

Tight relations and equivalences between smooth relative entropies

Bartosz Regula, Ludovico Lami, and Nilanjana Datta

Abstract—The precise one-shot characterisation of operational tasks in classical and quantum information theory relies on different forms of smooth entropic quantities. A particularly important connection is between the hypothesis testing relative entropy and the smoothed max-relative entropy, which together govern many operational settings. We first strengthen this connection into a type of equivalence: we show that the hypothesis testing relative entropy is equivalent to a variant of the smooth max-relative entropy based on the information spectrum divergence, which can be alternatively understood as a measured smooth max-relative entropy. Furthermore, we improve a fundamental lemma due to Datta and Renner that connects the different variants of the smoothed max-relative entropy, introducing a modified proof technique based on matrix geometric means and a tightened gentle measurement lemma. We use the unveiled connections and tools to strictly improve on previously known one-shot bounds and duality relations between the smooth max-relative entropy and the hypothesis testing relative entropy, establishing provably tight bounds between them. The results then allow us to refine other divergence inequalities, in particular sharpening bounds that connect the max-relative entropy with Rényi divergences.

I. INTRODUCTION

SMOOTH relative entropies [1] were defined to address the need for a precise understanding of the performance of various operational protocols in information theory in settings beyond the asymptotic i.i.d. one, where it is assumed that the underlying resources (sources, channels, or entangled states) employed in the protocols are available for asymptotically many uses. In this asymptotic setting, unique entropic quantities naturally emerge as the optimal rates of information-theoretic tasks. In contrast, when the assumptions of the asymptotic i.i.d. setting are lifted, i.e. in the so-called *one-shot information theory*, several different, inequivalent types of smooth entropies are required to fully characterise the performance of different tasks. Similarly to how many problems in classical information theory can be broadly divided into

two types, packing-type and covering-type problems, the more general setting of quantum information can also be categorised in a similar way. On one side, there are problems that can be connected with hypothesis testing, including quantum channel coding [2], [3], [4], [5], [6], quantum data compression [7], [6], and quantum resource distillation [8], [9], [10]. Due to the underlying hypothesis testing structure, the one-shot performance of these protocols can be related with a quantity known as the *hypothesis testing relative entropy* D_H^ε [2], [3]. On the other side, covering-type problems include quantum channel simulation [11], [12], privacy amplification [1], [13], [14], decoupling [15], convex split [16], and quantum resource dilution [8], [9]. The one-shot characterisation of such protocols typically relies on a quantity known as the *smooth max-relative entropy* D_{\max}^ε [1], [17]. Understanding the precise relations between the two smooth relative entropies has thus been an important problem in the one-shot characterisation of quantum information theory [18], [13], [19], [20], [21].

Despite their many differences, the hypothesis testing relative entropy and the max-relative entropy are known to be quantitatively related, albeit in a complementary fashion. Informally, the value of D_{\max}^ε closely matches that of $D_H^{1-\varepsilon}$, up to asymptotically negligible factors [13], [18]. This means that the two smooth quantities satisfy a ‘weak/strong converse duality’: understanding the behaviour of D_H^ε in the small ε regime (weak converse) closely mirrors the understanding the large ε (strong converse) behaviour of D_{\max}^ε , and vice versa. This fact was crucial in obtaining results such as asymptotic equipartition theorems for smooth entropic quantities [22], [23], [24], [18], [25], [26], [27] or in characterising the higher-order corrections of optimal rates in quantum information tasks [13], [28]. However, the precise one-shot bounds that connected D_{\max}^ε and $D_H^{1-\varepsilon}$ in the literature are often not tight, motivating us to look for alternative approaches and stronger links.

II. SUMMARY OF MAIN RESULTS

In this paper, we introduce new techniques for the study of the relations between D_{\max}^ε and $D_H^{1-\varepsilon}$, strengthening both the conceptual and the quantitative connections. The framework introduced here is strictly stronger than all prior techniques: we strengthen or recover all previously known bounds, and many of our contributions are major improvements over the state of the art, yielding provably tight bounds. Our approach employs an intermediate quantity, $\tilde{D}_{\max}^\varepsilon$, which is related to the smooth max-relative entropy and was originally introduced in [28] as a type of information spectrum divergence [13] (see Section III

B. Regula is with the RIKEN Pioneering Research Institute (PRI) and RIKEN Center for Quantum Computing (RQC), Wako, Saitama 351-0198, Japan. He was partially supported by the Japan Science and Technology Agency (JST) PRESTO grant no. JPMJPR25FB. Email: bartosz.regula@gmail.com.

L. Lami was with QuSoft, Korteweg–de Vries Institute for Mathematics, and Institute for Theoretical Physics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands. He is now with Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy. He was supported by MIUR (Ministero dell’Istruzione, dell’Università e della Ricerca) through the project ‘Dipartimenti di Eccellenza 2023–2027’ of the ‘Classe di Scienze’ department at the Scuola Normale Superiore and by the European Union under the European Research Council (ERC Grant Agreement No. 101165230). Email: ludovico.lami@gmail.com.

N. Datta is with the Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, United Kingdom. Email: n.datta@damtp.cam.ac.uk.

for definitions). This variant of the max-relative entropy was previously used as a technical tool in many proofs and also found use in characterising quantum differential privacy. Our closer investigation in Section V reveals that it is intrinsically connected to both D_{\max}^ε and $D_H^{1-\varepsilon}$. Specifically, we show that $\tilde{D}_{\max}^\varepsilon$ corresponds to a measured variant of the smoothed max-relative entropy (Proposition 3), and that it is in fact equivalent to $D_H^{1-\varepsilon}$ in a precise sense: in Theorem 4 we prove that, for all pairs of states ρ and σ and for all $\varepsilon \in (0, 1)$,

$$\begin{aligned} D_H^{1-\varepsilon}(\rho\|\sigma) &= \inf_{\mu \in (0, \varepsilon]} \left[\tilde{D}_{\max}^{\varepsilon-\mu}(\rho\|\sigma) + \log \frac{1}{\mu} \right], \\ \tilde{D}_{\max}^{1-\varepsilon}(\rho\|\sigma) &= \sup_{\mu \in (0, \varepsilon]} \left[D_H^{\varepsilon-\mu}(\rho\|\sigma) - \log \frac{1}{\mu} \right], \end{aligned} \quad (1)$$

implying that one can reconstruct the hypothesis testing relative entropy function $D_H^{1-\varepsilon}$ from $\tilde{D}_{\max}^\varepsilon$, and vice versa.

This finding allows us to establish new tight bounds between the hypothesis testing relative entropy and $\tilde{D}_{\max}^\varepsilon$ (Lemma 7). To relate the latter with the operationally important quantity D_{\max}^ε , previous works relied on an important lemma originally shown by Datta and Renner [29], which, however, is not tight for several types of distance measures employed in the smoothing. We introduce an improved proof technique based on the operator geometric mean that leads to a tightening of the previous statement, in particular when the smoothing is over normalised quantum states. Namely, in Theorem 5 we show that given a state ρ and two positive semi-definite operators $A, Q \geq 0$ with $\text{Tr } Q \leq \varepsilon < 1$, if ρ is approximately dominated by A with ‘remainder’ Q , in the sense that the operator inequality $\rho \leq A + Q$ holds, then we can find a smoothing of ρ that is *exactly* dominated by A up to a small rescaling: formally, we can find a normalised state ρ' such that

$$F(\rho, \rho') \geq 1 - \varepsilon, \quad \frac{1}{2} \|\rho - \rho'\|_1 \leq \sqrt{\varepsilon}, \quad \rho' \leq \frac{A}{1 - \varepsilon}. \quad (2)$$

We also obtain an improved statement of this result for smoothing over subnormalised quantum states, which involves a tightening of another key tool in quantum information theory, the gentle measurement lemma (see Lemma 6). Our proof method is also easy to generalise, which we exemplify with a multi-partite extension of the lemma to a setting of simultaneous state smoothing (Corollary 19).

We apply our results to obtain a set of tight bounds that elucidate the duality relations between D_{\max}^ε and $D_H^{1-\varepsilon}$ in Section VII. We consider both of the commonly used definitions of smoothing in quantum information — based on either the trace distance or the purified distance — and in the process characterise the connections and differences between them. Our main result in Theorem 12 shows that the following relations hold for all states ρ and σ , all $\varepsilon \in (0, 1)$, and all $\mu \in (0, \varepsilon]$: for the trace distance smoothing, we have

$$\begin{aligned} D_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma) + \log \frac{1}{\varepsilon} &\leq D_H^{1-\varepsilon}(\rho\|\sigma) \\ &\leq D_{\max}^{\varepsilon-\mu}(\rho\|\sigma) + \log \frac{1}{\mu}, \end{aligned} \quad (3)$$

and for the purified smoothing we get

$$\begin{aligned} D_{\max}^{\sqrt{\varepsilon}}(\rho\|\sigma) + \log \frac{1}{\varepsilon} &\leq D_H^{1-\varepsilon}(\rho\|\sigma) \\ &\leq D_{\max}^{\sqrt{\varepsilon-\mu}}(\rho\|\sigma) + \log \frac{F_2(1-\varepsilon, \varepsilon-\mu)}{\mu^2}, \end{aligned} \quad (4)$$

where $F_2(p, q) := \left(\sqrt{pq} + \sqrt{(1-p)(1-q)} \right)^2 \leq 1$ denotes the binary fidelity. Both of the lower bounds on $D_H^{1-\varepsilon}$ are a significant improvement over previously known one-shot relations, and our purified distance upper bound in (4) also refines all known results of this type. Crucially, all of the inequalities are tight: none of the ε dependencies can be improved in general, with the lower bounds tight in the case $\sigma = \rho$, the upper bound of (3) tight for classical (commuting) states, and the upper bound of (4) tight for pure states.

Going further, in Section VIII-A we employ our techniques to strengthen also the bounds connecting the smooth max-relative entropy with the Rényi divergences, leading also to a bound for the hypothesis testing relative entropy. To wit, in Corollaries 14 and 15 we show that, for all $\varepsilon \in (0, 1)$, all $\alpha > 1$, and all $\beta \in (0, 1)$,

$$\begin{aligned} D_{\max}^\varepsilon(\rho\|\sigma) &\leq \tilde{D}_\alpha(\rho\|\sigma) + \frac{1}{\alpha-1} \log \frac{1}{\varepsilon^2}, \\ D_H^\varepsilon(\rho\|\sigma) &\geq D_\beta(\rho\|\sigma) - \frac{\beta}{1-\beta} \log \frac{1}{\varepsilon} + \log \frac{1}{1-\varepsilon}, \end{aligned} \quad (5)$$

where \tilde{D}_α and D_β denote, respectively, the sandwiched Rényi relative entropy (given by (106)) and the Petz–Rényi relative entropy (105). The bounds obtained in this way give tight constraints on the asymptotic expansion of several of the smooth divergence variants discussed here.

We also obtain other auxiliary results which may be of independent interest: (i) improved connections between the hypothesis testing relative entropy D_H^ε and the information spectrum divergence D_s^ε (Proposition 17); (ii) a tightening of the so-called quantum substate theorem (Corollary 20); (iii) a variant (Theorem 21) of a recent important result, namely, Frenkel’s integral representation for the Umegaki relative entropy, but with the integrand being a function of $\tilde{D}_{\max}^\varepsilon$, and (iv) a weak/strong converse duality relation between $D_H^{1-\varepsilon}$ and the Hilbert projective metric (Corollary D.2).

III. SMOOTH DIVERGENCES

A. Hypothesis testing relative entropy and max-relative entropy

We use Greek letters (ρ, σ) to denote quantum states, which we understand to be positive semidefinite operators of trace one acting on a finite-dimensional Hilbert space. We will sometimes specialise our results to classical probability distributions, which are denoted by Latin letters (p, q) ; we can equivalently interpret them as commuting quantum states — that is, states that are diagonal in a common basis. All logarithms are to base 2 unless otherwise stated.

Smooth divergences are broadly understood as divergences (relative entropies) between quantum states that incorporate

the allowance for some form of error or uncertainty about the states under consideration, quantified by a smoothing parameter ε .

Perhaps the most fundamental divergence in one-shot quantum information theory is the ***hypothesis testing relative entropy***, defined for two quantum states ρ and σ as [2], [3]

$$D_H^\varepsilon(\rho\|\sigma) := -\log \inf \left\{ \text{Tr } M\sigma \mid 0 \leq M \leq \mathbb{1}, \text{Tr}(\mathbb{1}-M)\rho \leq \varepsilon \right\}, \quad (6)$$

where $\varepsilon \in [0, 1]$. Equivalently, $D_H^\varepsilon(\rho\|\sigma) = -\log \beta_\varepsilon(\rho, \sigma)$, where $\beta_\varepsilon(\rho, \sigma)$ denotes the optimal probability of the type II error of hypothesis testing between ρ and σ when the type I error probability is constrained to be at most ε . The divergence is conceptually very closely related to smooth relative entropies, although it is defined slightly differently than conventional smooth divergences [1] — rather than employing an optimisation over a neighbourhood of states, it can be regarded a smoothed variant of the Petz–Rényi divergence of order 0 (min-relative entropy [17]) under a notion of ‘operator smoothing’ [3].

Another fundamental quantity that we focus on here is the ***max-relative entropy*** [17]

$$D_{\max}(\rho\|\sigma) := \log \inf \left\{ \lambda \geq 0 \mid \rho \leq \lambda\sigma \right\}, \quad (7)$$

where $\inf \emptyset = \infty$ and we take $\log(\infty) := \infty$ for consistency. This can be understood as the sandwiched Rényi divergence \tilde{D}_α of order ∞ [30]. The smooth variant of this quantity is defined not by evaluating $D_{\max}(\rho\|\sigma)$ exactly, but instead by minimising over all states $\rho' \approx_\varepsilon \rho$ in an ε -ball around ρ [1], [17]. This definition crucially depends on how exactly we define the smoothing neighbourhood, and in particular on how we quantify the distance between quantum states.

Two of the most fundamental measures of distance are the ***trace (total variation) distance*** $\frac{1}{2} \|\rho - \rho'\|_1$, where $\|X\|_1 = \text{Tr} \sqrt{X^\dagger X}$ is the Schatten 1-norm, as well as the ***purified distance*** $P(\rho, \rho') := \sqrt{1 - F(\rho, \rho')}$, where $F(\rho, \rho')$ denotes the Uhlmann fidelity

$$F(\rho, \rho') := \left\| \sqrt{\rho} \sqrt{\rho'} \right\|_1^2 = \left(\text{Tr} \sqrt{\sqrt{\rho} \rho' \sqrt{\rho}} \right)^2. \quad (8)$$

The definition of the distance measures is often extended to cover the case where ρ' is not necessarily a normalised quantum state, but rather a *subnormalised* positive operator with $\text{Tr} \rho' \leq 1$. As long as ρ is normalised, the definition of fidelity in Eq. (8) does not change [25]; however, the

form of the trace distance needs to be adjusted to account for subnormalised states. For ρ and ρ' with $\text{Tr} \rho = 1$ and $\text{Tr} \rho' \leq 1$, we define the ***generalised trace distance*** as [25]

$$\|\rho - \rho'\|_+ := \frac{1}{2} \|\rho - \rho'\|_1 + \frac{1}{2} \text{Tr}(\rho - \rho') = \text{Tr}(\rho - \rho')_+, \quad (9)$$

where $\text{Tr}(\cdot)_+$ denotes the trace of the positive part of a Hermitian operator. Using a variational characterisation of $\text{Tr}(\cdot)_+$, we can rewrite this as

$$\begin{aligned} \|\rho - \rho'\|_+ &= \max_{0 \leq M \leq \mathbb{1}} \text{Tr} [M(\rho - \rho')] \\ &= \min \left\{ \text{Tr } X \mid \rho - \rho' \leq X, X \geq 0 \right\}, \end{aligned} \quad (10)$$

where the formulation as a maximum neatly expresses the quantum Neyman–Pearson lemma [31], while the dual formulation as a minimum will often find use in this paper.

We then define the ***smooth max-relative entropy*** as

$$D_{\max}^{\varepsilon, \Delta}(\rho\|\sigma) := \inf_{\rho' \in \mathcal{B}_\Delta^\varepsilon(\rho)} D_{\max}(\rho'\|\sigma), \quad (11)$$

where Δ denotes one of the choices of smoothing defined in Table I.

To relate the two types of smoothing based on trace and purified distance, from the Fuchs–van de Graaf inequalities [32], [25]

$$1 - \sqrt{F(\rho, \rho')} \leq \|\rho - \rho'\|_+ \leq \sqrt{1 - F(\rho, \rho')}, \quad (12)$$

one derives immediately that

$$D_{\max}^{\sqrt{\varepsilon(2-\varepsilon)}, P, =}(\rho\|\sigma) \leq D_{\max}^{\varepsilon, T, =}(\rho\|\sigma) \leq D_{\max}^{\varepsilon, P, =}(\rho\|\sigma) \quad (13)$$

and likewise for the subnormalised variants.

The smoothed quantities D_H^ε and D_{\max}^ε are not a priori related. A fundamental connection between them arises from the fact that their difference becomes negligible when one considers many i.i.d. copies of quantum states, and it is known that [33], [34], [22]

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} D_H^\varepsilon(\rho^{\otimes n} \|\sigma^{\otimes n}) &= \lim_{n \rightarrow \infty} \frac{1}{n} D_{\max}^{\varepsilon, P, =}(\rho^{\otimes n} \|\sigma^{\otimes n}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} D_{\max}^{\varepsilon, T, =}(\rho^{\otimes n} \|\sigma^{\otimes n}) \\ &= D(\rho\|\sigma) \end{aligned} \quad (14)$$

for all $\varepsilon \in (0, 1)$, where

$$D(\rho\|\sigma) := \text{Tr} \rho (\log \rho - \log \sigma) \quad (15)$$

TABLE I
TYPES OF SMOOTHING

Notation	Smoothing metric	Definition of smoothing ball
$T, =$	Trace distance, normalised	$\mathcal{B}_{T,=}^\varepsilon(\rho) := \left\{ \rho' \mid \frac{1}{2} \ \rho - \rho'\ _1 \leq \varepsilon, \rho' \geq 0, \text{Tr} \rho' = 1 \right\}$
T, \leq	Trace distance, subnormalised	$\mathcal{B}_{T,\leq}^\varepsilon(\rho) := \left\{ \rho' \mid \ \rho - \rho'\ _+ \leq \varepsilon, \rho' \geq 0, \text{Tr} \rho' \leq 1 \right\}$
$P, =$	Purified distance, normalised	$\mathcal{B}_{P,=}^\varepsilon(\rho) := \left\{ \rho' \mid \sqrt{1 - F(\rho, \rho')} \leq \varepsilon, \rho' \geq 0, \text{Tr} \rho' = 1 \right\}$
P, \leq	Purified distance, subnormalised	$\mathcal{B}_{P,\leq}^\varepsilon(\rho) := \left\{ \rho' \mid \sqrt{1 - F(\rho, \rho')} \leq \varepsilon, \rho' \geq 0, \text{Tr} \rho' \leq 1 \right\}$

denotes the (Umegaki) quantum relative entropy. In (15), one sets $D(\rho\|\sigma) = \infty$ if $\text{supp } \rho \not\subseteq \text{supp } \sigma$. This was later tightened to a one-shot relation in [13], namely

$$\begin{aligned} D_{\max}^{\sqrt{\varepsilon}, P, \leq}(\rho\|\sigma) - \log \frac{|\text{spec}(\sigma)|}{\varepsilon} \\ \leq D_H^{1-\varepsilon}(\rho\|\sigma) \\ \leq D_{\max}^{\sqrt{\varepsilon-\mu}, P, \leq}(\rho\|\sigma) + \log \frac{27(1-\varepsilon+\mu)}{\mu^3} \end{aligned} \quad (16)$$

for all $\varepsilon \in (0, 1)$ and all $\mu \in (0, \varepsilon]$. This showed that the values of the two divergences can be quantitatively related in complementary error regimes — a property that we refer to as the ‘weak/strong converse duality’ between D_H^ε and D_{\max}^ε . We note that this terminology is not necessarily meant to imply a rigorous connection with an operational task, but rather to merely evoke the intuitive idea that the weak converse regime is concerned with small ε while the strong converse regime with large ε . These bounds were subsequently improved in many works [18], [19], [20], but the problem of finding tight one-shot relations and bounds remained open.

