16, Corollary 5.5]. Next, we give a bound on the maximum output trace distance of LDP channels. See Appendix A-D for a proof:

Lemma III.4. *We have,*

$$\sup_{\mathcal{A}\in\mathrm{LDP}_\epsilon}\ \sup_{\rho,\sigma\in\mathcal{S}_A} E_1(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \frac{e^{-\epsilon}(e^\epsilon-1)^2}{e^\epsilon-e^{-\epsilon}}. \tag{III.8}$$

We are now ready to prove the main result of this section (see Appendix A-E for details).

Proposition III.5. *For any quantum channel $\mathcal{A} \in \mathrm{LDP}_\epsilon$,*

$$\eta_{\mathrm{Tr}}(\mathcal{A}) \leq \frac{(e^\epsilon-1)}{(e^\epsilon+1)}. \tag{III.9}$$

Alternatively we could have used the previously known bound [15] $\eta_{\mathrm{Tr}}(\mathcal{A}) \leq 1-e^{-\epsilon}$, which would simplify the proof a lot, but unfortunately only gives the weaker bound $\frac{e^\epsilon-1}{e^\epsilon} \geq \frac{e^\epsilon-1}{e^\epsilon+1} \equiv \sqrt{\Upsilon_\epsilon}$.

Now the proposition implies directly that

$$\eta_f(\mathcal{A}) \leq \eta_{\mathrm{Tr}}(\mathcal{A}) \leq \sqrt{\Upsilon_\epsilon}. \tag{III.10}$$

which then implies

$$\sup_{\mathcal{A}\in\mathrm{LDP}_\epsilon} H_{\frac{1}{2}}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \sqrt{\Upsilon_\epsilon}H_{\frac{1}{2}}(\rho\|\sigma). \tag{III.11}$$

Note that, in the classical setting, [2] proves the stronger

$$\eta_f(\mathcal{A}) \leq \Upsilon_\epsilon, \tag{III.12}$$

for operator convex f. There they start with (modulo notation)

$$\begin{aligned}&\eta_{\mathrm{KL}}(\mathcal{A})\\ &\leq \sup_{x,x'} H_{\frac{1}{2}}(\mathcal{A}(\cdot|x)\|\mathcal{A}(\cdot|x')) - \frac{1}{4}H_{\frac{1}{2}}(\mathcal{A}(\cdot|x)\|\mathcal{A}(\cdot|x'))^2,\end{aligned}$$

where $\eta_{\mathrm{KL}}(\mathcal{A})$ stands for the contraction coefficient for the Kullback–Leibler divergence. Proving a quantum version of this bound remains an interesting open problem. Applying the above results to the problem of sample complexity, we get the following result.

Theorem III.6 (Converse I of LDP hypothesis testing)**.** *For any two states $\rho, \sigma \in \mathcal{S}_A$,*

$$\sup_{\mathcal{A}\in\mathrm{LDP}_\epsilon(A,B)} H_{\frac{1}{2}}(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \sqrt{\Upsilon_\epsilon}H_{\frac{1}{2}}(\rho\|\sigma)\,.$$

and hence,

$$\mathrm{SC}_\epsilon(\rho,\sigma) \geq \frac{(e^\epsilon+1)\log 2.5}{2(e^\epsilon-1)H_{\frac{1}{2}}(\rho\|\sigma)}. \tag{III.13}$$

Proof. This follows directly from Equation (III.10). □

b) Converse, Part II: This proof is based on the χ^2 divergence for which we have the following equivalent definitions [16],

$$\begin{aligned}\chi^2(\rho\|\sigma) &\equiv H_2(\rho\|\sigma)\\ &= 2\int_1^\infty (E_\gamma(\rho\|\sigma) + \gamma^{-3}E_\gamma(\sigma\|\rho))\mathrm{d}\gamma\\ &= \int_0^\infty \operatorname{Tr}[(\rho-\sigma)(\sigma+s\mathbb{1})^{-1}(\rho-\sigma)(\sigma+s\mathbb{1})^{-1}]\,\mathrm{d}s.\end{aligned}$$

The next Lemma is proved in Appendix A-F and is the core technical ingredient of this section.

Lemma III.7. *For an arbitrary quantum channel $\mathcal{N}$ and two input states $\rho, \sigma \in \mathcal{S}_A$,*

$$\chi^2(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq 2E_1(\rho\|\sigma)^2 \max_{\Psi,\Phi} \chi^2(\mathcal{N}(\Psi)\|\mathcal{N}(\Phi)).$$

Note that this improves also the known classical result by a factor 2. Now, similar to the classical case, we have

$$\max_{\Psi,\Phi} \chi^2(\mathcal{N}(\Psi)\|\mathcal{N}(\Phi)) \leq \max_{\substack{\tau,\theta\\ E_{e^\epsilon}(\tau\|\theta)=0\\ E_{e^\epsilon}(\theta\|\tau)=0}} \chi^2(\tau\|\theta), \quad \text{(III.14)}$$

motivating the need for the following observation.

Lemma III.8. *We have*

$$\max_{\substack{\tau,\theta\\ E_{e^\epsilon}(\tau\|\theta)=0\\ E_{e^\epsilon}(\theta\|\tau)=0}} E_1(\tau\|\theta) = e^{-\epsilon}\frac{(e^\epsilon-1)^2}{e^\epsilon - e^{-\epsilon}},$$

and hence,

$$\max_{\substack{\tau,\theta\\ E_{e^\epsilon}(\tau\|\theta)=0\\ E_{e^\epsilon}(\theta\|\tau)=0}} D_f(\tau\|\theta) \leq \frac{f(e^\epsilon)+e^\epsilon f(e^{-\epsilon})}{e^\epsilon - 1} e^{-\epsilon}\frac{(e^\epsilon-1)^2}{e^\epsilon - e^{-\epsilon}}.$$

Proof. We start with the first statement for the trace distance. Define the measurement $\{\Pi_+, \Pi_- = \mathbb{1} - \Pi_+\}$ and the probabilities $\{p = \operatorname{Tr}\Pi_+\tau, 1-p\}$ and $\{q = \operatorname{Tr}\Pi_+\theta, 1-q\}$. It is known that this measurement achieves the trace distance and hence $E_1(\tau\|\theta) = E_1(p\|q)$, where the right hand side is the classical binary total variation distance. Furthermore, we have $E_\gamma(p\|q) \leq E_\gamma(\tau\|\theta)$. Putting everything together we get,

$$\max_{\substack{\tau,\theta\\ E_{e^\epsilon}(\tau\|\theta)=0\\ E_{e^\epsilon}(\theta\|\tau)=0}} E_1(\tau\|\theta) = \max_{\substack{0\leq p,q\leq 1\\ E_{e^\epsilon}(p\|q)=0\\ E_{e^\epsilon}(q\|p)=0}} E_1(p\|q). \quad \text{(III.15)}$$

It has therefore taken the same expression as in the classical case for which the solution was derived in [2, Equation (45)]. The second statement for f-divergences follows from the reverse Pinsker inequality proven in [16, Proposition 5.2]. □

As a special case for $f(x) = x^2 - 1$, we have

$$\max_{\substack{\tau,\theta\\ E_{e^\epsilon}(\tau\|\theta)=0\\ E_{e^\epsilon}(\theta\|\tau)=0}} \chi^2(\tau\|\theta) \leq e^{-\epsilon}(e^\epsilon - 1)^2, \quad \text{(III.16)}$$

which matches exactly the classical case. In summary, we have shown

$$\max_{\mathcal{N}\in \mathrm{LDP}_\epsilon} \chi^2(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq 2e^{-\epsilon}(e^\epsilon-1)^2 E_1(\rho\|\sigma)^2.$$

Theorem III.9 (Converse II of LDP hypothesis testing)**.** *For any two states $\rho, \sigma \in \mathcal{S}_A$,*

$$\mathrm{SC}_\epsilon(\rho,\sigma) \geq \frac{16}{25}\frac{1}{e^{-\epsilon}(e^\epsilon-1)^2\, E_1(\rho\|\sigma)^2}$$