B. Modified max-relative entropy $\tilde{D}_{\max}^\varepsilon$

An important quantity in our approach will be a variant of the smooth max-relative entropy, formalised by Datta and Leditzky [28] as

$$\tilde{D}_{\max}^\varepsilon(\rho\|\sigma) := \log \inf \left\{ \lambda \geq 0 \mid \text{Tr}(\rho - \lambda\sigma)_+ \leq \varepsilon \right\}. \quad (17)$$

In fact, it was first introduced as an alternative to the information spectrum divergence D_s^ε [13] and denoted $\overline{D}_s^\varepsilon$ in [28]. Here we prefer to relate it to the max-relative entropy, the connection with which becomes clearer once we use the variational form of $\text{Tr}(\cdot)_+$ to write [35]

$$\tilde{D}_{\max}^\varepsilon(\rho\|\sigma) = \log \inf \left\{ \lambda \mid \rho \leq \lambda\sigma + Q, Q \geq 0, \text{Tr } Q \leq \varepsilon \right\}, \quad (18)$$

which more closely resembles the optimisation problem that defines D_{\max} . Indeed, we notice that $\tilde{D}_{\max}^0(\rho\|\sigma) = D_{\max}(\rho\|\sigma)$, which justifies regarding $\tilde{D}_{\max}^\varepsilon$ as a smoothed variant of D_{\max} .

As we show in detail in Appendix B, for any two fixed states ρ, σ the function $\varepsilon \mapsto \tilde{D}_{\max}^\varepsilon(\rho\|\sigma)$ is continuous and strictly decreasing on the interval $\varepsilon \in (1 - \text{Tr } \rho\Pi_\sigma, 1]$, where Π_σ denotes the projector onto the support of σ .

The modified max-relative entropy $\tilde{D}_{\max}^\varepsilon$ made implicit appearances in many early works on quantum hypothesis testing already [34], [36], [23], [37] through its connection with the information spectrum method [36], [29]. By Eq. (17), it can be understood as a generalised inverse function of the hockey-stick divergence $E_\lambda(\rho\|\sigma) := \text{Tr}(\rho - \lambda\sigma)_+$, which has also found applications in various aspects of one-shot information theory [38], [39], [40], [41], [42]. $\tilde{D}_{\max}^\varepsilon$ was first formalised as a divergence in [28], where it was used to study second-order expansions for optimal rates of certain quantum information-theoretic tasks in the i.i.d. setting. Some of its basic properties, including the data processing inequality, were established there. The classical analogue of this quantity independently found use in the study of differential

privacy [43], [44], and a generalisation of these results later also led to $\tilde{D}_{\max}^\varepsilon$ playing a role in characterising quantum differential privacy [35].¹ A conditional entropy variant of the divergence also made appearances in the study of classical privacy amplification [46], [47], where it was connected with the smooth min-entropy [48].

Although $\tilde{D}_{\max}^\varepsilon$ was employed mostly as a technical tool in one-shot quantum information theory, we will show that the quantity has many precise connections with both the hypothesis testing relative entropy D_H^ε and with the standard smoothed max-relative entropy $D_{\max}^{\varepsilon, T, =}$ that were not previously known, ultimately leading to tighter bounds between all of these quantities.

IV. ON NORMALISED AND SUBNORMALISED SMOOTHING

As the first step of our investigation, we will clarify a basic property of the smooth divergences that will help consolidate the many possible choices of smoothing encountered in the literature.

A natural question in the definition of the smooth max-relative entropy is why the ostensibly less physical subnormalised states are considered among the smoothing variants. In operational settings, it may indeed seem desirable to optimise only over normalised states, as these are the only states that can result from the application of a deterministic state transformation (quantum channel). However, some technical benefits can be gained by relaxing the smoothing to subnormalised states — notably, the invariance of the resulting conditional entropic quantities under isometries, which leads to useful duality relations [49], [25] — making this a frequently made choice in many works on one-shot quantum information theory.

We can show, however, that the two smoothing notions are equivalent whenever the normalised smooth max-relative entropy is non-zero, or, equivalently, when the subnormalised quantity takes positive values. Naturally, most of the questions about the behaviour of these quantities is concerned with regimes in which they are larger than zero, and hence it is useful to realise that the dichotomy between the normalised and subnormalised definitions is effectively irrelevant in such cases.

Lemma 1. *For any two quantum states ρ and σ and any $\varepsilon \in [0, 1]$, it holds that*

$$\begin{aligned} D_{\max}^{\varepsilon, T, =}(\rho\|\sigma) &= \max \left\{ D_{\max}^{\varepsilon, T, \leq}(\rho\|\sigma), 0 \right\}, \\ D_{\max}^{\varepsilon, P, =}(\rho\|\sigma) &= \max \left\{ D_{\max}^{\varepsilon, P, \leq}(\rho\|\sigma), 0 \right\}. \end{aligned} \quad (19)$$

As a consequence, $D_{\max}^{\varepsilon, T, =}(\rho\|\sigma) = D_{\max}^{\varepsilon, T, \leq}(\rho\|\sigma)$ if and only if $\varepsilon \leq \frac{1}{2} \|\rho - \sigma\|_1$, and $D_{\max}^{\varepsilon, P, =}(\rho\|\sigma) = D_{\max}^{\varepsilon, P, \leq}(\rho\|\sigma)$ if and only if $\varepsilon \leq P(\rho, \sigma)$.

Proof. Note first that, since $D_{\max}(\rho\|\sigma) \geq 0$ for any normalised operators ρ and σ , it must be the case that $D_{\max}^{\varepsilon, T, =}(\rho\|\sigma)$ and $D_{\max}^{\varepsilon, P, =}(\rho\|\sigma)$ are non-negative. Now, if

¹The use of $\tilde{D}_{\max}^\varepsilon$ in the context of quantum differential privacy can in fact be traced back to the earlier works [45], [41]. In particular, some results of [41] are stated in terms of an inverse function of E_λ . The same quantity implicitly appeared in [45], although it is not immediately clear that the definition there corresponds to $\tilde{D}_{\max}^\varepsilon$. We clarify the equivalence of the different definitions in Appendix A.

$D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma) < 0$, then there exists a subnormalised state ρ' such that $\rho' < \sigma$, where we assumed that $\sigma > 0$ since we can always restrict the space down to the support of σ without loss of generality. This implies that $\rho - \sigma < \rho - \rho' \leq (\rho - \rho')_+$. But then the assumption that $\|\rho - \rho'\|_+ \leq \varepsilon$ means that $\frac{1}{2}\|\rho - \sigma\|_1 = \text{Tr}(\rho - \sigma)_+ < \text{Tr}(\rho - \rho')_+ \leq \varepsilon$, and so $D_{\max}^{\varepsilon, T, =}(\rho \| \sigma) \leq D_{\max}(\sigma \| \sigma) = 0$. In the case of $D_{\max}^{\varepsilon, P, \leq}(\rho \| \sigma)$ we similarly have that $\rho' < \sigma$ for some subnormalised state such that $F(\rho, \rho') \geq 1 - \varepsilon^2$. Using the strict operator monotonicity of the square root (see e.g. [50, Proposition 1.2.9]), we get that

$$F(\rho, \sigma) = \left(\text{Tr} \sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right)^2 > \left(\text{Tr} \sqrt{\sqrt{\rho}\rho'\sqrt{\rho}} \right)^2 = F(\rho, \rho'), \quad (20)$$

so that again $D_{\max}^{\varepsilon, P, =}(\rho \| \sigma) \leq D_{\max}(\sigma \| \sigma) = 0$.

We can in fact make a stronger statement here, namely, $D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma) < 0 \iff \varepsilon > \frac{1}{2}\|\rho - \sigma\|_1$, and analogously for the purified distance. We have already shown the \Rightarrow implication. To see the other direction, assume that $\varepsilon > \frac{1}{2}\|\rho - \sigma\|_1$. Clearly, $D_{\max}^{\varepsilon, T, =}(\rho \| \sigma) = 0$ because $\sigma \in \mathcal{B}_{T, =}^{\varepsilon}(\rho)$, so $D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma)$ can be at most zero. But if it were exactly zero, then we would have that $\rho' \leq \sigma$ for some subnormalised state ρ' with $\|\rho - \rho'\|_+ \leq \varepsilon$, and hence $\frac{1}{2}\|\rho - \sigma\|_1 \leq \text{Tr}(\rho - \rho')_+ \leq \varepsilon$, contradicting our assumption. Hence $D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma) < 0$.

We will now show that if $D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma) \geq 0$, then $D_{\max}^{\varepsilon, T, =}(\rho \| \sigma) = D_{\max}^{\varepsilon, T, =}(\rho \| \sigma)$. Let ρ' be a feasible subnormalised state such that $\rho' \leq \lambda\sigma$ and $\|\rho - \rho'\|_+ \leq \varepsilon$ for some $\lambda \geq 1$. To avoid trivial cases we will assume that $\text{Tr} \rho' < 1$. Denoting by $\{|i\rangle\}_i$ an orthonormal basis that diagonalises the operator $\lambda\sigma - \rho'$, let $\rho'_{ij} := \langle i | \rho' | j \rangle$ and $\sigma_{ij} := \langle i | \sigma | j \rangle$. Define the sets of indices

$$S := \left\{ i \mid \sigma_{ii} < \rho'_{ii} \leq \lambda\sigma_{ii} \right\}, \quad S^\perp := \left\{ i \mid \sigma_{ii} \geq \rho'_{ii} \right\}. \quad (21)$$

For some value of $\mu \in [0, 1]$ to be fixed later, we now define

$$\begin{aligned}
 \rho'' &:= \rho' + \mu \sum_{i \in S^\perp} (\sigma_{ii} - \rho'_{ii}) |i\rangle\langle i| \\
 &\stackrel{(i)}{\leq} \rho' + \sum_{i \in S^\perp} (\sigma_{ii} - \rho'_{ii}) |i\rangle\langle i| \\
 &= \sum_{i \in S} \rho'_{ii} |i\rangle\langle i| + \sum_{i \neq j} \rho'_{ij} |i\rangle\langle j| + \sum_{i \in S^\perp} \sigma_{ii} |i\rangle\langle i| \quad (22) \\
 &\stackrel{(ii)}{\leq} \sum_{i \in S} \lambda\sigma_{ii} |i\rangle\langle i| + \sum_{i \neq j} \lambda\sigma_{ij} |i\rangle\langle j| + \sum_{i \in S^\perp} \sigma_{ii} |i\rangle\langle i| \\
 &\stackrel{(iii)}{\leq} \lambda\sigma,
 \end{aligned}$$

where: (i) follows since $\mu \leq 1$ and $\sigma_{ii} \geq \rho'_{ii}$ for all $i \in S^\perp$ by definition; (ii) follows as $\rho'_{ii} \leq \lambda\sigma_{ii}$ for all $i \in S$, and $\lambda\sigma_{ij} = \rho'_{ij}$ because $\{|i\rangle\}_i$ diagonalises $\lambda\sigma - \rho'$; finally, (iii) is simply since $\lambda \geq 1$.

Observe that there must exist at least one element in S^\perp such that $\rho'_{ii} < \sigma_{ii}$, as otherwise our assumption that $\text{Tr} \rho' < 1$

would be contradicted. We can then always choose a value of μ so that $\text{Tr} \rho'' = 1$, namely

$$\mu := \frac{1 - \text{Tr} \rho'}{\sum_{i \in S^\perp} (\sigma_{ii} - \rho'_{ii})}. \quad (23)$$

The fact that $\mu \leq 1$ can be seen by noticing that

$$1 - \text{Tr} \rho' = \sum_{i \in S \cup S^\perp} (\sigma_{ii} - \rho'_{ii}) \leq \sum_{i \in S^\perp} (\sigma_{ii} - \rho'_{ii}) \quad (24)$$

where we used that $\sigma_{ii} < \rho'_{ii}$ for all $i \in S$.

Now, clearly, $\rho'' \geq \rho'$. This implies that $\rho - \rho'' \leq \rho - \rho'$ and hence that

$$\frac{1}{2}\|\rho - \rho''\|_1 = \text{Tr}(\rho - \rho'')_+ \leq \text{Tr}(\rho - \rho')_+ \leq \varepsilon. \quad (25)$$

Thus $D_{\max}^{\varepsilon, T, =}(\rho \| \sigma) \leq D_{\max}(\rho'' \| \sigma) \leq \log \lambda$ by (22). Since this holds for all feasible λ , we have that $D_{\max}^{\varepsilon, T, =}(\rho \| \sigma) \leq D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma)$ and thus the two quantities must be equal.

The proof for the purified distance is completely analogous. We now start with a subnormalised state ρ' such that $\sqrt{1 - F(\rho, \rho')} \leq \varepsilon$ and using again the operator monotonicity of the square root we obtain

$$F(\rho, \rho'') = \left(\text{Tr} \sqrt{\sqrt{\rho}\rho''\sqrt{\rho}} \right)^2 \geq \left(\text{Tr} \sqrt{\sqrt{\rho}\rho'\sqrt{\rho}} \right)^2 = F(\rho, \rho'), \quad (26)$$

from which we get that $D_{\max}^{\varepsilon, P, =}(\rho \| \sigma) \leq D_{\max}(\rho'' \| \sigma) \leq \log \lambda$. \square

Another way to relate the normalised and subnormalised quantities without having to distinguish the cases when the subnormalised smoothed divergences take negative values is to take any subnormalised state ρ' and normalise it as $\rho'' := \frac{\rho'}{\text{Tr} \rho'}$; it is not difficult to show that $\frac{1}{2}\|\rho - \rho''\|_1 \leq \|\rho - \rho'\|_+$ and $F(\rho, \rho'') \geq F(\rho, \rho')$, meaning that ρ'' is a feasible state for the normalised smooth D_{\max} . Accounting for the renormalisation factor, $\text{Tr} \rho'$, gives

$$\begin{aligned}
 D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma) &\leq D_{\max}^{\varepsilon, T, =}(\rho \| \sigma) \\
 &\leq D_{\max}^{\varepsilon, T, \leq}(\rho \| \sigma) + \log \frac{1}{1 - \varepsilon}, \quad (27)
 \end{aligned}$$

where we used that $\text{Tr} \rho' \geq 1 - \varepsilon$ for any ρ' with $\|\rho - \rho'\|_+ \leq \varepsilon$. The purified distance relation is analogous.

V. EQUIVALENCES

A. Classical $\tilde{D}_{\max}^\varepsilon$ and D_{\max}^ε

For classical systems, there is in fact a complete equivalence between the modified max-relative entropy $\tilde{D}_{\max}^\varepsilon$ and the standard smooth max-relative entropy D_{\max}^ε . This was first noticed in [43], [44], where a classical definition of $\tilde{D}_{\max}^\varepsilon$ was introduced in the context of characterising differential privacy.

Lemma 2 ([44], [47]). *For all classical probability distributions p, q (or commuting quantum states), we have*

$$\tilde{D}_{\max}^\varepsilon(p \| q) = D_{\max}^{\varepsilon, T, \leq}(p \| q). \quad (28)$$

As a consequence,

$$D_{\max}^{\varepsilon, T, =}(p \| q) = \max \left\{ \tilde{D}_{\max}^\varepsilon(p \| q), 0 \right\}. \quad (29)$$

The classical result in (29) was first shown in [44, Lemma 3.17] (see also [43]), without studying the subnormalised smoothing variant. It was previously claimed that (29) holds also in the quantum case [45], but the proof contains a gap [41] (and indeed this claim can be easily verified to be false numerically; see also Appendix A). The classical equality $D_{\max}^{\varepsilon, T, \leq} = \tilde{D}_{\max}^{\varepsilon}$ was also shown in [47, Proposition 2] (see also [51], [46]), without connecting it to the normalised $D_{\max}^{\varepsilon, T, =}$. For completeness, we give a proof here, essentially following [47].

Proof of Lemma 2. We will understand p and q as diagonal density operators $p = \sum_i p_i |i\rangle\langle i|$, $q = \sum_i q_i |i\rangle\langle i|$ in some orthonormal basis. Consider then any feasible solution for $\tilde{D}_{\max}^{\varepsilon}(p||q)$, that is, $p \leq \lambda q + Q$ for some positive semidefinite operator Q with $\text{Tr } Q \leq \varepsilon$. Without loss of generality, Q can be taken to be diagonal in the basis $\{|i\rangle\}$, since we can simply dephase any feasible Q in this basis while maintaining its feasibility. Furthermore, we can in fact assume that $p \geq Q$: should it be the case that $p_i < Q_i$ for some i , we can define $Q' \leq p$ through $Q'_i := \min\{Q_i, p_i\}$, which is feasible with the same objective value λ as

$$\lambda q_i \geq \max\{p_i - Q_i, 0\} = p_i - Q'_i \quad \forall i. \quad (30)$$

But then $p' := p - Q \leq \lambda q$ is a subnormalised state such that $\|p - p'\|_+ = \|Q\|_+ = \text{Tr } Q \leq \varepsilon$, implying that

$$D_{\max}^{\varepsilon, T, \leq}(p||q) \leq D_{\max}(p'||q) \leq \log \lambda. \quad (31)$$

Since this holds for any feasible solution λ , it follows that $D_{\max}^{\varepsilon, T, \leq}(p||q) \leq \tilde{D}_{\max}^{\varepsilon}(p||q)$. The opposite inequality between the quantities follows easily by noting that any feasible p' such that $p' \leq \lambda q$ and $\text{Tr}(p - p')_+ \leq \varepsilon$ gives a feasible solution for $\tilde{D}_{\max}^{\varepsilon}(p||q)$ through $p \leq p' + \text{Tr}(p - p')_+ \leq \lambda q + \text{Tr}(p - p')_+$. The quantities thus must be equal, and Eq. (29) then follows by Lemma 1. \square

B. $\tilde{D}_{\max}^{\varepsilon}$ as measured D_{\max}^{ε}

A natural way to define quantum divergences from classical ones is to first perform a measurement and then compute the classical divergence of the resulting probability distributions; optimising over all measurements then gives a well-defined quantum relative entropy [52], [33], [53]. Following this idea, we define the *measured smooth max-relative entropy* as

$$\begin{aligned} D_{\max, \mathbb{M}}^{\varepsilon, \Delta}(\rho||\sigma) \\ := \sup \left\{ D_{\max}^{\varepsilon, \Delta}(p_{\rho, M}||p_{\sigma, M}) \mid M = (M_i)_{i=1}^n \in \mathbb{M}, n \in \mathbb{N}_+ \right\}, \end{aligned} \quad (32)$$

where the choices of distance Δ are as in Section III, $p_{\rho, M}$ denotes the probability distribution of the measurement outcomes, $p_{\rho, M}(i) = \text{Tr } M_i \rho$, and \mathbb{M} is the set of all finite-outcome quantum measurements.