Proof. Consider the error probability of symmetric hypothesis testing,

$$p_e(\rho,\sigma) = \frac{1}{2}(1 - E_1(\rho\|\sigma)).$$

After rewriting that, we can continue with

$$2(1-p_e(\rho,\sigma))^2 = 2E_1(\rho\|\sigma)^2 \leq D(\rho\|\sigma) \leq \chi^2(\rho\|\sigma),$$

where the first inequality is Pinsker's inequality and the second is from [22]. Applying this to n copies and using additivity of the relative entropy we get,

$$\begin{aligned}&2(1-2p_e(\mathcal{A}(\rho)^{\otimes n}, \mathcal{A}(\sigma)^{\otimes n}))^2\\ &\quad\leq n\chi^2(\mathcal{A}(\rho)\|\mathcal{A}(\sigma))\\ &\quad\leq 2nE_1(\rho\|\sigma)^2 \max_{\Psi,\Phi}\chi^2(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi))\\ &\quad\leq 2nE_1(\rho\|\sigma)^2 e^{-\epsilon}(e^\epsilon-1)^2,\end{aligned}$$

where the second inequality is Lemma III.7 and the third Equation (III.16). Choosing the error probability as 0.1, this is gives

$$n \geq \frac{16}{25}\frac{1}{e^{-\epsilon}(e^\epsilon-1)^2\, E_1(\rho\|\sigma)^2},$$

from which the claim follows. □

Note that this is a factor 8 better than the classical result in [2]: A factor 2 because of the improvement in Lemma III.7 and a factor 4 because of a suboptimal use of Pinskers inequality in [2, Lemma 2].

Acknowledgement

HC is supported by Grants No. NSTC 112-2636-E-002-009, No. NSTC 112-2119-M-007-006, No. NSTC 112-2119-M-001-006, No. NSTC 112-2124-M-002-003, No. NTU-112V1904-4, NTU-CC-113L891605, and NTU-113L900702.

References

[1] A primer on private statistics. *arXiv preprint arXiv:2005.00010*, 2020.

[2] S. Asoodeh and H. Zhang. Contraction of locally differentially private mechanisms. *arXiv preprint arXiv:2210.13386*, 2022.

[3] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete. Discriminating states: The quantum Chernoff bound. *Physical Review Letters*, 98:160501, Apr 2007.

[4] K. M. R. Audenaert, M. Nussbaum, A. Szkoła, and F. Verstraete. Asymptotic error rates in quantum hypothesis testing. *Communications in Mathematical Physics*, 279(1):251–283, Feb 2008.

[5] Z. Bar-Yossef. *The complexity of massive data set computations*. University of California, Berkeley, 2002.

[6] C. L. Canonne, G. Kamath, A. McMillan, A. Smith, and J. Ullman. The structure of optimal private tests for simple hypotheses. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 310–321, 2019.

[7] H.-C. Cheng, N. Datta, N. Liu, T. Nuradha, R. Salzmann, and M. M. Wilde. *arXiv preprint arXiv:2403.17868*.

[8] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*. IEEE, Oct. 2013. arXiv preprint arXiv:1302.3203.

[9] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

[10] A. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, SIGMOD/PODS03. ACM, June 2003.

[11] A. Evfimievski, J. Gehrke, and R. Srikant. Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 211–222, 2003.

[12] C. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, may 1999.

[13] F. Hiai and M. B. Ruskai. Contraction coefficients for noisy quantum channels. *Journal of Mathematical Physics*, 57(1), 2016.

[14] C. Hirche, C. Rouzé, and D. S. França. On contraction coefficients, partial orders and approximation of capacities for quantum channels. *Quantum*, 6:862, 2022.

[15] C. Hirche, C. Rouzé, and D. S. França. Quantum differential privacy: An information theory perspective. *IEEE Transactions on Information Theory*, 2023.

[16] C. Hirche and M. Tomamichel. Quantum Rényi and f-divergences from integral representations. *arXiv preprint arXiv:2306.12343*, 2023.

[17] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, Oct. 2008.

[18] S. P. Kasiviswanathan, H. K. Lee, K. Nissim, S. Raskhodnikova, and A. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

[19] Y. Li, V. Y. Tan, and M. Tomamichel. Optimal adaptive strategies for sequential quantum hypothesis testing. *Communications in Mathematical Physics*, 392(3):993–1027, 2022.

[20] A. Pensia, A. R. Asadi, V. Jog, and P.-L. Loh. Simple binary hypothesis testing under local differential privacy and communication constraints. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 3229–3230. PMLR, 2023.

[21] M. B. Ruskai. Beyond strong subadditivity? improved bounds on the contraction of generalized relative entropy. *Reviews in Mathematical Physics*, 6(05a):1147–1161, 1994.

[22] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete. The χ^2-divergence and mixing times of quantum markov processes. *Journal of Mathematical Physics*, 51(12), 2010.