We now show that $\tilde{D}_{\max}^{\varepsilon}$ can be understood precisely as a measured variant of the smooth max-relative entropy $D_{\max}^{\varepsilon, T, \leq}$, giving the former quantity a new interpretation and more closely connecting the two divergences.

Proposition 3. *For all quantum states ρ and σ and for all $\varepsilon \in [0, 1]$ it holds that*

$$\tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) = D_{\max, \mathbb{M}}^{\varepsilon, T, \leq}(\rho||\sigma) = \tilde{D}_{\max, \mathbb{M}}^{\varepsilon}(\rho||\sigma). \quad (33)$$

It also follows as a consequence that $D_{\max, \mathbb{M}}^{\varepsilon, T, =}(\rho||\sigma) = \max\{\tilde{D}_{\max}^{\varepsilon}(\rho||\sigma), 0\}$.

Proof. By the data processing inequality for $\tilde{D}_{\max}^{\varepsilon}$, we have that $\tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) \geq \tilde{D}_{\max}^{\varepsilon}(p_{\rho, M}||p_{\sigma, M})$ for any measurement M , and hence

$$\begin{aligned} \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) &\geq \sup_{M \in \mathbb{M}} \tilde{D}_{\max}^{\varepsilon}(p_{\rho, M}||p_{\sigma, M}) \\ &= D_{\max, \mathbb{M}}^{\varepsilon, T, \leq}(\rho||\sigma) \end{aligned} \quad (34)$$

where the last equality is by Lemma 2, using the fact that $p_{\rho, M}$ and $p_{\sigma, M}$ are classical probability distributions and so $\tilde{D}_{\max}^{\varepsilon}(p_{\rho, M}||p_{\sigma, M}) = D_{\max, \mathbb{M}}^{\varepsilon, T, \leq}(p_{\rho, M}||p_{\sigma, M})$.

To show the opposite inequality, we begin by using convex duality to write (see also [35])

$$\begin{aligned} \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) &= \log \inf \left\{ \lambda \mid \rho \leq \lambda \sigma + Q, Q \geq 0, \text{Tr } Q \leq \varepsilon \right\} \\ &= \log \sup \left\{ \text{Tr } W \rho - y \varepsilon \mid 0 \leq W \leq y \mathbb{1}, \right. \\ &\quad \left. \text{Tr } W \sigma = 1, y \geq 0 \right\} \\ &= \log \sup \left\{ \frac{\text{Tr } W' \rho - \varepsilon}{\text{Tr } W' \sigma} \mid 0 \leq W' \leq \mathbb{1} \right\}, \end{aligned} \quad (35)$$

where the second equality follows by strong Lagrange duality, noting that $W = \mathbb{1}$, $y = 1 + \delta$ is a strictly feasible solution pair for any $\delta > 0$, and in last line we made the change of variables $\frac{1}{y}W \mapsto W'$. Let us then take any W' feasible for the optimisation on the rightmost side of (35). Choosing $M' = (W', \mathbb{1} - W')$ we get, using the definition of the measured smooth max-relative entropy in (32) and the classical equivalence between $D_{\max}^{\varepsilon, T, \leq}$ and $\tilde{D}_{\max}^{\varepsilon}$ shown in Lemma 2,

$$\begin{aligned} D_{\max, \mathbb{M}}^{\varepsilon, T, \leq}(\rho||\sigma) &\geq D_{\max}^{\varepsilon, T, \leq}(p_{\rho, M'}||p_{\sigma, M'}) \\ &= \tilde{D}_{\max}^{\varepsilon}(p_{\rho, M'}||p_{\sigma, M'}) \\ &= \log \sup \left\{ \frac{\text{Tr } V p_{\rho, M'} - \varepsilon}{\text{Tr } V p_{\sigma, M'}} \mid 0 \leq V \leq \mathbb{1} \right\} \\ &\geq \log \sup \frac{\text{Tr } W' \rho - \varepsilon}{\text{Tr } W' \sigma}, \end{aligned} \quad (36)$$

where in the last line we made the choice of V as projecting onto the first element of the two-element distribution $p_{\rho, M'}$. Optimising over all choices of W' gives us $D_{\max, \mathbb{M}}^{\varepsilon, T, \leq}(\rho||\sigma) \geq \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma)$, and so the two quantities must be equal.

The result for the variant $D_{\max, \mathbb{M}}^{\varepsilon, T, =}$ with normalised smoothing follows since

$$\begin{aligned} D_{\max, \mathbb{M}}^{\varepsilon, T, =}(\rho||\sigma) &= \sup_{M \in \mathbb{M}} D_{\max}^{\varepsilon, T, =}(p_{\rho, M}||p_{\sigma, M}) \\ &= \sup_{M \in \mathbb{M}} \max \left\{ 0, D_{\max}^{\varepsilon, T, \leq}(p_{\rho, M}||p_{\sigma, M}) \right\} \\ &= \max \left\{ 0, D_{\max, \mathbb{M}}^{\varepsilon, T, \leq}(\rho||\sigma) \right\} \end{aligned} \quad (37)$$

by Lemma 2. \square

C. D_H^ε as optimised $\tilde{D}_{\max}^\varepsilon$

Although the connection between D_{\max}^ε and $D_H^{1-\varepsilon}$ is now a well-known duality in one-shot quantum information theory, the precise bounds between the two are often rather loose — they match in the asymptotic regime, and indeed also when the second-order asymptotics are considered [13], but large gaps are expected between them at the single-copy level (cf. (16)). Due to this, no exact equivalence has been obtained between the two quantities.

Surprisingly, here we show that, as long as one considers the modified max-relative entropy variant $\tilde{D}_{\max}^\varepsilon$, the hypothesis testing relative entropy and the smooth max-relative entropy are in a precise sense equivalent to each other — either of them can be obtained from the other. This will allow us to obtain much tighter bounds between the two quantities than were previously known.

Theorem 4. *For all quantum states ρ and σ and for all $\varepsilon \in (0, 1)$ it holds that*

$$D_H^{1-\varepsilon}(\rho\|\sigma) = \inf_{\delta \in [0, \varepsilon]} \left[\tilde{D}_{\max}^\delta(\rho\|\sigma) - \log(\varepsilon - \delta) \right], \quad (38)$$

$$\tilde{D}_{\max}^\varepsilon(\rho\|\sigma) = \sup_{\delta \in (\varepsilon, 1]} \left[D_H^{1-\delta}(\rho\|\sigma) + \log(\delta - \varepsilon) \right]. \quad (39)$$

Eq. (38) holds also for $\varepsilon = 1$, while Eq. (39) for $\varepsilon = 0$. Without loss of generality, the range of the optimisation can be restricted to $\delta \in (0, \varepsilon)$ in (38) and $\delta \in (\varepsilon, 1)$ in (39).

If ρ and σ commute, \tilde{D}_{\max}^δ can be replaced with $D_{\max}^{\delta, T, \leq}$ in the above.

Proof. Take any $\varepsilon \in (0, 1]$. We first rewrite $D_H^{1-\varepsilon}(\rho\|\sigma)$ in a more convenient form using the change of variables $\frac{1}{\text{Tr } M\sigma} \mapsto z$, $\frac{1}{\text{Tr } M\sigma} M \mapsto M'$:

$$\begin{aligned} D_H^{1-\varepsilon}(\rho\|\sigma) &= \log \sup \left\{ \frac{1}{\text{Tr } M\sigma} \mid 0 \leq M \leq \mathbb{1}, \text{Tr } M\rho \geq \varepsilon \right\} \\ &= \log \sup_{z, M'} \left\{ z \mid 0 \leq M' \leq z\mathbb{1}, \text{Tr } M'\rho \geq z\varepsilon, \text{Tr } M'\sigma = 1 \right\} \\ &= \log \sup_{z, M'} \left\{ z \mid 0 \leq M' \leq z\mathbb{1}, \text{Tr } M'\rho \geq z\varepsilon, \text{Tr } M'\sigma \leq 1 \right\}, \end{aligned} \quad (40)$$

observing in the last line that the constraint $\text{Tr } M'\sigma = 1$ can be relaxed to an inequality without loss of generality, as any feasible M' with $\text{Tr } M'\sigma < 1$ can be rescaled to satisfy $\text{Tr } M'\sigma = 1$ while only increasing the feasible optimal value. One can then readily compute the Lagrange dual as

$$\begin{aligned} D_H^{1-\varepsilon}(\rho\|\sigma) &= \log \inf_{h, y, P} \left\{ h \mid y\rho \leq h\sigma + P, P \geq 0, \right. \\ &\quad \left. h, y \geq 0, 1 + \text{Tr } P \leq y\varepsilon \right\}, \end{aligned} \quad (41)$$

where equality follows by strong duality.² Substituting $\mu = \frac{1}{y}$, $Q = \mu P$, and $\lambda = h\mu$ we get

$$\begin{aligned} D_H^{1-\varepsilon}(\rho\|\sigma) &= \log \inf_{h, \mu, P} \left\{ h \mid \rho \leq h\mu\sigma + \mu P, P \geq 0, \right. \\ &\quad \left. h, \frac{1}{\mu} \geq 0, 1 + \text{Tr } P \leq \frac{1}{\mu}\varepsilon \right\} \\ &= \log \inf_{h, \mu, Q} \left\{ h \mid \rho \leq h\mu\sigma + Q, Q \geq 0, \right. \\ &\quad \left. h \geq 0, \mu > 0, \text{Tr } Q \leq \varepsilon - \mu \right\} \\ &= \log \inf_{\mu \in (0, \varepsilon]} \frac{1}{\mu} \inf_{\lambda, Q} \left\{ \lambda \mid \rho \leq \lambda\sigma + Q, Q \geq 0, \right. \\ &\quad \left. \text{Tr } Q \leq \varepsilon - \mu \right\} \\ &= \inf_{\mu \in (0, \varepsilon]} \left[\tilde{D}_{\max}^{\varepsilon-\mu}(\rho\|\sigma) - \log \mu \right], \end{aligned} \quad (42)$$

where in the last line we recalled the definition of $\tilde{D}_{\max}^{\varepsilon-\mu}$ in (18). This is precisely the expression claimed in (38). Because of its definition as an infimum, the function $\delta \mapsto \tilde{D}_{\max}^\delta$ is lower semi-continuous; since it is also easily verified to be non-increasing, it follows that it must be continuous from the right. This means that the endpoint $\mu = \varepsilon$ can be excluded from the optimisation interval without affecting the optimal value. (For completeness, we prove the continuity properties of \tilde{D}_{\max}^δ in full detail in Appendix B.)

Let us now move on to the second equality of the theorem, namely (39). Take any $\varepsilon \in [0, 1)$. Using the dual form of $\tilde{D}_{\max}^\varepsilon$ in the last line of (35), we have

$$\begin{aligned} \tilde{D}_{\max}^\varepsilon(\rho\|\sigma) &= \log \sup \left\{ \frac{\text{Tr } W'\rho - \varepsilon}{\text{Tr } W'\sigma} \mid 0 \leq W' \leq \mathbb{1} \right\} \\ &\stackrel{(i)}{=} \log \sup \left\{ \frac{\text{Tr } W'\rho - \varepsilon}{\text{Tr } W'\sigma} \mid 0 \leq W' \leq \mathbb{1}, \text{Tr } W'\rho > \varepsilon \right\} \\ &= \log \sup_{\delta \in (\varepsilon, 1]} \sup \left\{ \frac{\text{Tr } W'\rho - \varepsilon}{\text{Tr } W'\sigma} \mid 0 \leq W' \leq \mathbb{1}, \text{Tr } W'\rho = \delta \right\} \\ &= \log \sup_{\delta \in (\varepsilon, 1]} \sup \left\{ \frac{\delta - \varepsilon}{\text{Tr } W'\sigma} \mid 0 \leq W' \leq \mathbb{1}, \text{Tr } W'\rho = \delta \right\} \end{aligned} \quad (43)$$

²Strong duality is particularly easy to see for $\varepsilon \in (0, 1)$: a choice of $z \in (\varepsilon, 1)$ and $M' = \varepsilon\mathbb{1}$ forms a strictly feasible solution pair to the primal problem (40), and so Slater's criterion applies (see e.g. [54, Sec. 5.9]). In the edge case of $\varepsilon = 1$, where the value of the hypothesis testing relative entropy reduces to $D_H^0(\rho\|\sigma) = -\log \text{Tr } \Pi_\rho \sigma$ with Π_ρ denoting the projection onto the support of ρ , the primal problem in (40) is no longer strictly feasible — instead, we can argue the strong feasibility of the dual in (41). To this end, assume that ρ and σ are not orthogonal, as otherwise the problem trivialises. We would now like to find a $P > 0$ such that $y\rho < h\sigma + P$ and $1 + \text{Tr } P < y$. Let us decompose the underlying Hilbert space \mathcal{H} into $\mathcal{H} = \text{supp}(\sigma) \oplus \ker(\sigma)$ and write $h\sigma + P - y\rho = \begin{pmatrix} h\sigma + \Pi_\sigma(P - y\rho)\Pi_\sigma & \Pi_\sigma(P - y\rho)\Pi_\sigma^\perp \\ \Pi_\sigma^\perp(P - y\rho)\Pi_\sigma & \Pi_\sigma^\perp(P - y\rho)\Pi_\sigma^\perp \end{pmatrix}$. For any fixed $P > 0$ and $y > 0$, taking the Schur complement with respect to the upper-left block tells us that, for all sufficiently large h , the matrix is positive definite if and only if $\Pi_\sigma^\perp(P - y\rho)\Pi_\sigma^\perp > 0$ (understood as an operator acting on $\ker(\sigma)$). Let us then choose $P := \Pi_\sigma^\perp \rho \Pi_\sigma^\perp + \delta\mathbb{1}$ for some $\delta > 0$. Then $1 + \text{Tr } P = 1 + \delta \text{Tr } \mathbb{1} + y \text{Tr } \Pi_\sigma^\perp \rho$ which, for sufficiently large y , satisfies $1 + \text{Tr } P < y$ since $\text{Tr } \Pi_\sigma^\perp \rho < 1$ by the assumption of non-orthogonality of ρ and σ . The choice of P , h , and y constructed in this way forms a strictly feasible solution to (41), ensuring that strong duality holds by Slater's criterion.

$$\begin{aligned} & \stackrel{\text{(ii)}}{=} \log \sup_{\delta \in (\varepsilon, 1]} \sup \left\{ \frac{\delta - \varepsilon}{\text{Tr } W' \rho} \mid 0 \leq W' \leq \mathbb{1}, \text{Tr } W' \rho \geq \delta \right\} \\ & \stackrel{\text{(iii)}}{=} \sup_{\delta \in (\varepsilon, 1]} [D_H^{1-\delta}(\rho \parallel \sigma) + \log(\delta - \varepsilon)] \end{aligned}$$

where in (i) we observed that, since the optimal value of the supremum must be positive (consider that $W' = \mathbb{1}$ is feasible), we can restrict to operators W' such that $\text{Tr } W' \rho > \varepsilon$ without loss of generality; in (ii), we used the fact that the condition $\text{Tr } W' \rho = \delta$ can be relaxed to $\text{Tr } W' \rho \geq \delta$ without loss of generality: for any W'' with $\text{Tr } W'' \rho > \delta$ we can scale it down to a solution $W' \leq W''$ which has $\text{Tr } W' \rho = \delta$ and for which the optimal value cannot be smaller; finally, (iii) is by definition of $D_H^{1-\delta}$. The fact that it suffices to optimise over $\delta \in (\varepsilon, 1)$ is argued as before: since the function $\delta \mapsto D_H^{1-\delta}$ is non-increasing and, because of its definition as a supremum, upper semi-continuous,³ it must be continuous from the left.

The classical (commuting) case follows by Lemma 2. \square

From the proof it is evident that this relation can be understood as a very direct consequence of Lagrange duality, making it conceptually related to the quantum Neyman–Pearson lemma [31]. Some related quantitative connections were observed before. This includes a dual expression of D_H^ε in terms of the hockey-stick divergence $\text{Tr}(\rho - \lambda\sigma)_+$ [55, Proposition 3.2], as well as the observation that (39) holds in the classical case [46, Proposition 2].

We also note an interesting connection with one-shot results in information theory that appear in the form $D_H^{\varepsilon-\delta}(\rho \parallel \sigma) - \log \frac{1}{\delta}$, notably in the recent one-shot achievability bounds of [6]. By Theorem 4, optimising over δ allows such results to be understood instead in terms of $\tilde{D}_{\max}^{1-\varepsilon}$. More precisely, the smooth relative entropy bounds of [6] concern bounding the achievable size of a code M in terms of the demanded error probability ε , and for instance [6, Proposition 2] can be understood as stating the bound

$$\log M \geq \tilde{D}_{\max}^{1-\varepsilon}(\rho_{XB} \parallel \rho_X \otimes \rho_B), \quad (44)$$

with an analogous rephrasing applying to the other bounds of [6]. Indeed, this could already be deduced from the proof approach of [6] that directly employs the quantity $\text{Tr}[\rho \wedge \lambda\sigma] = 1 - \text{Tr}(\rho - \lambda\sigma)_+$, as was recently observed also in [56]. The rewriting in terms of $\tilde{D}_{\max}^{1-\varepsilon}$ removes the avoidable error terms δ in the bound and, through the results of this work, can help obtain insights into the properties of this achievability result as well as improved connections with other smooth relative entropies. In the classical case, analogous achievability results were obtained in [57, Theorem 17], and similar quantitative connections were discussed in [58, §2.4].