[23] E. M. Vargas, C. Hirche, G. Sentís, M. Skotiniotis, M. Carrizo, R. Muñoz-Tapia, and J. Calsamiglia. Quantum sequential hypothesis testing. *Physical review letters*, 126(18):180502, 2021.

Appendix A
Proofs

In this section we provide the proofs missing in the main text.

A. *Proof of Proposition II.2*

We first show the upper bound. By Lemma II.1,

$$\begin{aligned} \delta &= \frac{1}{2}\left(1 - E_1\left(\rho^{\otimes n} \| \sigma^{\otimes n}\right)\right) && \text{(A.1)} \\ &\leq \frac{1}{2} F\left(\rho^{\otimes n} \| \sigma^{\otimes n}\right) && \text{(A.2)} \\ &= \frac{1}{2} F\left(\rho \| \sigma\right)^n . && \text{(A.3)} \end{aligned}$$

We obtain

$$\mathrm{SC}(\rho,\sigma) \leq \frac{-\log(2\delta)}{-\log F(\rho,\sigma)} \leq \frac{-\log(2\delta)}{1-F(\rho,\sigma)}, \tag{A.4}$$

where the last inequality follows from $\log(1+x) \leq x$ for all $x \geq 0$.

Next, we show the lower bound. Following the idea in [5, Theorem 4.7] and using Lemma II.1, we have

$$\begin{aligned} F(\rho,\sigma)^n &= F\left(\rho^{\otimes n}, \sigma^{\otimes n}\right) && \text{(A.5)} \\ &\leq \sqrt{1 - E_1\left(\rho^{\otimes n}, \sigma^{\otimes n}\right)^2} && \text{(A.6)} \\ &\leq \sqrt{1-(1-2\delta)^2} && \text{(A.7)} \\ &= \sqrt{4\delta(1-\delta)} && \text{(A.8)} \\ &\leq \sqrt{4\delta}. && \text{(A.9)} \end{aligned}$$

Hence,

$$\begin{aligned} \mathrm{SC}(\rho,\sigma) &\geq -\frac{\log\frac{1}{4\delta}}{2\log F(\rho,\sigma)} && \text{(A.10)} \\ &\geq -\frac{\log\frac{1}{4\delta}}{2\log\left(1-\frac{H_{\frac{1}{2}}(\rho,\sigma)}{2}\right)} && \text{(A.11)} \\ &\geq \frac{\log\frac{1}{4\delta}}{2H_{\frac{1}{2}}(\rho,\sigma)}, && \text{(A.12)} \end{aligned}$$

where the last inequality follows from $\log(1-x) \geq -2x$, for all $0 \leq x \leq \frac{1}{2}$ provided that $H_{\frac{1}{2}}(\rho,\sigma) \leq 1$.

B. *Proof of Theorem III.2*

Proof. We resort to Lemma II.1 and observe that,

$$E_1(\rho\|\sigma)^2 \leq 1 - F(\rho,\sigma)^2 \leq 2(1-F(\rho,\sigma)). \tag{A.13}$$

It remains to show that there exists an LDP algorithm $\mathcal{A} : \mathcal{L}_A \to \mathcal{L}_B$ achieving the scaling for optimal sample complexity. Inspired by the use of classical binary algorithm in [2], we introduce the following channel for $\kappa \in [0,1]$:

$$\begin{aligned} \mathcal{B}(\cdot) =& |0\rangle\langle 0| \left(\kappa \operatorname{Tr}(\{\rho \geq \sigma\}\cdot) + (1-\kappa)\operatorname{Tr}(\{\rho < \sigma\}\cdot)\right) \\ &+ |1\rangle\langle 1| \left((1-\kappa)\operatorname{Tr}(\{\rho \geq \sigma\}\cdot) + \kappa \operatorname{Tr}(\{\rho < \sigma\}\cdot)\right) . \end{aligned}$$

We can easily verify that

$$\begin{aligned}
E_1(\mathcal{B}(\rho)\|\mathcal{B}(\sigma)) &= \frac{1}{2}\operatorname{Tr}|\mathcal{B}(\rho)-\mathcal{B}(\sigma)| \\
&= \frac{1}{2}|\kappa \operatorname{Tr}(\{\rho \geq \sigma\}(\rho-\sigma)) + (1-\kappa)\operatorname{Tr}(\{\rho < \sigma\}(\rho-\sigma))| \\
&\quad + \frac{1}{2}|(1-\kappa)\operatorname{Tr}(\{\rho \geq \sigma\}(\rho-\sigma)) + \kappa \operatorname{Tr}(\{\rho < \sigma\}(\rho-\sigma))| \\
&= |(2\kappa - 1)E_1(\rho\|\sigma)|.
\end{aligned}$$