A noteworthy aspect of the classical case of the above result is that it gives a direct correspondence between the hypothesis testing relative entropy and the smooth max-relative entropy

³Indeed, consider a sequence of values $(\varepsilon_n)_n$, with $\lim_n \varepsilon_n = \varepsilon$. Set $\lambda_n := D_H^{\varepsilon_n}(\rho \parallel \sigma)$, and posit, up to extracting a subsequence, that $\lim_n \lambda_n = \limsup_n \lambda_n = \lambda$. By compactness, for all n we can find a test M_n such that $0 \leq M_n \leq \mathbb{1}$, $\text{Tr } M_n \rho \geq 1 - \varepsilon_n$, and $\text{Tr } M_n \sigma = \exp[-\lambda_n]$. Up to extracting yet another subsequence, we can further assume that $(M_n)_n$ has a limit, i.e. $\lim_n M_n = M$. Naturally, M will also be a valid test, and $\text{Tr } M \rho = \lim_n \text{Tr } M_n \rho \geq \lim_n (1 - \varepsilon_n) = 1 - \varepsilon$. Hence, $D_H^\varepsilon(\rho \parallel \sigma) \geq -\log \text{Tr } M \sigma = -\lim_n \log \text{Tr } M_n \sigma = \limsup_n D_H^{\varepsilon_n}(\rho \parallel \sigma)$.

$D_{\max}^{\varepsilon, T, \leq}$. Moreover, if $\varepsilon \leq \frac{1}{2} \|p - q\|_1$, then we know from Lemma 1 that we can equivalently use the normalised smooth max-relative entropy:

$$\begin{aligned} D_H^{1-\varepsilon}(p \parallel q) &= \inf_{\delta \in [0, \varepsilon]} [D_{\max}^{\delta, T, =}(p \parallel q) - \log(\varepsilon - \delta)], \\ D_{\max}^{\varepsilon, T, =}(p \parallel q) &= \sup_{\delta \in (\varepsilon, 1]} [D_H^{1-\delta}(p \parallel q) + \log(\delta - \varepsilon)]. \end{aligned} \quad (45)$$

In the quantum case, the equivalence we previously showed in Proposition 3 tells us that Theorem 4 can be understood as a correspondence between the hypothesis testing relative entropy and the *measured* smooth max-relative entropy. In Appendix A we also discuss other possible interpretations of $\tilde{D}_{\max}^\varepsilon$. However, in general $\tilde{D}_{\max}^\varepsilon(\rho \parallel \sigma) \neq D_{\max}^{\varepsilon, T, \leq}(\rho \parallel \sigma)$, so the equivalence relation of Theorem 4 does not extend to the standard smooth max-relative entropy D_{\max}^ε itself. Nevertheless, we will see in Section VII that the result will directly lead to the strengthening of multiple bounds involving D_{\max}^ε .

Another immediate consequence of Theorem 4 is that, given two pairs of states (ρ, σ) and (ρ', σ') , it holds that $D_H^\varepsilon(\rho \parallel \sigma) \geq D_H^\varepsilon(\rho' \parallel \sigma') \forall \varepsilon \in [0, 1]$ if and only if $\tilde{D}_{\max}^\varepsilon(\rho \parallel \sigma) \geq \tilde{D}_{\max}^\varepsilon(\rho' \parallel \sigma') \forall \varepsilon \in [0, 1]$. (Note that the point $\varepsilon = 1$ is trivial in both cases, while for $\varepsilon = 0$ we have shown that the expressions in Theorem 4 still apply.) This is related to the result of [59, Theorem 2] and the classical [60, Theorem 3]. Statements of this kind are important in the context of statistical comparison of experiments in the sense of Blackwell [61] and its quantum generalisations [62], [63], [59].

VI. IMPROVING A USEFUL LEMMA

Although we have already shown a very close relation between D_H^ε and the modified quantity $\tilde{D}_{\max}^\varepsilon$, this is so far insufficient to tightly connect the hypothesis testing relative entropy with the smooth max-relative entropy D_{\max}^ε itself. The issue is that, although one can straightforwardly upper bound $\tilde{D}_{\max}^\varepsilon$ with $D_{\max}^{\varepsilon, T, \leq}$ and other variants of D_{\max}^ε (see Eq. (71)), the other direction is much less obvious: $\tilde{D}_{\max}^\varepsilon$ involves an optimisation over values of λ such that $\rho \leq \lambda\sigma + Q$, where Q is a positive semidefinite operator of sufficiently small trace, but how can this operator inequality yield a (possibly subnormalised) state $\rho' \approx_\varepsilon \rho$ that can be used as a feasible solution for the smooth max-relative entropy?

A solution to this conundrum was first given in a fundamental lemma by Datta and Renner [29, Lemma 3]. It found use in many results that studied the asymptotics of the max-relative entropy, e.g. [22], [23], [24], [18], [25], [26], [27]. Here we introduce a modified proof approach, allowing us to give better estimates in particular for smoothing over normalised states.

We first state the lemma in a rather general form; the specific applications to max-relative entropy bounds will be discussed in Section VII.

Theorem 5 (Tightened Datta–Renner lemma). *Let ρ be a state, $A, Q \geq 0$ positive semi-definite with $\text{Tr } Q \leq \varepsilon < 1$, and assume that $\rho \leq A + Q$ holds. Then there exists a subnormalised state ρ' such that*

$$\rho' \leq A, \quad \frac{\rho'}{\text{Tr } \rho'} \leq \frac{A}{1 - \varepsilon}; \quad (46)$$

moreover, ρ' is close to ρ in the following senses:

$$F(\rho, \rho') \geq (1 - \varepsilon)^2, \quad (47)$$

$$F\left(\rho, \frac{\rho'}{\text{Tr } \rho'}\right) \geq 1 - \varepsilon, \quad (48)$$

$$\frac{1}{2} \left\| \rho - \frac{\rho'}{\text{Tr } \rho'} \right\|_1 \leq \sqrt{\varepsilon}, \quad (49)$$

$$\|\rho - \rho'\|_+ \leq \sqrt{\varepsilon \left(1 - \frac{3\varepsilon}{4}\right)} + \frac{\varepsilon}{2} \leq \sqrt{\varepsilon(2 - \varepsilon)}. \quad (50)$$

Here, the fidelity bound with unnormalised states in (47) is the same as the one found in [29], [25], but the other bounds improve on known estimates (e.g. [29], [23]), and in particular on the statement of the lemma with normalised smoothing found in [23, Lemma C.5].

Before proving this result, let us discuss the simple but consequential difference between our approach and the one found in previous works. Since our aim is to use the matrix inequality $\rho \leq A + Q$ to construct a subnormalised state ρ' such that $\rho' \leq A$, a natural idea is to define $T := A^{1/2}(A + Q)^{-1/2}$ and take the ansatz $\rho' := T\rho T^\dagger$, which is easily verified to satisfy the desired inequality. This is indeed the approach that the original proofs of the Datta–Renner lemma took [29], [23], [25]. But then how can we determine how close ρ' is to ρ in terms of distance measures such as the trace distance or fidelity? One may be tempted to use the celebrated gentle measurement lemma [64], [65], which tells us that if M is a POVM operator such that $\text{Tr } M\rho$ is close to 1, then the subnormalised state $\sqrt{M}\rho\sqrt{M}$ is close to ρ . The issue, however, is that this lemma applies only to positive semidefinite operators, which the ansatz T is not. This is indeed a difficulty that arises in the previous proof methods, and the bounds derived therein — in particular, those for a normalised smoothing state — are looser than what one would have obtained from applying the gentle measurement lemma. Is there, then, a way to instead pick a positive semidefinite operator in this approach?

Notice that the operator $G := A^{1/2}U(A + Q)^{-1/2}$, where U is some unitary, still satisfies $G\rho G^\dagger \leq A$. The question then becomes whether one can choose U in a way that makes G positive semidefinite. This is indeed always possible, and if $A > 0$, then such a solution is unique [50, Proposition 4.1.8]: U equals $(A^{-1/2}(A + Q)^{-1}A^{-1/2})^{1/2}A^{1/2}(A + Q)^{1/2}$, and G becomes the **geometric mean** of A and $(A + Q)^{-1}$.

More generally, for $A, B \geq 0$ one defines the operator geometric mean $A \# B$ as

$$A \# B := A^{1/2} \left(A^{-1/2} B A^{-1/2} \right)^{1/2} A^{1/2}. \quad (51)$$

The properties of the geometric mean that will be relevant to us is that it is positive semidefinite, and that it is monotone non-increasing in either argument: in particular, $A \leq C$ implies that $A \# B \leq C \# B$ (see e.g. [50, Ch. 4.1]).

It thus looks as if we are in a position to apply the gentle measurement lemma to $\rho' := G\rho G$. Another issue transpires, however: although the known formulations of the gentle measurement lemma give tight estimates on the distance of the normalised state $\frac{G\rho G}{\text{Tr } G\rho G}$ to ρ , it turns out that the

previous bounds on the error of the subnormalised state $G\rho G$ were not tight in trace distance. We thus first introduce the following improvement.

Lemma 6 (A gentler measurement lemma). *Let $M \in [0, \mathbb{1}]$ be a measurement operator and ρ be an arbitrary state on the same system. If $\text{Tr } M\rho \geq 1 - \varepsilon$ for some $\varepsilon \in [0, 1]$, then*

$$F\left(\rho, \sqrt{M}\rho\sqrt{M}\right) \geq (1 - \varepsilon)^2, \quad (52)$$

$$F\left(\rho, \frac{\sqrt{M}\rho\sqrt{M}}{\text{Tr } M\rho}\right) \geq 1 - \varepsilon, \quad (53)$$

$$\frac{1}{2} \left\| \rho - \frac{\sqrt{M}\rho\sqrt{M}}{\text{Tr } M\rho} \right\|_1 \leq \sqrt{\varepsilon}, \quad (54)$$

$$\frac{1}{2} \left\| \rho - \sqrt{M}\rho\sqrt{M} \right\|_1 \leq \begin{cases} \sqrt{\varepsilon \left(1 - \frac{3\varepsilon}{4}\right)} & \text{if } \varepsilon \leq 2/3, \\ 1/\sqrt{3} & \text{if } \varepsilon > 2/3 \end{cases} \quad (55)$$

$$\leq \sqrt{\varepsilon},$$

$$\left\| \rho - \sqrt{M}\rho\sqrt{M} \right\|_+ \leq \sqrt{\varepsilon \left(1 - \frac{3\varepsilon}{4}\right)} + \frac{\varepsilon}{2} \quad (56)$$

$$\leq \sqrt{\varepsilon(2 - \varepsilon)}.$$

All of the primary bounds are tight for all $\varepsilon \in [0, 1]$.

Here, the bounds (52)–(54) were known (see e.g. [66, Lemma 9.4.1]), while (55) and (56) improve on previous estimates [64], [65], [67]. To streamline the discussion here, we defer the proof to Appendix C.

We can now prove Theorem 5 using the reasoning outlined above.

Proof of Theorem 5. Up to projecting down onto the support of $A + Q$, we can assume without loss of generality that $A + Q > 0$ is invertible. Define

$$G := A \# ((A + Q)^{-1})$$

$$= (A + Q)^{-1/2} \left((A + Q)^{1/2} A (A + Q)^{1/2} \right)^{1/2} (A + Q)^{-1/2}. \quad (57)$$

Due to the monotonicity of the operator geometric mean we have that

$$0 \leq G \leq (A + Q) \# ((A + Q)^{-1}) = \mathbb{1}. \quad (58)$$

Conjugating by G , from $\rho \leq A + Q$ we deduce that

$$\rho' := G\rho G \leq G(A + Q)G = A, \quad (59)$$

as a simple calculation using the formula on the rightmost side of (57) reveals. We now estimate

$$1 - \text{Tr } \rho G^2 = \text{Tr } \rho (\mathbb{1} - G^2) \leq \text{Tr}(A + Q) (\mathbb{1} - G^2)$$

$$= \text{Tr}(A + Q) - \text{Tr } G(A + Q)G \quad (60)$$

$$= \text{Tr } Q \leq \varepsilon,$$

where the first inequality holds due to the fact that $\mathbb{1} - G^2 \geq 0$. Applying the gentler measurement lemma (Lemma 6) gives the claimed bounds. \square

An advantage of the proof approach using the operator geometric mean is that it can be easily adapted to more

general scenarios. In particular, we can generalise it to a multi-partite setting where the smoothing is performed over the marginals over a global state, inspired by the notion of ‘simultaneous smoothing’ considered in the context of the one-shot multipartity typicality conjecture by Drescher and Fawzi [68] and encountered earlier in [20]. We discuss this in detail in Section VIII-C.

VII. TIGHTENED BOUNDS AND RELATIONS

A. Relating $\tilde{D}_{\max}^\varepsilon$ and D_H^ε

We are now ready to tackle the question of relating the two fundamental operational quantities, D_{\max}^ε and D_H^ε .

The first key ingredient will be tight bounds between D_H^ε and the modified smooth max-relative entropy $\tilde{D}_{\max}^\varepsilon$, obtained from the precise connection between the two that we established in Theorem 4.

Lemma 7. *For all $\varepsilon \in (0, 1)$ and all $\mu \in (0, \varepsilon]$, it holds that*

$$\begin{aligned} \tilde{D}_{\max}^\varepsilon(\rho\|\sigma) + \log \frac{1}{\varepsilon(1-\varepsilon)} &\leq D_H^{1-\varepsilon}(\rho\|\sigma) \\ &\leq \tilde{D}_{\max}^{\varepsilon-\mu}(\rho\|\sigma) + \log \frac{1}{\mu} \end{aligned} \quad (61)$$

The first inequality improves over the previously known bound of [28, Proposition 4.7], and indeed also over a stronger bound that was implicit in the proof of [18, Theorem 11]. The second inequality was known [28].

Proof. The upper bound is an immediate consequence of (38) in Theorem 4.

Let us now fix some $\delta \in (\varepsilon, 1]$. Using again the same equation (38) tells us that for any $\zeta \in (0, \delta]$ we have

$$D_H^{1-\delta}(\rho\|\sigma) \leq \tilde{D}_{\max}^{\delta-\zeta}(\rho\|\sigma) + \log \frac{1}{\zeta}. \quad (62)$$

With the choice $\zeta = \delta - \varepsilon + \mu$ for some fixed $\mu \in (0, \varepsilon]$, this gives

$$\begin{aligned} D_H^{1-\delta}(\rho\|\sigma) + \log \frac{\delta - \varepsilon}{\varepsilon(1-\varepsilon)} &\leq \tilde{D}_{\max}^{\varepsilon-\mu}(\rho\|\sigma) + \log \frac{\delta - \varepsilon}{\varepsilon(1-\varepsilon)(\delta - \varepsilon + \mu)} \\ &\leq \tilde{D}_{\max}^{\varepsilon-\mu}(\rho\|\sigma) + \log \frac{1 - \varepsilon}{\varepsilon(1-\varepsilon)(1 - \varepsilon + \mu)} \\ &= \tilde{D}_{\max}^{\varepsilon-\mu}(\rho\|\sigma) + \log \frac{1}{\varepsilon(1 - \varepsilon + \mu)}, \end{aligned} \quad (63)$$

where in the second line we made use of the fact that $\frac{\delta - \varepsilon}{\delta - \varepsilon + \mu} = \left(1 + \frac{\mu}{\delta - \varepsilon}\right)^{-1}$, which is clearly monotonic in δ . Consider now that

$$\varepsilon(1 - \varepsilon + \mu) = \varepsilon(1 - \varepsilon) + \varepsilon\mu \geq \mu(1 - \varepsilon) + \varepsilon\mu = \mu \quad (64)$$

since $\varepsilon \geq \mu$ by assumption. Thus

$$D_H^{1-\delta}(\rho\|\sigma) + \log \frac{\delta - \varepsilon}{\varepsilon(1 - \varepsilon)} \leq \tilde{D}_{\max}^{\varepsilon-\mu}(\rho\|\sigma) + \log \frac{1}{\mu}. \quad (65)$$

Taking the supremum over all $\delta \in (\varepsilon, 1]$ and infimum over all $\mu \in (0, \varepsilon]$ yields

$$\tilde{D}_{\max}^\varepsilon(\rho\|\sigma) + \log \frac{1}{\varepsilon(1 - \varepsilon)} \leq D_H^{1-\varepsilon}(\rho\|\sigma) \quad (66)$$

by Theorem 4. \square

A point of note here is that the bounds can be verified to be tight in many ways. The tightness of the error term $\log \frac{1}{\varepsilon(1-\varepsilon)}$ is particularly easy to see in the trivial case $\rho = \sigma$, where it holds that $D_H^{1-\varepsilon}(\rho\|\rho) = \log \frac{1}{\varepsilon}$ and $\tilde{D}_{\max}^\varepsilon(\rho\|\rho) = \log(1 - \varepsilon)$. In light of Theorem 4, we also see that the upper bound is as tight as possible, since it holds with equality by taking the infimum over μ . The lower bound is additionally tight in an i.i.d. asymptotic sense at the level of exponents, as we will shortly see in Sec. VIII-A.

B. Relating D_{\max}^ε and $\tilde{D}_{\max}^\varepsilon$

Applying the improved Datta–Renner lemma immediately gives the following bounds.

Corollary 8. *For all quantum states ρ and σ and for all $\varepsilon \in [0, 1)$, it holds that*

$$D_{\max}^{\sqrt{\varepsilon}, T, =}(\rho\|\sigma) - \log \frac{1}{1 - \varepsilon} \leq \tilde{D}_{\max}^\varepsilon(\rho\|\sigma), \quad (67)$$

$$D_{\max}^{\sqrt{\varepsilon(1-\frac{3\varepsilon}{4})+\frac{\varepsilon}{2}}, T, \leq}(\rho\|\sigma) \leq \tilde{D}_{\max}^\varepsilon(\rho\|\sigma), \quad (68)$$

$$D_{\max}^{\sqrt{\varepsilon}, P, =}(\rho\|\sigma) - \log \frac{1}{1 - \varepsilon} \leq \tilde{D}_{\max}^\varepsilon(\rho\|\sigma), \quad (69)$$

$$D_{\max}^{\sqrt{\varepsilon(2-\varepsilon)}, P, \leq}(\rho\|\sigma) \leq \tilde{D}_{\max}^\varepsilon(\rho\|\sigma). \quad (70)$$

Proof. Recall that any feasible solution for $\tilde{D}_{\max}^\varepsilon(\rho\|\sigma)$ corresponds to an operator inequality $\rho \leq \lambda\sigma + Q$ for some $Q \geq 0$ with $\text{Tr } Q \leq \varepsilon$. The result is then a direct consequence of Theorem 5 with the choice $A = \lambda\sigma$, as the result gives us precisely either normalised or subnormalised states that are feasible solutions for the different variants of $D_{\max}^\varepsilon(\rho\|\sigma)$. \square

To relate the modified max-relative entropy and the standard smoothed variants, we now need inequalities in the opposite direction. Indeed, one such relation follows immediately from the variational form of the trace distance in Eq. (10). We state this here for completeness.

Lemma 9. *For all $\varepsilon \in (0, 1)$, it holds that*

$$\tilde{D}_{\max}^\varepsilon(\rho\|\sigma) \leq D_{\max}^{\varepsilon, T, \leq}(\rho\|\sigma) \leq D_{\max}^{\varepsilon, P, \leq}(\rho\|\sigma), \quad (71)$$

and analogously for the quantities $D_{\max}^{\varepsilon, T, =}$ and $D_{\max}^{\varepsilon, P, =}$ with normalised smoothing.