We now choose $\kappa := \frac{e^\epsilon}{1+e^\epsilon}$. Therefore,

$$\begin{aligned}
E_1(\mathcal{B}(\rho)\|\mathcal{B}(\sigma)) &= \left|\Big(\frac{2e^\epsilon}{1+e^\epsilon} - 1\Big)E_1(\rho\|\sigma)\right| \\
&= \frac{e^\epsilon - 1}{e^\epsilon + 1}E_1(\rho\|\sigma)\,.
\end{aligned}$$

Combining with (A.13), we get

$$\begin{aligned}
\sup_{\mathcal{A}\in \mathrm{LDP}_\epsilon(A,B)} 1 - F(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) &\geq \frac{1}{2}\sup_{\mathcal{A}\in \mathrm{LDP}_\epsilon(A,B)} E_1(\mathcal{A}(\rho)\|\mathcal{A}(\sigma))^2 \\
&\geq \frac{1}{2}\left(\frac{e^\epsilon - 1}{e^\epsilon + 1}\right)^2 E_1(\rho\|\sigma)^2.
\end{aligned}$$

Combining with the upper bound in Proposition II.2, the result follows. □

C. Proof of Lemma III.3

Proof. By specializing Equation (II.3) we get,

$$\begin{aligned}
&H_{\frac{1}{2}}(\rho\|\sigma) \qquad\qquad (\text{A.14})\\
&= \frac{1}{2}\int_1^\infty \gamma^{-\frac{3}{2}}\Big(E_\gamma(\rho\|\sigma) + E_\gamma(\sigma\|\rho)\Big)d\gamma \qquad\qquad (\text{A.15})\\
&= \frac{1}{2}\int_1^{e^{D_\infty(\rho\|\sigma)}} \gamma^{-\frac{3}{2}}E_\gamma(\rho\|\sigma)d\gamma + \frac{1}{2}\int_1^{e^{D_\infty(\sigma\|\rho)}} \gamma^{-\frac{3}{2}}E_\gamma(\sigma\|\rho)d\gamma \qquad\qquad (\text{A.16})\\
&\leq \frac{1}{2}\int_1^{e^{D_\infty(\rho\|\sigma)}} \gamma^{-\frac{3}{2}}\frac{e^{D_\infty(\rho\|\sigma)} - \gamma}{e^{D_\infty(\rho\|\sigma)} - 1}E_1(\rho\|\sigma)d\gamma + \frac{1}{2}\int_1^{e^{D_\infty(\sigma\|\rho)}} \gamma^{-\frac{3}{2}}\frac{e^{D_\infty(\sigma\|\rho)} - \gamma}{e^{D_\infty(\sigma\|\rho)} - 1}E_1(\sigma\|\rho)d\gamma \qquad\qquad (\text{A.17})\\
&= \frac{(e^{\frac{1}{2}D_\infty(\rho\|\sigma)} - 1)^2}{e^{D_\infty(\rho\|\sigma)} - 1}E_1(\rho\|\sigma) + \frac{(e^{\frac{1}{2}D_\infty(\sigma\|\rho)} - 1)^2}{e^{D_\infty(\sigma\|\rho)} - 1}E_1(\sigma\|\rho), \qquad\qquad (\text{A.18})
\end{aligned}$$

where the second equality holds because $E_\gamma(\rho\|\sigma) = 0$ for $\gamma \geq D_\infty(\rho\|\sigma)$, the inequality by convexity of E_γ and the last equality by evaluating the integral. □