Proof. Given any $\rho' \leq \lambda\sigma$, we have that $\rho \leq \lambda\sigma + (\rho - \rho')_+$, and by definition $\text{Tr}(\rho - \rho')_+ = \|\rho - \rho'\|_+$. Thus any feasible solution for $D_{\max}^{\varepsilon, T, \leq}$ yields a feasible solution for $\tilde{D}_{\max}^\varepsilon$ with the same feasible optimal value. Together with one of the Fuchs–van de Graaf inequalities (12), this gives the stated bounds. \square

This bound is tight for the trace distance, as can be seen from the classical case in Lemma 2. However, it can already be deduced from prior works such as [13, Proposition 13] that for the purified distance, tighter bounds can be obtained — in particular, ones that feature $\sqrt{\varepsilon}$ rather than ε in the upper bound. Here we improve on such bounds.

Lemma 10. *Let ρ, σ be two states. For all $\varepsilon \in (0, 1)$ and $\delta \in (0, 1 - \varepsilon)$, we have*

$$\tilde{D}_{\max}^{\varepsilon+\delta}(\rho\|\sigma) \leq D_{\max}^{\sqrt{\varepsilon}, P, \leq}(\rho\|\sigma) + \log \frac{(\varepsilon + \delta)(1 - \varepsilon - \delta)}{\delta}. \quad (72)$$

As a consequence, for any $c \in (1, \frac{1}{\varepsilon})$ it also holds that

$$\tilde{D}_{\max}^{c\varepsilon}(\rho\|\sigma) \leq D_{\max}^{\sqrt{\varepsilon}, P, \leq}(\rho\|\sigma) + \log \frac{c}{c-1}. \quad (73)$$

Proof. Let ρ' be a smoothing of ρ with the property that

$$P(\rho, \rho') \leq \sqrt{\varepsilon}, \quad \rho' \leq \lambda\sigma, \quad \log \lambda = D_{\max}^{\sqrt{\varepsilon}, P}(\rho\|\sigma). \quad (74)$$

By Uhlmann's theorem, this means that there are two purifications $|\psi\rangle$ of ρ and $|\psi'\rangle$ of ρ' (not necessarily normalised) such that

$$\text{Tr} \psi\psi' = F(\rho, \rho') \geq 1 - \varepsilon. \quad (75)$$

We would now like to estimate $\tilde{D}_{\max}^{\varepsilon+\delta}(\psi\|\psi')$. To this end, consider for some $k > 0$ the operator $\psi - k\psi'$. We will use the fact that any rank-two Hermitian operator X satisfies

$$\begin{aligned} \|X\|_1 &= \sqrt{\text{Tr}(X^2) + 2|\det(X)|} \\ &= \sqrt{\text{Tr}(X^2) + |\text{Tr}(X^2) - \text{Tr}(X)^2|}. \end{aligned} \quad (76)$$

In our case, $X = \psi - k\psi'$ has one positive and one negative eigenvalue, which implies that $\text{Tr}(X^2) - \text{Tr}(X)^2 \geq 0$. Thus

$$\begin{aligned} \|\psi - k\psi'\|_1 &= \sqrt{2 \text{Tr}((\psi - k\psi')^2) - (\text{Tr}(\psi - k\psi'))^2} \\ &= \sqrt{(\text{Tr} \psi + k \text{Tr} \psi')^2 - 4k \text{Tr} \psi\psi'} \\ &\leq \sqrt{(1+k)^2 - 4k(1-\varepsilon)} \\ &= \sqrt{(1-k)^2 + 4k\varepsilon}. \end{aligned} \quad (77)$$

This gives

$$\begin{aligned} \text{Tr}(\psi - k\psi')_+ &= \frac{1}{2} \text{Tr}(\psi - k\psi') + \frac{1}{2} \|\psi - k\psi'\|_1 \\ &\leq \frac{1}{2} \left(1 - k + \sqrt{(1-k)^2 + 4k\varepsilon} \right). \end{aligned} \quad (78)$$

Setting the rightmost side to be equal to $\varepsilon + \delta$ and solving for k , we obtain

$$k = \frac{(\varepsilon + \delta)(1 - \varepsilon - \delta)}{\delta}. \quad (79)$$

By the data processing inequality, we then infer that

$$\tilde{D}_{\max}^{\varepsilon+\delta}(\rho\|\rho') \leq \tilde{D}_{\max}^{\varepsilon+\delta}(\psi\|\psi') \leq \log \frac{(\varepsilon + \delta)(1 - \varepsilon - \delta)}{\delta}, \quad (80)$$

which implies that there exists an operator $Q \geq 0$ with $\text{Tr} Q \leq \varepsilon + \delta$ such that $\rho \leq \frac{(\varepsilon+\delta)(1-\varepsilon-\delta)}{\delta}\rho' + Q$. Combined with the assumption that $\rho' \leq \lambda\sigma$ this gives

$$\tilde{D}_{\max}^{\varepsilon+\delta}(\rho\|\sigma) \leq \log \lambda \frac{(\varepsilon + \delta)(1 - \varepsilon - \delta)}{\delta} \quad (81)$$

as claimed.

The second part of the Lemma follows by fixing $\delta = (c-1)\varepsilon$ and writing

$$\begin{aligned} \tilde{D}_{\max}^{c\varepsilon}(\rho\|\sigma) &\leq D_{\max}^{\sqrt{\varepsilon}, P, \leq}(\rho\|\sigma) + \log \frac{c\varepsilon(1 - c\varepsilon)}{(c-1)\varepsilon} \\ &\leq D_{\max}^{\sqrt{\varepsilon}, P, \leq}(\rho\|\sigma) + \log \frac{c}{c-1}. \end{aligned} \quad (82)$$

This concludes the proof. \square

We observe that the approach of the proof readily gives an exact expression for the value of $\tilde{D}_{\max}^{\varepsilon}$ between any two pure states; this can be used to obtain an exact expression also for the pure-state hypothesis testing relative entropy D_H^{ε} . Such an expression can be found for instance in [31, Ch. IV, Eq. (2.33)], although we do not believe this to be well known in the literature. We derive and state the result here as it will be useful in establishing the tightness of our bounds.

Lemma 11. *Consider any two pure states $\psi_1 = |\psi_1\rangle\langle\psi_1|$ and $\psi_2 = |\psi_2\rangle\langle\psi_2|$, and let $f = \text{Tr} \psi_1\psi_2 = |\langle\psi_1|\psi_2\rangle|^2$. Then*

$$\tilde{D}_{\max}^{\varepsilon}(\psi_1\|\psi_2) = \begin{cases} \log \frac{\varepsilon(1-\varepsilon)}{f-(1-\varepsilon)} & \text{if } \varepsilon > 1-f \\ \infty & \text{otherwise,} \end{cases} \quad (83)$$

$$D_H^{\varepsilon}(\psi_1\|\psi_2) = \begin{cases} \log \frac{F_2(1-\varepsilon, f)}{(f-\varepsilon)^2} & \text{if } \varepsilon < f \\ \infty & \text{otherwise} \end{cases} \quad (84)$$

$$= \begin{cases} -\log[1 - F_2(\varepsilon, f)] & \text{if } \varepsilon < f \\ \infty & \text{otherwise,} \end{cases} \quad (85)$$

where $F_2 : [0, 1] \times [0, 1] \rightarrow [0, 1]$ denotes the binary fidelity (Bhattacharyya coefficient), defined as

$$F_2(p, q) := \left(\sqrt{pq} + \sqrt{(1-p)(1-q)} \right)^2. \quad (86)$$

Proof. Using Eq. (77), where we now write an equality instead of an inequality, we have

$$\text{Tr}(\psi_1 - \lambda\psi_2)_+ = \frac{1}{2} \left(1 - \lambda + \sqrt{(1+\lambda)^2 - 4\lambda f} \right). \quad (87)$$

For the right-hand side to equal ε , and hence for λ to be a feasible solution for $\tilde{D}_{\max}^{\varepsilon}$, we need

$$\lambda = \frac{\varepsilon(1-\varepsilon)}{f-(1-\varepsilon)}, \quad (88)$$

which is precisely the claimed expression. If $f \leq 1 - \varepsilon$, this is impossible to satisfy for any $\lambda \in (0, \infty)$, leading to a diverging optimal value.

For D_H^{ε} , consider first that if $\varepsilon \geq f$, then $M = \mathbb{1} - \psi_2$ satisfies $\text{Tr} M\psi_1 \geq 1 - \varepsilon$ and hence is a feasible measurement operator for the definition of the hypothesis testing relative entropy (6), giving $D_H^{\varepsilon}(\psi_1\|\psi_2) \geq -\log \text{Tr} M\psi_2 = \infty$. Let us then assume that $\varepsilon < f$. In this case, take $M = |\phi\rangle\langle\phi|$, where

$$\begin{aligned} |\phi\rangle &= \sqrt{1-\varepsilon}|\psi_1\rangle - \sqrt{\varepsilon}|\psi_1^{\perp}\rangle, \\ |\psi_1^{\perp}\rangle &= \frac{(\mathbb{1} - \psi_1)|\psi_2\rangle}{\|(\mathbb{1} - \psi_1)|\psi_2\rangle\|} = \frac{(\mathbb{1} - \psi_1)|\psi_2\rangle}{\sqrt{1-f}}. \end{aligned} \quad (89)$$

This operator clearly satisfies $\text{Tr} M\psi_1 = 1 - \varepsilon$, meaning that $D_H^{\varepsilon}(\psi_1\|\psi_2) \geq -\log \text{Tr} M\psi_2$, where

$$\begin{aligned} \text{Tr} M\psi_2 &= \left(\sqrt{1-\varepsilon}\sqrt{f} - \sqrt{1-f}\sqrt{\varepsilon} \right)^2 \\ &= \frac{(f-\varepsilon)^2}{F_2(1-\varepsilon, f)}, \end{aligned} \quad (90)$$

using that $(a - b)^2 = \frac{(a^2 - b^2)^2}{(a+b)^2}$. The expression appearing in (85) is a rewriting of the above using the convenient identity

$$\begin{aligned} & \left(\sqrt{p(1-q)} - \sqrt{(1-p)q} \right)^2 \\ &= 1 - \left(\sqrt{pq} + \sqrt{(1-p)(1-q)} \right)^2. \end{aligned} \quad (91)$$

For the opposite inequality, consider that Lemma 7 gives

$$\begin{aligned} D_H^\varepsilon(\psi_1 \parallel \psi_2) &\leq \tilde{D}_{\max}^{1-\varepsilon-\mu}(\psi_1 \parallel \psi_2) + \log \frac{1}{\mu} \\ &= \log \frac{(1-\varepsilon-\mu)(\varepsilon+\mu)}{(f-\varepsilon-\mu)\mu} \end{aligned} \quad (92)$$

for any $\mu \in (0, f-\varepsilon)$, where in the second line we used the pure-state formula for $\tilde{D}_{\max}^{\varepsilon-\mu}$ from Eq. (88). Choosing

$$\mu = \frac{f-\varepsilon}{1 + \sqrt{\frac{f(1-f)}{\varepsilon(1-\varepsilon)}}} = \frac{\varepsilon(1-\varepsilon) - \sqrt{\varepsilon f(1-\varepsilon)(1-f)}}{f - (1-\varepsilon)}, \quad (93)$$

some cumbersome yet straightforward algebra gives exactly

$$\frac{(f-\varepsilon-\mu)\mu}{(1-\varepsilon-\mu)(\varepsilon+\mu)} = \left(\sqrt{1-\varepsilon}\sqrt{f} - \sqrt{1-f}\sqrt{\varepsilon} \right)^2, \quad (94)$$

meaning that the upper bound of (92) matches the lower bound of (90). \square

C. Improved bounds between D_{\max}^ε and D_H^ε

Putting our findings together, we obtain upper and lower bounds that directly connect D_{\max}^ε with $D_H^{1-\varepsilon}$, recovering the known ‘weak/strong converse duality’ between the two quantities and improving on the previously known quantitative bounds between them.

Theorem 12 (Tight weak/strong converse duality between D_{\max}^ε and $D_H^{1-\varepsilon}$). *For all quantum states ρ and σ , all $\varepsilon \in (0, 1)$, and all $\mu \in (0, \varepsilon]$, it holds that*

$$D_{\max}^{\sqrt{\varepsilon}, T, =}(\rho \parallel \sigma) + \log \frac{1}{\varepsilon} \quad (95a)$$

$$\leq D_H^{1-\varepsilon}(\rho \parallel \sigma) \quad (95b)$$

$$\leq D_{\max}^{\varepsilon-\mu, T, =}(\rho \parallel \sigma) + \log \frac{1}{\mu}, \quad (95c)$$

$$D_{\max}^{\sqrt{\varepsilon}, P, =}(\rho \parallel \sigma) + \log \frac{1}{\varepsilon} \quad (96a)$$

$$\leq D_H^{1-\varepsilon}(\rho \parallel \sigma) \quad (96b)$$

$$\leq D_{\max}^{\sqrt{\varepsilon-\mu}, P, =}(\rho \parallel \sigma) + \log \frac{F_2(1-\varepsilon, \varepsilon-\mu)}{\mu^2} \quad (96c)$$

$$\leq D_{\max}^{\varepsilon-\mu, P, =}(\rho \parallel \sigma) + \log \frac{1}{\mu^2}, \quad (97)$$

where we recall that

$$F_2(1-\varepsilon, \varepsilon-\mu) = \left(\sqrt{(1-\varepsilon)(\varepsilon-\mu)} + \sqrt{\varepsilon(1-\varepsilon+\mu)} \right)^2.$$

For classical systems or commuting quantum states, a stronger trace distance bound holds:

$$\begin{aligned} & D_{\max}^{\varepsilon, T, =}(p \parallel q) + \log \frac{1}{\varepsilon} \\ & \leq D_H^{1-\varepsilon}(p \parallel q) \\ & \leq D_{\max}^{\varepsilon-\mu, T, =}(p \parallel q) + \log \frac{1}{\mu}. \end{aligned} \quad (98)$$

All of the primary bounds of this theorem are tight: namely, for all $0 < \mu \leq \varepsilon < 1$, we can find two states saturating any of the inequalities in (95), (96), and (98). The bounds also hold when normalised smoothing is replaced with subnormalised smoothing.

Proof. The lower bounds on the hypothesis testing relative entropy in Eqs. (95)–(96) follow directly by combining Lemma 7 and Corollary 8. The upper bound in (95) is from Lemma 9.

The upper bound in Eq. (96) requires more elaboration. Combining Lemma 7 with Lemma 10 gives, for any $\delta \in (0, \mu)$, that

$$\begin{aligned} & D_H^{1-\varepsilon}(\rho \parallel \sigma) \\ & \leq \tilde{D}_{\max}^{\varepsilon+\delta-\mu}(\rho \parallel \sigma) + \log \frac{1}{\mu-\delta} \\ & \leq D_{\max}^{\sqrt{\varepsilon-\mu}, P, \leq}(\rho \parallel \sigma) + \log \frac{(\varepsilon+\delta-\mu)(1-\varepsilon-\delta+\mu)}{\delta(\mu-\delta)}. \end{aligned} \quad (99)$$

For the choice of $\delta = \mu \left(1 + \sqrt{\frac{\varepsilon(1-\varepsilon)}{(\varepsilon-\mu)(1-\varepsilon+\mu)}} \right)^{-1}$, as in the proof of Lemma 11 we get exactly (96), which can be verified to be minimal.

For the classical case, we can use Lemma 1 as well as the equivalence between $D_{\max}^{\varepsilon, T, =}(p \parallel q)$ and $\tilde{D}_{\max}^\varepsilon(p \parallel q)$ in Lemma 2 to get

$$\begin{aligned} D_{\max}^{\varepsilon, T, =}(p \parallel q) + \frac{1}{\varepsilon} &= \max \left\{ \tilde{D}_{\max}^\varepsilon(p \parallel q) + \log \frac{1}{\varepsilon}, \log \frac{1}{\varepsilon} \right\} \\ &\leq D_H^{1-\varepsilon}(p \parallel q), \end{aligned} \quad (100)$$

where the inequality follows from Lemma 7 as well as the fact that $D_H^{1-\varepsilon}(\rho \parallel \sigma) \geq \log \frac{1}{\varepsilon}$ for all states, easily verified by choosing $M = \varepsilon \mathbb{1}$ in the definition (6). \square

Almost all of the inequalities in Theorem 12 improve over state-of-the-art bounds, often significantly so. The first inequalities in Eqs. (95) and (96) are a major improvement over the best known bound of [20, Theorem 4]: the left-hand side of our bound is larger by an additive term of $\log \frac{1}{\varepsilon(1-\varepsilon)}$. This result improves also on bounds stated in [19, Proposition 4.1] and [18, Theorem 11], which had tighter error terms but worse smoothing terms than the bound of [20]. The second inequality in (95) recovers a result of [18, Theorem 11]. Our upper bound for the purified distance in (96) improves previously known results, the strongest of which was the upper bound of [20, Theorem 4].

Once again, the bounds are the strongest possible. Even in the simplest case $\rho = \sigma$, previous results did not give tight bounds; in contrast, our bounds are, to the best of our knowledge, the first statement of the weak/strong converse duality between D_H^ε and D_{\max}^ε that gives tight error terms in this sense. This follows since $D_{\max}^{\sqrt{\varepsilon}, T, =}(\rho \parallel \rho) = D_{\max}^{\sqrt{\varepsilon}, P, =}(\rho \parallel \rho) = 0$ while $D_H^{1-\varepsilon}(\rho \parallel \rho) = \log \frac{1}{\varepsilon}$, which immediately yields a saturation of the lower bounds on $D_H^{1-\varepsilon}$ in Eqs. (95)–(96). The lower bounds also give a tight constraint on the asymptotic error exponent of $D_{\max}^{\varepsilon, P, \leq}$ [69] and on the exponent of $D_{\max}^{\varepsilon, T, \leq}$ for classical systems, which we will discuss in more detail in the next section.

The upper bounds on the hypothesis testing relative entropy in (95) and (98) are saturated, for example, by the classical distributions $p = (\varepsilon, 1-\varepsilon)$ and $q = (\mu, 1-\mu)$, since one easily verifies that $\|p - q\|_+ = \varepsilon - \mu$ and so $D_{\max}^{\varepsilon-\mu, T, =}(p\|q) = 0$, while the test $M = (1, 0)$ satisfies $\text{Tr } Mp = \varepsilon$ and $\text{Tr } Mq = \mu$, meaning that $D_H^{1-\varepsilon}(p\|q) \geq -\log \mu$. For the upper bound in (96), consider two pure states $\psi = |\psi\rangle\langle\psi|$ and $\phi = |\phi\rangle\langle\phi|$ such that $P(\psi, \phi) = \sqrt{\varepsilon - \mu}$ and hence $D_{\max}^{\sqrt{\varepsilon-\mu}, P, =}(\psi\|\phi) = 0$. By Lemma 11, $D_H^{1-\varepsilon}(\psi\|\phi) = -\log(1 - F_2(\varepsilon, \varepsilon - \mu))$, which exactly matches the rightmost side of the bound in (96).

The upper bound in (98) is also tight for any p, q , and $\varepsilon \leq \|p - q\|_+$, in the following sense: by Theorem 4 and Lemma 1, minimising the rightmost side of (95) or (98) over μ gives exact equality in the bound. This means that the upper bound can be saturated for all classical distributions with the optimal choice of μ .

We observe that the bounds in (95) have a mismatch in the functional dependence of the smoothing parameter: the lower bound on $D_H^{1-\varepsilon}$ features smooth max-relative entropy with smoothing of order $\sqrt{\varepsilon}$ (effectively due to the use of the gentle measurement lemma in Theorem 5), while the smoothing parameter in the upper bounds is of order ε . This contrasts with the bounds for purified distance in (96), which both share the same order of $\sqrt{\varepsilon}$ — this is a crucial property that allowed e.g. the computation of the second-order expansion of the max-relative entropy for the purified distance [13] and the evaluation of the error and strong converse exponents for $D_{\max}^{\varepsilon, P, \leq}$ [69], [70]. It is clear that the trace distance smoothing exhibits a different behaviour than the purified distance, as indicated by the fact that the classical case of (98) features matching scaling of order ε rather than $\sqrt{\varepsilon}$. Indeed, because of this classical case, we know that the term ε in the upper bound cannot be improved to $\sqrt{\varepsilon}$. However, one could still ask: is it possible that we could instead improve the lower bound for the trace distance, e.g. by establishing that $D_{\max}^{\varepsilon, T, =} + \log \frac{1}{\varepsilon} \leq D_H^{1-\varepsilon}$ holds for all quantum states? This is in fact impossible, as we now argue. Consider two pure states ψ and ϕ with trace distance $\|\psi - \phi\|_+ = \sqrt{1 - |\langle\psi|\phi\rangle|^2} = \sqrt{\varepsilon}$. From Lemma 11 we know that $D_H^\delta(\psi\|\phi)$ is infinite iff $\delta \geq 1 - \varepsilon$, while it is not difficult to verify that $D_{\max}^{\delta, T, =}(\psi\|\phi)$ is infinite iff $\delta < \sqrt{\varepsilon}$. Hence, any relation of the form

$$D_{\max}^{\delta, T, =}(\psi\|\phi) + g(\delta) \leq D_H^{f(\delta)}(\psi\|\phi) \quad (101)$$

must be such that $f(\delta) \geq 1 - \varepsilon$ for all $\delta < \sqrt{\varepsilon}$. The best case scenario is therefore $f(\sqrt{\varepsilon}) = 1 - \varepsilon$, which precisely corresponds to the bound in (95). The fact that the choice of $g(\sqrt{\varepsilon}) = \log \frac{1}{\varepsilon}$ is optimal in general can be verified by considering the trivial case $\psi = \phi$, where $D_{\max}^{\varepsilon, T, =}(\psi\|\psi) = 0$ but $D_H^{1-\varepsilon}(\psi\|\psi) = \log \frac{1}{\varepsilon}$. Our bounds are therefore tight also in this sense, and the smoothing of $D_{\max}^{\varepsilon, T, =}$ exhibits an inherent difference in the scaling of the error with respect to $D_H^{1-\varepsilon}$ between pure states and classical states.

We additionally note that, in light of Lemma 1, potentially tighter one-shot restrictions can be obtained by using the subnormalised smoothing variants of the max-relative entropy. However, the scaling and asymptotic behaviour of

these bounds is essentially the same as the ones given in Theorem 12. We state the bounds here for completeness.

Corollary 13. *For all quantum states ρ and σ and all $\varepsilon \in (0, 1)$, it holds that*

$$D_{\max}^{\sqrt{\varepsilon(1-\frac{3\varepsilon}{4})+\frac{\varepsilon}{2}}, T, \leq}(\rho\|\sigma) + \log \frac{1}{\varepsilon(1-\varepsilon)} \leq D_H^{1-\varepsilon}(\rho\|\sigma), \quad (102)$$

$$D_{\max}^{\sqrt{\varepsilon(2-\varepsilon)}, P, \leq}(\rho\|\sigma) + \log \frac{1}{\varepsilon(1-\varepsilon)} \leq D_H^{1-\varepsilon}(\rho\|\sigma). \quad (103)$$

For classical systems or commuting quantum states, we also have

$$D_{\max}^{\varepsilon, T, \leq}(p\|q) + \log \frac{1}{\varepsilon(1-\varepsilon)} \leq D_H^{1-\varepsilon}(p\|q). \quad (104)$$

These bounds are a direct consequence of Lemma 7 and Corollary 8, with the classical result using Lemma 2. They improve on previously known inequalities that used subnormalised smoothing, e.g. [19].

We remark that an expression of the same exact form as our upper bound on the rightmost side of (96) previously appeared in [71, Lemma III.8] as a tight upper bound on the smooth min-relative entropy (sandwiched Rényi divergence of order $1/2$). It is unclear to us if either of the bounds implies the other or if they are incomparable in general.

VIII. CONSEQUENCES AND OTHER INEQUALITIES

A. Inequalities with Rényi relative entropies

The Petz–Rényi relative entropies D_α [72] and the sandwiched Rényi relative entropies \tilde{D}_α [30], [73] are defined, respectively, as

$$D_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \text{Tr}(\rho^\alpha \sigma^{1-\alpha}), \quad (105)$$

$$\tilde{D}_\alpha(\rho\|\sigma) := \frac{1}{\alpha - 1} \log \text{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha. \quad (106)$$

For $\alpha \in (0, 1)$, it is sufficient that ρ and σ not be orthogonal to ensure that $D_\alpha(\rho\|\sigma)$ and $\tilde{D}_\alpha(\rho\|\sigma)$ are finite; for $\alpha > 1$, the quantities are set to infinity whenever $\text{supp}(\rho) \not\subseteq \text{supp}(\sigma)$. Both of the relative entropies are additive under tensor products: $D_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) = nD_\alpha(\rho\|\sigma)$ and $\tilde{D}_\alpha(\rho^{\otimes n}\|\sigma^{\otimes n}) = n\tilde{D}_\alpha(\rho\|\sigma)$.

We will also employ the measured variant of the quantities, given, in analogy with (32), by

$$D_{\alpha, \mathbb{M}}(\rho\|\sigma) := \sup \left\{ D_\alpha(p_{\rho, M}\|p_{\sigma, M}) \mid M = (M_i)_{i=1}^n \in \mathbb{M}, n \in \mathbb{N}_+ \right\}, \quad (107)$$

where $p_{\rho, M}(i) = \text{Tr } M_i \rho$. Note that the two different quantum definitions (105) and (106) lead to the same notion of measured Rényi relative entropy (107), because they coincide for all classical (commuting) states.

An important property of \tilde{D}_α is that it is asymptotically attained by measurements (see [37, Corollary III.8] and [74, Corollary 4]):

$$\tilde{D}_\alpha(\rho\|\sigma) = \lim_{n \rightarrow \infty} \frac{1}{n} D_{\alpha, \mathbb{M}}(\rho^{\otimes n}\|\sigma^{\otimes n}). \quad (108)$$

Several bounds were given in the literature that connect smooth entropies with Rényi α divergences. Here we discuss how they can be improved using the relations established in this work.

First, we obtain upper bounds.

Corollary 14. *For all $\varepsilon \in (0, 1)$ and all $\alpha > 1$, it holds that*

$$\begin{aligned} & \tilde{D}_{\max}^{\varepsilon}(\rho\|\sigma) + \log \frac{1}{1-\varepsilon} \\ & \leq D_{\alpha, \mathbb{M}}(\rho\|\sigma) + \frac{1}{\alpha-1} \log \frac{1}{\varepsilon} \\ & \leq \tilde{D}_{\alpha}(\rho\|\sigma) + \frac{1}{\alpha-1} \log \frac{1}{\varepsilon}. \end{aligned} \quad (109)$$

As a result,

$$\begin{aligned} D_{\max}^{\varepsilon, T, =}(\rho\|\sigma) & \leq D_{\max}^{\varepsilon, P, =}(\rho\|\sigma) \\ & \leq D_{\alpha, \mathbb{M}}(\rho\|\sigma) + \frac{1}{\alpha-1} \log \frac{1}{\varepsilon^2} \\ & \leq \tilde{D}_{\alpha}(\rho\|\sigma) + \frac{1}{\alpha-1} \log \frac{1}{\varepsilon^2}. \end{aligned} \quad (110)$$

The bound in (110) improves over previously known bounds in [20, Theorem 3] (see also [21, Proposition 6]) and in [25, Proposition 6.22], losing a superfluous additive factor of $\log \frac{1}{1-\varepsilon^2}$ in the former and tightening the smoothing term in the latter. We remark here that some of the previous bounds were stated in terms of the looser sandwiched Rényi relative entropies, but it is clear from their proofs that they apply to $D_{\alpha, \mathbb{M}}$ too.

Proof of Corollary 14. By Lemma 7 we have that

$$\begin{aligned} & \tilde{D}_{\max}^{\varepsilon}(\rho\|\sigma) + \log \frac{1}{\varepsilon(1-\varepsilon)} \\ & \leq D_H^{1-\varepsilon}(\rho\|\sigma) \\ & \leq D_{\alpha, \mathbb{M}}(\rho\|\sigma) + \frac{\alpha}{\alpha-1} \log \frac{1}{\varepsilon} \end{aligned} \quad (111)$$

for all $\alpha > 1$, where the second inequality is a standard argument based on the data processing of the Rényi divergences (see e.g. [37, Lemma IV.7]). Using the Datta–Renner lemma (in particular, Eq. (69) in Corollary 8) then gives the bound for the smooth max-relative entropy, noting that in (110) we changed the smoothing parameter from ε to ε^2 . The fact that $D_{\alpha, \mathbb{M}}(\rho\|\sigma) \leq \tilde{D}_{\alpha}(\rho\|\sigma)$ is a consequence of the data processing inequality for the sandwiched Rényi divergence for all $\alpha > 1$ [75], [76]. \square

To investigate the tightness of the bounds, we turn to a large-deviation-style analysis in terms of error exponents. Here we will use this latter term in an abstract, rather than operational, sense: we do not study any particular information-theoretic task, but rather assume that the smoothing parameter ε of $\tilde{D}_{\max}^{\varepsilon}$ can be understood as the error of some relevant problem. We can then look at the ‘error exponent’ of the quantity $\tilde{D}_{\max}^{\varepsilon}$, that is, the largest exponent E such that $\tilde{D}_{\max}^{\exp(-nE+o(n))}(\rho^{\otimes n}\|\sigma^{\otimes n}) = nR$ for some fixed rate $R > 0$.

Plugging $\varepsilon = \exp(-nE + o(n))$ into (109) and dividing by n , we have in the limit $n \rightarrow \infty$ that

$$\begin{aligned} E & \geq \sup_{\alpha > 1} (\alpha - 1) \left(R - \lim_{n \rightarrow \infty} \frac{1}{n} D_{\alpha, \mathbb{M}}(\rho^{\otimes n}\|\sigma^{\otimes n}) \right) \\ & = \sup_{\alpha > 1} (\alpha - 1) \left(R - \tilde{D}_{\alpha}(\rho\|\sigma) \right), \end{aligned} \quad (112)$$

where we recalled (108). This asymptotic bound is in fact known to be tight: this was established in [37, Theorem IV.4] as a key step in the derivation of the strong converse exponent of quantum hypothesis testing.

For the smooth max-relative entropy, (110) gives an asymptotically tight bound on the error exponent of $D_{\max}^{\varepsilon, P, \leq}$ (and hence, by Lemma 1, also of $D_{\max}^{\varepsilon, P, =}$). To see this, notice that combining Corollary 8 with Lemma 10 says that for any constant $c \in (1, \frac{1}{\varepsilon})$ we have

$$\begin{aligned} D_{\max}^{\sqrt{2c\varepsilon}, P, \leq}(\rho\|\sigma) & \leq \tilde{D}_{\max}^{c\varepsilon}(\rho\|\sigma) \\ & \leq D_{\max}^{\sqrt{\varepsilon}, P, \leq}(\rho\|\sigma) + \log \frac{c}{c-1}. \end{aligned} \quad (113)$$

As constant factors do not affect the asymptotic behaviour when the error vanishes exponentially fast, this shows that the error exponent of $D_{\max}^{\varepsilon, P, \leq}$ is exactly half that of $\tilde{D}_{\max}^{\varepsilon}$, i.e. half of (112), as previously shown in [69]. From Lemma 2 and the above discussion, we also know that the bound of (112) gives exactly the error exponent of $D_{\max}^{\varepsilon, T, \leq}$ when ρ and σ commute.

We can also give a lower bound.

Corollary 15. *For all states ρ, σ , all $\varepsilon \in (0, 1)$, and all $\alpha \in (0, 1)$, it holds that*

$$\begin{aligned} D_{\max}^{\varepsilon, P, \leq}(\rho\|\sigma) & \geq D_{\max}^{\varepsilon, T, \leq}(\rho\|\sigma) \geq \tilde{D}_{\max}^{\varepsilon}(\rho\|\sigma) \\ & \geq D_{\alpha}(\rho\|\sigma) - \frac{1}{1-\alpha} \log \frac{1}{1-\varepsilon}. \end{aligned} \quad (114)$$

Consequently,

$$D_H^{\varepsilon}(\rho\|\sigma) \geq D_{\alpha}(\rho\|\sigma) - \frac{\alpha}{1-\alpha} \log \frac{1}{\varepsilon} + \log \frac{1}{1-\varepsilon}. \quad (115)$$

As a bound on the smooth max-relative entropy, Eq. (114) improves on the bound given in [21, Proposition 4]. The bound on the hypothesis testing relative entropy in (115) is in general incomparable with the best previously known bound of [55, Proposition 3.2], which lacks the final $\log \frac{1}{1-\varepsilon}$ term but instead features an extra α -dependent term. The additional $\log \frac{1}{1-\varepsilon}$ term can, for instance, improve results that employ D_H^{ε} as an intermediate quantity for connections with Rényi divergences, such as a bound on the smooth divergence $\tilde{D}_{1/2}$ in [77, Remark 4].

Proof of Corollary 15. When $\rho \perp \sigma$, all of the terms diverge, and the statement is trivial; let us thus assume otherwise. The first two inequalities are then immediate from the definitions (see Lemma 9). The last one is an application of the inequality of Audenaert et al. [78], which states that $\text{Tr}(A - B)_+ \geq$

$\text{Tr } A - \text{Tr } A^\alpha B^{1-\alpha}$ for all operators $A, B \geq 0$ and all $\alpha \in (0, 1)$. Specifically,

$$\begin{aligned}
 & \tilde{D}_{\max}^\varepsilon(\rho \parallel \sigma) \\
 &= \inf \left\{ \log \lambda \mid \text{Tr}(\rho - \lambda \sigma)_+ \leq \varepsilon \right\} \\
 &\geq \inf \left\{ \log \lambda \mid 1 - \lambda^{1-\alpha} \text{Tr} \rho^\alpha \sigma^{1-\alpha} \leq \varepsilon \right\} \\
 &= \inf \left\{ \log \lambda \mid \log \lambda \geq \frac{1}{1-\alpha} \log(1-\varepsilon) \right. \\
 &\quad \left. - \frac{1}{1-\alpha} \log \text{Tr} \rho^\alpha \sigma^{1-\alpha} \right\} \quad (116) \\
 &= \frac{1}{\alpha-1} \log \text{Tr} \rho^\alpha \sigma^{1-\alpha} - \frac{1}{1-\alpha} \log \frac{1}{1-\varepsilon}.
 \end{aligned}$$

The bound on D_H^ε in (115) then follows by combining this result with Lemma 7. \square

Once again, to study the tightness of the bound, let us look at exponents — now the strong converse exponent of $\tilde{D}_{\max}^\varepsilon$, namely the least E_{sc} such that $\tilde{D}_{\max}^{1-\exp(-nE_{\text{sc}}+o(n))}(\rho^{\otimes n} \parallel \sigma^{\otimes n}) = nR$ for some fixed $R > 0$. Corollary 15 gives a lower bound on this exponent as

$$E_{\text{sc}} \geq \sup_{\alpha \in (0,1)} (\alpha-1)(R - D_\alpha(\rho \parallel \sigma)). \quad (117)$$

This is indeed tight, as can be shown by contradiction. Take $\tilde{D}_{\max}^{1-\exp(-nE_{\text{sc}}+o(n))}(\rho^{\otimes n} \parallel \sigma^{\otimes n}) = nR$ and assume that the inequality in (117) is strict, namely that $E_{\text{sc}} \geq (\alpha-1)(R - D_\alpha(\rho \parallel \sigma)) + \delta$ for all $\alpha \in (0, 1)$ and some fixed $\delta > 0$. Relaxing Lemma 7 gives us the inequality $\tilde{D}_{\max}^{1-\varepsilon}(\rho \parallel \sigma) + \log \frac{1}{\varepsilon} \leq D_H^\varepsilon(\rho \parallel \sigma)$, which implies that $nR + nE_{\text{sc}} + o(n) \leq D_H^{\exp(-nE_{\text{sc}}+o(n))}(\rho^{\otimes n} \parallel \sigma^{\otimes n})$. But then, the converse for the error exponent of hypothesis testing [79], [78] gives

$$\begin{aligned}
 E_{\text{sc}} &\leq \sup_{\alpha \in (0,1)} \frac{\alpha-1}{\alpha} \left(R + E_{\text{sc}} - D_\alpha(\rho \parallel \sigma) \right) + o(1) \\
 &= \sup_{\alpha \in (0,1)} \frac{1}{\alpha} \left((\alpha-1)(R - D_\alpha(\rho \parallel \sigma)) + (\alpha-1)E_{\text{sc}} \right) + o(1) \\
 &\leq \sup_{\alpha \in (0,1)} \frac{1}{\alpha} (E_{\text{sc}} + (\alpha-1)E_{\text{sc}} - \delta) + o(1) \\
 &= E_{\text{sc}} - \delta + o(1), \quad (118)
 \end{aligned}$$

which, in the limit $n \rightarrow \infty$, leads to a contradiction.

For the smooth max-relative entropy, the bound from Corollary 15 does not give a tight lower bound on the strong converse exponent of $D_{\max}^{\varepsilon, P, \leq}$ in general [70]. However, since we have seen the bound to be tight for the strong converse exponent of $\tilde{D}_{\max}^\varepsilon$, by Lemma 2 it is also tight for the strong converse exponent of $D_{\max}^{\varepsilon, T, \leq}$ when ρ and σ are commuting states — this was already shown earlier in [80].

We also remark that an application of Corollary 15 together with the tighter upper bound of Lemma 10 gives another variant of a lower bound for the purified smoothing: for any $c \in (0, 1)$, we have

$$D_{\max}^{\sqrt{c\varepsilon}, P, \leq}(\rho \parallel \sigma) \geq D_\alpha(\rho \parallel \sigma) - \frac{\alpha}{1-\alpha} \log \frac{1}{1-\varepsilon} - \log \frac{1}{1-c}. \quad (119)$$

For clarity, let us give a formal statement of the asymptotic results discussed in this section.