D. Proof of Lemma III.4

Proof. We follow closely the classical proof in [2, Appendix C]. First, observe,

$$\sup_{\mathcal{A}}\sup_{\rho,\sigma} E_1(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \sup_{\substack{\rho,\sigma \\ E_{e^\epsilon}(\rho\|\sigma)=0 \\ E_{e^\epsilon}(\sigma\|\rho)=0}} E_1(\rho\|\sigma), \tag{A.19}$$

which holds because the channel outputs are always close in the corresponding Hockey-stick divergence by definition. Next, define the measurement,

$$\mathcal{M}(\cdot) = |0\rangle\langle 0|\operatorname{Tr}\{P_+\cdot\} + |1\rangle\langle 1|\operatorname{Tr}\{(\mathrm{id} - P_+)\cdot\}, \tag{A.20}$$

where $P_+ = \{\rho \geq 0\sigma\}$ It is now easy to check that

$$E_1(\rho\|\sigma) = E_1(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)), \tag{A.21}$$

and, by data processing,

$$E_\gamma(\rho\|\sigma) = 0 \quad \Rightarrow \quad E_\gamma(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) = 0. \tag{A.22}$$

The output of the measurement can simply be seen as a binary probability distribution and hence,

$$\sup_{\substack{\rho,\sigma \\ E_{e^\epsilon}(\rho\|\sigma)=0 \\ E_{e^\epsilon}(\sigma\|\rho)=0}} E_1(\rho\|\sigma) = \sup_{\substack{p,q \\ E_{e^\epsilon}(p\|q)=0 \\ E_{e^\epsilon}(q\|p)=0}} E_1(p\|q) = \frac{e^{-\epsilon}(e^\epsilon - 1)^2}{e^\epsilon - e^{-\epsilon}}, \tag{A.23}$$

where p, q are said binary probability distributions and the final equality was shown in [2, Equation (43)]. Inserting this back into Equation (A.19) gives the claimed result. □

E. Proof of Proposition III.5

Proof. We start with,

$$\begin{aligned}\eta_{\mathrm{Tr}}(\mathcal{A}) &= \sup_{\Psi\perp\Phi} E_1(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi)) &\text{(A.24)}\\ &\leq \sup_{\Psi\perp\Phi} \sqrt{1 - F(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi))^2}, &\text{(A.25)}\\ &\leq \sup_{\Psi\perp\Phi} \sqrt{H_{\frac{1}{2}}(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi)) - \frac{1}{4}H_{\frac{1}{2}}(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi))^2}, &\text{(A.26)}\end{aligned}$$

which follows from the Fuchs-van-de-Graaf inequality [12] and then the first inequality, $1 - F(\cdot,\cdot) \leq \frac{H_{\frac{1}{2}}(\cdot,\cdot)}{2}$, in Lemma II.1. As observed in [2], due to the monotonicity of $t \to t(1 - \frac{1}{4}t)$ in $[0, 2]$, it is sufficient to continue with

$$\sup_{\Psi\perp\Phi} H_{\frac{1}{2}}(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi)) \leq 2\frac{(e^{\frac{1}{2}\epsilon} - 1)^2}{e^\epsilon - 1} \sup_{\Psi\perp\Phi} E_1(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi)), \tag{A.27}$$

which follows from Lemma III.3 because $\mathcal{A}$ being LDP implies that $D_\infty(\mathcal{A}(\rho)\|\mathcal{A}(\sigma)) \leq \epsilon$ and $D_\infty(\mathcal{A}(\sigma)\|\mathcal{A}(\rho)) \leq \epsilon$. By then applying Lemma III.4, we can further bound this by,

$$\begin{aligned}\sup_{\Psi\perp\Phi} H_{\frac{1}{2}}(\mathcal{A}(\Psi)\|\mathcal{A}(\Phi)) &\leq 2\frac{(e^{\frac{1}{2}\epsilon} - 1)^2}{e^\epsilon - 1}\frac{e^{-\epsilon}(e^\epsilon - 1)^2}{e^\epsilon - e^{-\epsilon}} &\text{(A.28)}\\ &= 2\frac{(e^{\frac{1}{2}\epsilon} - 1)^2(1 - e^{-\epsilon})}{e^\epsilon - e^{-\epsilon}}. &\text{(A.29)}\end{aligned}$$