Corollary 16. *For any two states ρ, σ and any $R > 0$, let*

$$\varepsilon_n(R) := \inf \left\{ \varepsilon \in (0, 1) \mid \tilde{D}_{\max}^\varepsilon(\rho^{\otimes n} \parallel \sigma^{\otimes n}) \leq nR \right\}. \quad (120)$$

Then

$$\begin{aligned}
 \lim_{n \rightarrow \infty} -\frac{1}{n} \log \varepsilon_n(R) &= \sup_{\alpha > 1} (\alpha-1) \left(R - \tilde{D}_\alpha(\rho \parallel \sigma) \right), \\
 \lim_{n \rightarrow \infty} -\frac{1}{n} \log (1 - \varepsilon_n(R)) &= \sup_{\alpha \in (0,1)} (\alpha-1) \left(R - D_\alpha(\rho \parallel \sigma) \right). \quad (121)
 \end{aligned}$$

As discussed above, understanding the asymptotic properties of $\tilde{D}_{\max}^\varepsilon$ can find direct use in characterising D_{\max}^ε , which then can be used to study tasks whose performance is connected to this quantity, such as quantum privacy amplification [69], [70], [80]. As another example of how the results here can be related with the analysis of the performance of operational problems, let us consider again the one-shot achievability bounds of [6]. Recall from (44) that [6] gives lower (achievable) bounds on code size of the form $\log M \geq \tilde{D}_{\max}^{1-\varepsilon}(\rho_{XB} \parallel \rho_X \otimes \rho_B)$ where ε represents the error. The error exponent of D_{\max}^ε then corresponds to an achievability result for the strong converse exponent of the tasks studied in [6], and, vice versa, the strong converse exponent of $\tilde{D}_{\max}^\varepsilon$ gives achievability of the error exponents for the tasks. The bound of Corollary 15 then exactly matches the one-shot bound on the error exponent of classical-quantum channel coding given in [6, Proposition 1], namely

$$\log \frac{1}{\varepsilon} \geq n \sup_{\alpha \in (0,1)} (\alpha-1) (\log M - D_\alpha(\rho_{XB} \parallel \rho_X \otimes \rho_B)). \quad (122)$$

On the other hand, the fact that the upper bound of Corollary 14 is asymptotically achievable, as per [37, Theorem IV.4], then yields an achievable strong converse exponent:

$$\log \frac{1}{1-\varepsilon} \geq n \sup_{\alpha > 1} (\alpha-1) \left(\log M - \tilde{D}_\alpha(\rho_{XB} \parallel \rho_X \otimes \rho_B) \right) + o(n). \quad (123)$$

This result did not appear in [6] and in fact it was asked there whether the one-shot achievability result can be used to deduce achievable strong converse exponents, which we see to indeed be possible. Although the exponents obtained in this way do not match the optimal ones, the achievability result is significantly simpler than the more involved proof approaches needed to establish the optimal error [81], [82], [83] and strong converse exponents [84] of classical-quantum channel coding. The fact that this achievability result is not sufficient to give the optimal exponential behaviour of the tasks can be seen from Lemma 7: if either $\varepsilon \sim \exp(-nE)$ or $1-\varepsilon \sim \exp(-nE)$ for some positive exponent E , there is always a non-vanishing gap between the asymptotic expansions of $\tilde{D}_{\max}^{1-\varepsilon}$ and D_H^ε due to the additive factor $\log \frac{1}{\varepsilon(1-\varepsilon)}$ in the lemma; however, the optimal asymptotic exponents of coding tasks generally match those of D_H^ε . Whether one-shot approaches based on smooth

relative entropies can be tightened to give optimal achievability is an interesting open problem.

B. Relation with information spectrum divergence D_s^ε

The divergence $\tilde{D}_{\max}^\varepsilon$ was originally introduced in [28] as an alternative to the *information spectrum divergence*⁴ D_s^ε , with the latter defined as [13]

$$\begin{aligned} D_s^\varepsilon(\rho\|\sigma) &= \log \sup \left\{ \gamma \mid \text{Tr} \rho\{\rho \leq \gamma\sigma\} \leq \varepsilon \right\} \\ &= \log \sup \left\{ \gamma \mid \text{Tr} \rho\{\rho > \gamma\sigma\} \geq 1 - \varepsilon \right\}, \end{aligned} \quad (124)$$

where $\{\rho \leq \gamma\sigma\}$ (respectively, $\{\rho > \gamma\sigma\}$) denotes the projector onto the eigenspace of $\gamma\sigma - \rho$ corresponding to the non-negative (respectively, negative) eigenvalues. The quantity D_s^ε was formalised in [13] to study the precise relations between the information spectrum method [85], [36] and quantum information-theoretic tasks such as quantum hypothesis testing. Here we improve on the known quantitative connections between the different functions.

Proposition 17. *For all ρ, σ , all $\varepsilon \in (0, 1)$, and all $\mu \in (0, 1 - \varepsilon)$, it holds that*

$$\tilde{D}_{\max}^{1-\varepsilon}(\rho\|\sigma) \leq D_s^\varepsilon(\rho\|\sigma) \leq \tilde{D}_{\max}^{1-\varepsilon-\mu}(\rho\|\sigma) + \log \frac{1-\varepsilon}{\mu}. \quad (125)$$

Consequently,

$$D_s^\varepsilon(\rho\|\sigma) + \log \frac{1}{1-\varepsilon} \leq D_H^\varepsilon(\rho\|\sigma) \leq D_s^{\varepsilon+\delta}(\rho\|\sigma) + \log \frac{1}{\delta}. \quad (126)$$

The first of the inequalities in (125) is known [28, Proposition 4.3], as is the second inequality in (126), originally observed in [13, Lemma 12]. Our upper bound of (125) and lower bound of (126) improve on the previous bounds by the additive factor of $\log \frac{1}{1-\varepsilon}$.

Proof of Proposition 17. For the first inequality in (125), take any γ such that $\log \gamma > D_s^\varepsilon$, so that $\text{Tr} \rho\{\rho \leq \gamma\sigma\} > \varepsilon$. This means that

$$1 - \varepsilon \geq \text{Tr} \rho\{\rho > \gamma\sigma\} \geq \text{Tr}(\rho - \gamma\sigma)\{\rho > \gamma\sigma\} = \text{Tr}(\rho - \gamma\sigma)_+ \quad (127)$$

and so γ is a feasible solution for the minimisation in the definition of $\tilde{D}_{\max}^\varepsilon(\rho\|\sigma)$, yielding the claimed inequality upon taking an infimum over such γ .

This immediately implies the second inequality in (126) through the upper bound on D_H^ε stated in Lemma 7.

For the second inequality in (125), we adapt a proof approach of Hayashi from [51, Lemma 19]. Take any λ such that $\text{Tr}(\rho - \lambda\sigma)_+ \leq 1 - \varepsilon - \mu$. Now, fix an arbitrary $\zeta \in (0, \mu)$, let $k = \frac{1-\varepsilon}{\mu-\zeta}$ and assume towards a contradiction that $\gamma = k\lambda$ is a feasible solution for the maximisation in the definition of $D_s^\varepsilon(\rho\|\sigma)$, meaning that $\text{Tr} \rho\{\rho > k\lambda\sigma\} \geq 1 - \varepsilon > 0$. But

$$\frac{1 - \varepsilon - \mu}{1 - \varepsilon} \text{Tr} \rho\{\rho > k\lambda\sigma\}$$

⁴Despite its commonly used name, D_s^ε is not actually a divergence, as it does not obey the data processing inequality.

$$\begin{aligned} &< \frac{1 - \varepsilon - \mu + \zeta}{1 - \varepsilon} \text{Tr} \rho\{\rho > k\lambda\sigma\} \\ &= \left(1 - \frac{1}{k}\right) \text{Tr} \rho\{\rho > k\lambda\sigma\} \\ &= \text{Tr} \rho\{\rho > k\lambda\sigma\} - \frac{1}{k} \text{Tr} \rho\{\rho > k\lambda\sigma\} \quad (128) \\ &\leq \text{Tr} \rho\{\rho > k\lambda\sigma\} - \text{Tr} \lambda\sigma\{\rho > k\lambda\sigma\} \\ &= \text{Tr}(\rho - \lambda\sigma)\{\rho > k\lambda\sigma\} \\ &\leq \text{Tr}(\rho - \lambda\sigma)_+ \\ &\leq 1 - \varepsilon - \mu, \end{aligned}$$

where the penultimate inequality follows from the variational form of $\text{Tr}(\cdot)_+$ (10). We thus get that $\text{Tr} \rho\{\rho > k\lambda\sigma\} < 1 - \varepsilon$, contradicting our assumption of the feasibility of $\gamma = k\lambda$; hence it must be the case that $\log(k\lambda) \geq D_s^\varepsilon(\rho\|\sigma)$, which upon minimisation over λ and taking $\zeta \rightarrow 0$ gives exactly (125).

Taking an infimum of both sides of the second inequality in (125) over all $\mu \in (0, 1 - \varepsilon)$, by Theorem 4 we get exactly the first inequality in (126), which was the last missing piece in the proof. \square

The result of Proposition 17 can help clarify the relations between various one-shot achievability bounds in quantum information, which often appear in terms of either D_s^ε or D_H^ε (or indeed in terms of $\tilde{D}_{\max}^{1-\varepsilon}$, as in (44)). For instance, we get that $\tilde{D}_{\max}^{1-\varepsilon} \geq D_H^{\varepsilon-\delta} - \log \frac{1}{\delta} \geq D_s^{\varepsilon-\delta} - \log \frac{1-\varepsilon}{\delta} + O(\delta)$, tightening the known relations between the one-shot achievability result of [6] and the prior bound of [86], of relevance to their higher-order asymptotic expansion [6, §3.1].

Our bounds on D_s^ε in Proposition 17 are also tight at the level of the asymptotic exponents, when ε converges exponentially fast to 0 or to 1. Indeed, the asymptotics of D_s^ε are nothing but the asymptotics of the quantum Neyman–Pearson tests, well studied in hypothesis testing [36], [55], [37]. To see how these results are implied by our bounds, observe that combining the bounds of Proposition 17 with the data-processing upper bound on D_H^ε recalled in (111) gives

$$\begin{aligned} \tilde{D}_{\max}^{1-\varepsilon}(\rho\|\sigma) &\leq D_s^\varepsilon(\rho\|\sigma) \\ &\leq \tilde{D}_\alpha(\rho\|\sigma) + \frac{1}{\alpha-1} \log \frac{1}{1-\varepsilon}. \end{aligned} \quad (129)$$

As we discussed in Section VIII-A, this upper bound is asymptotically tight for $\tilde{D}_{\max}^{1-\varepsilon}$ when $\varepsilon \rightarrow 0$ by [37, Theorem IV.4], making it asymptotically tight also for D_s^ε . On the other hand, invoking (125) with the choice $\mu = (1-c)\varepsilon$ for some $c \in (0, 1)$ gives

$$\tilde{D}_{\max}^\varepsilon(\rho\|\sigma) \leq D_s^{1-\varepsilon}(\rho\|\sigma) \leq \tilde{D}_{\max}^{c\varepsilon}(\rho\|\sigma) + \log \frac{1}{1-c}, \quad (130)$$

which means that the asymptotic behaviour of $D_s^{1-\varepsilon}$ when $\varepsilon \rightarrow 0$ matches that of $\tilde{D}_{\max}^\varepsilon$. The exponents given by Corollary 16 thus hold also when $\tilde{D}_{\max}^\varepsilon$ is replaced by $D_s^{1-\varepsilon}$.

Although in this work we have emphasised the close connections between $\tilde{D}_{\max}^\varepsilon$ and the smooth max-relative entropy D_{\max}^ε , the insights discussed here strengthen also the interpretation of $\tilde{D}_{\max}^\varepsilon$ as a quantity equivalent, in a sense, to the information spectrum divergence: not only do $\tilde{D}_{\max}^{1-\varepsilon}$ and D_s^ε

have the same second-order asymptotic expansion [28], they also share the same asymptotic behaviour in terms of error and strong converse exponents and similar, tight one-shot bounds.

C. Simultaneous smoothing

Let us now consider a multi-partite generalisation of the Datta–Renner lemma (Theorem 5). This involves smoothing all marginals of a global state at the same time, and it is reminiscent of the ‘simultaneous smoothing’ conjecture of Drescher and Fawzi [68, Conjecture III.1]. While the authors there were only concerned with min-entropy smoothing, intrinsically linked to smoothing of the max-relative entropy with respect to the maximally mixed state, we consider the more general case where the second state is arbitrary. However, an important limitation of our result is that we achieve simultaneous smoothing over non-overlapping sub-systems only, while Drescher and Fawzi are interested in smoothing simultaneously over *all* sets of sub-systems (cf. [68, Theorem VI.2]). The same joint smoothing setting that we consider here appeared in [20].

Proposition 18. *For some positive integer $m \in \mathbb{N}_+$ and some choice of parameters $\varepsilon_1, \dots, \varepsilon_m \in [0, 1]$ with $\sum_j \varepsilon_j < 1$, let $\rho_{1\dots m}$ be an m -partite state such that $\rho_i \leq A_i + Q_i$ holds for all $i = 1, \dots, m$, for some choice of operators $A_i, Q_i \geq 0$ with $\text{Tr } Q_i \leq \varepsilon_i$. Here, ρ_i denotes the marginal of ρ on the i^{th} system. Then there exists an m -partite normalised state $\rho'_{1\dots m}$ such that*

$$\begin{aligned} \rho'_i &\leq \frac{A_i}{1 - \sum_j \varepsilon_j} \quad \forall i = 1, \dots, m \\ \text{and } F(\rho, \rho') &\geq 1 - \sum_j \varepsilon_j, \\ \frac{1}{2} \|\rho - \rho'\|_1 &\leq \sqrt{\sum_j \varepsilon_j}. \end{aligned} \quad (131)$$

Analogous bounds hold for smoothing involving a subnormalised state $\rho'_{1\dots m}$ and its marginals (as in Theorem 5) but we omit them for brevity.

Proof. We generalise the proof of Theorem 5. Setting $G_i := A_i \# ((A_i + Q_i)^{-1})$ and $G := \bigotimes_i G_i$, by monotonicity of the matrix geometric mean, as in (58), we have that $0 \leq G_i \leq (A_i + Q_i) \# ((A_i + Q_i)^{-1}) = \mathbb{1}$. By the same reasoning as in (60), $\text{Tr } \rho(G_i^2 \otimes \mathbb{1}_{i^c}) = \text{Tr } \rho_i G_i^2 \geq 1 - \varepsilon_i$, where $\mathbb{1}_{i^c}$ denotes the identity operator over all sub-systems except the i^{th} . Since $[G_i^2 \otimes \mathbb{1}_{i^c}, G_j^2 \otimes \mathbb{1}_{j^c}] = 0$ for all i, j , it is not difficult to verify by induction that

$$G^2 = \bigotimes_i G_i^2 \geq \mathbb{1} - \sum_i (\mathbb{1} - G_i^2) \otimes \mathbb{1}_{i^c}, \quad (132)$$

from which it follows that

$$\text{Tr } \rho G^2 \geq 1 - \sum_i (1 - \text{Tr } \rho_i G_i^2) \geq 1 - \sum_i \varepsilon_i. \quad (133)$$

See also a very similar argument in [87, Eq. (1.3)].

Since $0 \leq G_i \leq \mathbb{1}$, by taking tensor products and squaring we infer that $0 \leq G^2 \leq \mathbb{1}$. Thus, from the gentle measurement lemma we conclude that the post-measurement state

$$\rho' := \frac{G\rho G}{\text{Tr } G^2 \rho} \quad (134)$$

satisfies that $F(\rho, \rho') \geq 1 - \sum_i \varepsilon_i$. Also,

$$\begin{aligned} \rho'_i &= \frac{1}{\text{Tr } G^2 \rho} G_i \text{Tr}_{i^c} \left[\rho \left(\bigotimes_{j:j \neq i} G_j^2 \right) \right] G_i \\ &\leq \frac{1}{\text{Tr } G^2 \rho} G_i \rho_i G_i \leq \frac{A_i}{\text{Tr } G^2 \rho} \leq \frac{A_i}{1 - \sum_i \varepsilon_i}, \end{aligned} \quad (135)$$

concluding the proof. \square

Combined with our improvements in the weak/strong converse duality, analogously to Theorem 12, we obtain the following improvement over the bound between D_{\max}^ε and $D_H^{1-\varepsilon}$ given in [20, Theorem 5].

Corollary 19. *Let $\rho_{1\dots m}$ be an m -partite state with marginals ρ_i . For all states σ_i acting on the i^{th} system, and for all $\varepsilon_i \in [0, 1)$ such that $\sum_i \varepsilon_i < 1$, there exists a state $\rho'_{1\dots m}$ such that*

$$\frac{1}{2} \|\rho_{1\dots m} - \rho'_{1\dots m}\|_1 \leq P(\rho_{1\dots m}, \rho'_{1\dots m}) \leq \sqrt{\sum_i \varepsilon_i} \quad (136)$$

and

$$D_{\max}(\rho'_i \|\sigma_i) + \log \frac{1 - \sum_j \varepsilon_j}{\varepsilon_i (1 - \varepsilon_i)} \leq D_H^{1-\varepsilon_i}(\rho_i \|\sigma_i) \quad (137)$$

for all $i \in \{1, \dots, m\}$.

Proof. Applying Proposition 18 with the choices of $A_i = \lambda_i \sigma_i$, where $\rho_i \leq \lambda_i \sigma_i + Q_i$ is any feasible solution for $\tilde{D}_{\max}^{\varepsilon_i}(\rho_i \|\sigma_i)$, guarantees the existence of the suitable state ρ' satisfying

$$D_{\max}(\rho'_i \|\sigma_i) \leq \tilde{D}_{\max}^{\varepsilon_i}(\rho_i \|\sigma_i) + \log \frac{1}{1 - \sum_i \varepsilon_i}. \quad (138)$$

From Lemma 7 we then get that $\tilde{D}_{\max}^{\varepsilon_i}(\rho_i \|\sigma_i) + \log \frac{1}{\varepsilon_i (1 - \varepsilon_i)} \leq D_H^{1-\varepsilon_i}(\rho_i \|\sigma_i)$ holds for each i . \square

D. Quantum substate theorem

Another result that can be immediately strengthened by using our bounds is the so-called *quantum substate theorem* [88], [89], which provides an upper bound on the smooth max-relative entropy in terms of the quantum relative entropy $D(\rho \|\sigma)$.