Inserting this back into Equation (A.26) gives, after some calculation,

$$\eta_{\mathrm{Tr}}(\mathcal{A}) \leq \frac{e^\epsilon - 1}{e^\epsilon + 1}, \tag{A.30}$$

which is was we set out to prove. □

F. Proof of Lemma III.7

Proof. Let $\sigma = \sum_i p_i |i\rangle\langle i|$. We observe the following,

$$\chi^2(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \tag{A.31}$$

$$= \int_0^\infty \mathrm{Tr}[(\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(\sigma) + s\mathbb{1})^{-1}(\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(\sigma) + s\mathbb{1})^{-1}]\mathrm{d}s \tag{A.32}$$

$$\leq \int_0^\infty \mathrm{Tr}[\sum_{i,j} p_i p_j (\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}(\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(|j\rangle\langle j|) + s\mathbb{1})^{-1}]\mathrm{d}s \tag{A.33}$$

$$\leq \sum_i p_i \int_0^\infty \mathrm{Tr}[(\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}(\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}]\mathrm{d}s \tag{A.34}$$

$$\leq \max_i \int_0^\infty \mathrm{Tr}[(\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}(\mathcal{N}(\rho) - \mathcal{N}(\sigma))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}]\mathrm{d}s, \tag{A.35}$$

where the first inequality is operator convexity of x^{-1} and the second for x^2 Now, define the replacer channel

$$\mathcal{R}(\cdot) = \mathcal{N}(|i\rangle\langle i|)\,\mathrm{Tr}(\cdot), \tag{A.36}$$

and set $\rho - \sigma = X = X_+ - X_-$, with

$$X_+ = \sum_n p_n |n\rangle\langle n|, \qquad X_- = \sum_m p_m |m\rangle\langle m|. \tag{A.37}$$

Finally define

$$\hat{X}_+ = \frac{X_+}{\mathrm{Tr}\, X_+}, \qquad \hat{X}_- = \frac{X_-}{\mathrm{Tr}\, X_-}, \tag{A.38}$$

where $\mathrm{Tr}\, X_+ = \mathrm{Tr}\, X_- = E_1(\rho\|\sigma)$. Note that

$$\mathcal{N}(\rho - \sigma) = (\mathcal{N} - \mathcal{R})(\rho - \sigma) = (\mathcal{N} - \mathcal{R})(X_+ - X_-) = (\mathcal{N} - \mathcal{R})((\mathrm{Tr}\, X_+)\hat{X}_+ - (\mathrm{Tr}\, X_-)\hat{X}_-) \tag{A.39}$$

$$= E_1(\rho\|\sigma)(\mathcal{N} - \mathcal{R})(\hat{X}_+ - \hat{X}_-). \tag{A.40}$$

With this, we get

$$\chi^2(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq E_1(\rho\|\sigma)^2 \max_i \int_0^\infty \mathrm{Tr}\left[\left(((\mathcal{N} - \mathcal{R})(\hat{X}_+ - \hat{X}_-))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}\right)^2\right]\mathrm{d}s \tag{A.41}$$

$$\leq E_1(\rho\|\sigma)^2 \max_i \left(\int_0^\infty \mathrm{Tr}\left[\left(((\mathcal{N} - \mathcal{R})(\hat{X}_+))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}\right)^2\right]\mathrm{d}s \right. \tag{A.42}$$

$$\left. + \int_0^\infty \mathrm{Tr}\left[\left(((\mathcal{N} - \mathcal{R})(\hat{X}_-))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}\right)^2\right]\mathrm{d}s\right), \tag{A.43}$$

$$\leq E_1(\rho\|\sigma)^2 \max_i \left(\sum_n p_n \int_0^\infty \mathrm{Tr}\left[\left(((\mathcal{N} - \mathcal{R})(|n\rangle\langle n|))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}\right)^2\right]\mathrm{d}s \right. \tag{A.44}$$

$$\left. + \sum_m p_m \int_0^\infty \mathrm{Tr}\left[\left(((\mathcal{N} - \mathcal{R})(|m\rangle\langle m|))(\mathcal{N}(|i\rangle\langle i|) + s\mathbb{1})^{-1}\right)^2\right]\mathrm{d}s\right), \tag{A.45}$$

$$\leq 2E_1(\rho\|\sigma)^2 \max_{\Psi,\Phi} \int_0^\infty \mathrm{Tr}\left[\left((\mathcal{N}(\Psi) - \mathcal{N}(\Phi))(\mathcal{N}(\Phi) + s\mathbb{1})^{-1}\right)^2\right]\mathrm{d}s \tag{A.46}$$

$$= 2E_1(\rho\|\sigma)^2 \max_{\Psi,\Phi} \chi^2(\mathcal{N}(\Psi)\|\mathcal{N}(\Phi)), \tag{A.47}$$

where the second inequality is by dropping all negative terms, the third by convexity and the forth by optimizing over general pure states. □