Corollary 20 (Tighter quantum substate theorem). *For all states ρ, σ , and all $\varepsilon \in (0, 1)$ we have*

$$\begin{aligned} D_{\max}^{\varepsilon, T, =}(\rho \|\sigma) &\leq D_{\max}^{\varepsilon, P, =}(\rho \|\sigma) \\ &\leq \frac{D(\rho \|\sigma) + 1}{\varepsilon^2} - \log \frac{1}{\varepsilon^2}. \end{aligned} \quad (139)$$

Proof. From (96) in Theorem 12 we get that $D_{\max}^{\varepsilon, P, =}(\rho \|\sigma) + \log \frac{1}{\varepsilon^2} \leq D_H^{1-\varepsilon^2}(\rho \|\sigma)$, and

$$D_H^{1-\varepsilon^2}(\rho \|\sigma) \leq \frac{D_{\mathbb{M}}(\rho \|\sigma) + h(\varepsilon^2)}{\varepsilon^2} \leq \frac{D(\rho \|\sigma) + 1}{\varepsilon^2} \quad (140)$$

is the standard weak converse bound in the quantum Stein’s lemma [33], with $h(\varepsilon^2)$ denoting the binary entropy function. \square

This improves over the inequality between $D_{\max}^{\varepsilon, P, =}$ and $D(\rho \|\sigma)$ that follows from [89, Theorem 1] by an additive

term of $\log \frac{1}{\varepsilon^2(1-\varepsilon^2)}$. Note that the quantum substate theorem was originally stated in terms of the ‘observational divergence’ $D_{\text{obs}}(\rho\|\sigma) := \sup \left\{ \text{Tr } M\rho \log \frac{\text{Tr } M\rho}{\text{Tr } M\sigma} \mid 0 \leq M \leq \mathbb{1} \right\}$. Using the dual form of $\tilde{D}_{\text{max}}^\varepsilon$ in (43) one can also straightforwardly show the relation $\tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma) \leq \frac{1}{\varepsilon} D_{\text{obs}}(\rho\|\sigma)$, which recovers exactly the statement in [89, Theorem 1] by Corollary 8. However, this leads to a weaker bound between $D_{\text{max}}^{\varepsilon, P, =}$ and $D(\rho\|\sigma)$. Another variant of an upper bound involving $D(\rho\|\sigma)$ was also given in [21, Proposition 5].

IX. AN ALTERNATIVE INTEGRAL REPRESENTATION OF THE UMEGAKI RELATIVE ENTROPY BY ROTATING FRENKEL’S INTEGRAL

Recently, a new integral representation of the Umegaki quantum relative entropy $D(\rho\|\sigma)$ was proposed by Frenkel [90], sparking a wave of interest that resulted in new fundamental insights on quantum relative entropies and their properties [91], [42], [92], [93]. Here, we show that it is possible to ‘rotate’ Frenkel’s integral formula so as to obtain a fundamentally new integral representation of the Umegaki relative entropy $D(\rho\|\sigma)$ that involves the modified version $\tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma)$ of the smooth max-relative entropy (see (17) for a definition). Our result is as follows.

Theorem 21. *For all pairs of quantum states ρ and σ on the same finite-dimensional system, it holds that*

$$\begin{aligned} D(\rho\|\sigma) &= \int_0^1 d\varepsilon \left(\tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma) + (\log e) \left(1 - \exp[-\tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma)] \right) \right)_+ \\ &= \int_0^{\frac{1}{2}\|\rho-\sigma\|_1} d\varepsilon \left(\tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma) + (\log e) \left(1 - \exp[-\tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma)] \right) \right). \end{aligned} \quad (141)$$

These identities hold for every choice of logarithm base, provided that \exp is taken to be the inverse of the log function.

Proof. It is clear that the claim holds for one choice of (positive) logarithm base if and only if it holds for all such choices. Therefore, we can choose without loss of generality the most convenient base to work with, i.e. Euler’s number e .

If $\text{supp } \rho \not\subseteq \text{supp } \sigma$ then all expressions diverge — in the integrals, we have that $\tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma) = +\infty$ for sufficiently small ε . Therefore, from now on we are going to assume that $\text{supp } \rho \subseteq \text{supp } \sigma$. We start from Frenkel’s integral formula [90], written here à la Hirche and Tomamichel [42, Corollary 2.3]:

$$\begin{aligned} D(\rho\|\sigma) &= \int_1^\infty \frac{d\gamma}{\gamma} \text{Tr}(\rho - \gamma\sigma)_+ + \int_1^\infty \frac{d\gamma}{\gamma^2} \text{Tr}(\sigma - \gamma\rho)_+ \\ &= \int_0^\infty dt \text{Tr}(\rho - e^t\sigma)_+ + \int_0^1 du \text{Tr}(\sigma - \frac{1}{u}\rho)_+. \end{aligned} \quad (142)$$

Here, in the second line we introduced two changes of variable: $\gamma = e^t$ for the first integral, and $\gamma = 1/u$ for the second. The function $[0, \infty) \ni t \mapsto \text{Tr}(\rho - e^t\sigma)_+$ is monotonically non-increasing, it evaluates to $\text{Tr}(\rho - \sigma)_+ = \frac{1}{2}\|\rho - \sigma\|_1$ for $t = 0$, and it becomes identically 0 for sufficiently large t . The integral $\int_0^\infty dt \text{Tr}(\rho - e^t\sigma)_+$ is simply the area under

its curve. This area can be computed also in another way, by adding up the areas of horizontal strips instead of vertical ones. For $0 \leq \varepsilon \leq \frac{1}{2}\|\rho - \sigma\|_1$ define

$$f(\varepsilon) := \min \left\{ t : \text{Tr}(\rho - e^t\sigma)_+ \leq \varepsilon \right\} = \tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma) \geq 0. \quad (143)$$

This is just the inverse function to $t \mapsto \text{Tr}(\rho - e^t\sigma)_+$. Then clearly

$$\begin{aligned} &\int_0^\infty dt \text{Tr}(\rho - e^t\sigma)_+ \\ &= \int_0^{\frac{1}{2}\|\rho-\sigma\|_1} d\varepsilon f(\varepsilon) \\ &= \int_0^{\frac{1}{2}\|\rho-\sigma\|_1} d\varepsilon \tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma). \end{aligned} \quad (144)$$

A similar reasoning can be repeated for the second integral. This time $[0, 1] \ni u \mapsto \text{Tr}(\sigma - \frac{1}{u}\rho)_+$ is monotonically non-decreasing, it evaluates to $\text{Tr}(\sigma - \rho)_+ = \frac{1}{2}\|\rho - \sigma\|_1$ for $u = 1$, and its inverse function is

$$\begin{aligned} g(\varepsilon) &:= \max \left\{ u \in [0, 1] : \text{Tr}(\sigma - \frac{1}{u}\rho)_+ \leq \varepsilon \right\} \\ &= \left(\min \left\{ s \in [1, \infty) : \text{Tr}(\sigma - s\rho)_+ \leq \varepsilon \right\} \right)^{-1} \\ &= \exp \left[-\tilde{D}_{\text{max}}^\varepsilon(\sigma\|\rho) \right]. \end{aligned} \quad (145)$$

Hence,

$$\begin{aligned} &\int_0^1 du \text{Tr}(\sigma - \frac{1}{u}\rho)_+ \\ &= \int_0^{\frac{1}{2}\|\rho-\sigma\|_1} d\varepsilon \left(1 - \exp \left[-\tilde{D}_{\text{max}}^\varepsilon(\sigma\|\rho) \right] \right). \end{aligned} \quad (146)$$

The claim is proved once one adds up these two integrals, observing that both integrands are non-negative in the same range $0 \leq \varepsilon \leq \frac{1}{2}\|\rho - \sigma\|_1$. \square

APPENDIX A EQUIVALENT DEFINITIONS OF $\tilde{D}_{\text{max}}^\varepsilon$

Recall that we followed [28], [35] in defining

$$\begin{aligned} \tilde{D}_{\text{max}}^\varepsilon(\rho\|\sigma) &= \log \inf \left\{ \lambda \mid \rho \leq \lambda\sigma + Q, Q \geq 0, \text{Tr } Q \leq \varepsilon \right\} \\ &= \log \sup \left\{ \frac{\text{Tr } W'\rho - \varepsilon}{\text{Tr } W'\sigma} \mid 0 \leq W' \leq \mathbb{1} \right\}. \end{aligned} \quad (A.1)$$

This also mirrors the definition of the classical ε -approximate max-divergence in [43].

In [41], a quantity that we will call the *max-relative entropy smoothed over unnormalised positive operators* was defined as

$$\begin{aligned} D_{\text{max}}^{\varepsilon, \text{pos}}(\rho\|\sigma) &:= \log \inf \left\{ \lambda \mid Z \leq \lambda\sigma, Z \geq 0, \text{Tr}(\rho - Z)_+ \leq \varepsilon \right\}. \end{aligned} \quad (A.2)$$

The trace of the operators Z here is completely unconstrained.

The *max-relative entropy smoothed over Hermitian operators* made an inadvertent and implicit appearance in [45] (cf.

the discussion of the smoothing issue in [41]). We can define it as

$$D_{\max}^{\varepsilon, \text{Herm}, \leq}(\rho||\sigma) := \log \inf \left\{ \lambda \mid X \leq \lambda\sigma, X = X^\dagger, \right. \\ \left. \text{Tr } X \leq 1, \|\rho - X\|_+ \leq \varepsilon \right\}. \quad (\text{A.3})$$

This resembles the definition of $D_{\max}^{\varepsilon, T, \leq}$, but the operators X are not required to be positive semidefinite.

Lemma A.1. *For all quantum states and all $\varepsilon \in [0, 1]$, it holds that*

$$D_{\max}^{\varepsilon, \text{pos}}(\rho||\sigma) = D_{\max}^{\varepsilon, \text{Herm}, \leq}(\rho||\sigma) = \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma). \quad (\text{A.4})$$

The equivalence of $D_{\max}^{\varepsilon, \text{pos}}$ with $\tilde{D}_{\max}^{\varepsilon}$ can already be deduced from [41, Lemma II.6], which showed that $D_{\max}^{\varepsilon, \text{pos}}$ corresponds to the inverse function of the hockey-stick divergence $E_\lambda(\rho||\sigma) = \text{Tr}(\rho - \lambda\sigma)_+$. More generally, all of the quantities can be deduced to be equivalent because they all have been shown to tightly characterise quantum approximate differential privacy [45], [41], [35]. Notwithstanding, we give a concise direct proof.

Proof. Consider any feasible solution for $\tilde{D}_{\max}^{\varepsilon}$ in the form of $\rho \leq \lambda\sigma + Q$ with $Q \geq 0$ and $\text{Tr } Q \leq \varepsilon$. Then the operator $Z = \lambda\sigma$ satisfies $Z \leq \lambda\sigma$ and $\rho - Z \leq Q$, hence $\text{Tr}(\rho - Z)_+ \leq \text{Tr } Q \leq \varepsilon$, implying that $D_{\max}^{\varepsilon, \text{pos}} \leq \log \lambda$. Furthermore, the operator $X = \rho - Q$ satisfies $X \leq \lambda\sigma$, $\text{Tr } X \leq \text{Tr } \rho = 1$ and $\|\rho - X\|_+ = \|Q\|_+ \leq \varepsilon$, and hence $D_{\max}^{\varepsilon, \text{Herm}, \leq} \leq \log \lambda$.

On the other hand, consider first a feasible solution for $D_{\max}^{\varepsilon, \text{pos}}$, namely an operator $Z \geq 0$ such that $Z \leq \lambda\sigma$ and $\text{Tr}(\rho - Z)_+ \leq \varepsilon$. Then $\rho \leq Z + (\rho - Z)_+ \leq \lambda\sigma + (\rho - Z)_+$, so defining $Q = (\rho - Z)_+$ we have a feasible solution for $\tilde{D}_{\max}^{\varepsilon}$. Very similarly, any feasible solution $D_{\max}^{\varepsilon, \text{Herm}, \leq}$, that is, a Hermitian operator X satisfying $X \leq \lambda\sigma$ and $\|\rho - X\|_+ \leq \varepsilon$, gives a feasible solution for $\tilde{D}_{\max}^{\varepsilon}$ as $\rho \leq \lambda\sigma + (\rho - X)_+$, where $\text{Tr}(\rho - X)_+ \leq \varepsilon$ by definition of the generalised trace distance. \square

Note that, in the classical case, all of these smoothing variants reduce to the standard smooth max-relative entropy $D_{\max}^{\varepsilon, T, \leq}$, as per Lemma 2.

One can also define a normalised variant of the Hermitian-smoothed max-relative entropy, $D_{\max}^{\varepsilon, \text{Herm}, =}$, and follow the proof of Lemma 1 to show that

$$D_{\max}^{\varepsilon, \text{Herm}, =}(\rho||\sigma) = \max \left\{ D_{\max}^{\varepsilon, \text{Herm}, \leq}(\rho||\sigma), 0 \right\} \\ = \max \left\{ \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma), 0 \right\}. \quad (\text{A.5})$$

APPENDIX B

CONTINUITY AND STRICT MONOTONICITY PROPERTIES OF $\tilde{D}_{\max}^{\varepsilon}$

This appendix is devoted to the investigation of some analytical properties of the function $\varepsilon \mapsto \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma)$ for any two fixed states ρ, σ . In what follows, we denote with Π_σ the projector onto the support of a state σ , and with $\{\rho \geq \gamma\sigma\}$ the projector onto the positive part of the operator $\rho - \gamma\sigma$. We start with the following simple result.

Lemma B.1. *Let ρ and σ be two finite-dimensional quantum states. Then*

$$\lim_{\gamma \rightarrow \infty} \{\rho \geq \gamma\sigma\} = \mathbb{1} - \Pi_\sigma. \quad (\text{B.1})$$

Proof. If $\Pi_\sigma = \mathbb{1}$, so that σ has no zero eigenvalues, then $\rho \leq \gamma\sigma$ for all sufficiently large γ , and there is nothing to prove. Therefore, let us assume that σ has at least one zero eigenvalue.

Setting $x := 1/\gamma$, we have $\{\rho \geq \gamma\sigma\} = \{\sigma - x\rho \leq 0\}$. Imagine sorting the spectra of σ and $\sigma - x\rho$ in (say) non-increasing order, repeating the eigenvalues according to their multiplicities. By [94, Theorem 5.1 and Eq. (5.2)], every eigenvalue of $\sigma - x\rho$ must converge to the corresponding eigenvalue of σ as $x \rightarrow 0$ (note that [94, Eq. (5.2)] prohibits changes in multiplicities). Now, consider an interval $I = [-\delta, \delta]$, so that δ is strictly smaller than the smallest non-zero eigenvalue of σ . Invoking again [94, Theorem 5.1], we see that the total projector $P_I(x)$ onto the span of all the eigenspaces of $\sigma - x\rho$ whose eigenvalues lie in I is a continuous function of x at $x = 0$. That is, $P_I(0) = \lim_{x \rightarrow 0^+} P_I(x) = \mathbb{1} - \Pi_\sigma$. By Weyl's monotonicity theorem [95, Corollary III.2.3], and sorting the spectra in non-increasing order as before, every eigenvalue of $\sigma - x\rho$ must be smaller than or equal to the corresponding eigenvalue of σ , when $x > 0$. This means that $P_I(x) = \{\sigma - x\rho \leq 0\}$ for all sufficiently small $x > 0$. Therefore, $\lim_{x \rightarrow 0^+} \{\sigma - x\rho \leq 0\} = \lim_{x \rightarrow 0^+} P_I(x) = P_I(0) = \mathbb{1} - \Pi_\sigma$. \square

We are now ready to prove the following.

Lemma B.2. *Let ρ and σ be two finite-dimensional quantum states. Then:*

(a) *If $[\rho, \Pi_\sigma] \neq 0$, the function $\gamma \mapsto \text{Tr}(\rho - \gamma\sigma)_+$ is continuous and strictly decreasing on the whole half-positive real line $\gamma \in [0, \infty)$; furthermore, it holds that*

$$\lim_{\gamma \rightarrow \infty} \text{Tr}(\rho - \gamma\sigma)_+ = 1 - \text{Tr } \rho \Pi_\sigma. \quad (\text{B.2})$$

(b) *If $[\rho, \Pi_\sigma] = 0$, the function $\gamma \mapsto \text{Tr}(\rho - \gamma\sigma)_+$ is continuous, strictly decreasing on the interval $\gamma \in [0, \exp[D_{\max}(\rho \Pi_\sigma || \sigma)]]$, and equal to $1 - \text{Tr } \rho \Pi_\sigma$ for all $\gamma \geq \exp[D_{\max}(\rho \Pi_\sigma || \sigma)]$.*

Consequently:

(c) *If $[\rho, \Pi_\sigma] \neq 0$, the function $\varepsilon \mapsto \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) \in [0, +\infty]$ is equal to $+\infty$ for $\varepsilon \in [0, 1 - \text{Tr } \rho \Pi_\sigma]$, continuous and strictly decreasing for $\varepsilon \in (1 - \text{Tr } \rho \Pi_\sigma, 1]$, and such that*

$$\lim_{\varepsilon \rightarrow (1 - \text{Tr } \rho \Pi_\sigma)^+} \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) = +\infty. \quad (\text{B.3})$$

(d) *If $[\rho, \Pi_\sigma] = 0$, the function $\varepsilon \mapsto \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) \in [0, +\infty]$ is equal to $+\infty$ if $\varepsilon \in [0, 1 - \text{Tr } \rho \Pi_\sigma)$ (this latter interval is understood to be empty if $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$), continuous and strictly decreasing for $\varepsilon \in [1 - \text{Tr } \rho \Pi_\sigma, 1]$, and such that*

$$\lim_{\varepsilon \rightarrow (1 - \text{Tr } \rho \Pi_\sigma)^+} \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) = D_{\max}(\rho \Pi_\sigma || \sigma). \quad (\text{B.4})$$

(e) *In particular, $\varepsilon \mapsto \tilde{D}_{\max}^{\varepsilon}(\rho||\sigma) \in [0, +\infty]$ is always continuous from the right.*

Proof. The continuity of the function $\gamma \mapsto \text{Tr}(\rho - \gamma\sigma)_+ =: E_\gamma(\rho\|\sigma)$ follows immediately from its convexity [42, Lemma 2.1(9)]. The fact that this function is monotonically non-increasing is also well known [36, eq. between Eqs. (17)–(18)], but it is helpful to reproduce the proof of this property here. Assume that $\gamma \leq \gamma'$. Then, as a consequence of the variational form of $\text{Tr}(\cdot)_+$ (Eq. (10)),

$$\begin{aligned} E_\gamma(\rho\|\sigma) &\geq \text{Tr}(\rho - \gamma\sigma)\{\rho \geq \gamma'\sigma\} \\ &= E_{\gamma'}(\rho\|\sigma) + (\gamma' - \gamma) \text{Tr} \sigma\{\rho \geq \gamma'\sigma\} \\ &\geq E_{\gamma'}(\rho\|\sigma). \end{aligned} \quad (\text{B.5})$$

If $\gamma < \gamma'$, equality is only possible when $\text{Tr} \sigma\{\rho \geq \gamma'\sigma\} = 0$. But then, for any $\gamma'' \geq \gamma'$,

$$\begin{aligned} E_{\gamma''}(\rho\|\sigma) &\geq \text{Tr}(\rho - \gamma''\sigma)\{\rho \geq \gamma'\sigma\} \\ &= \text{Tr} \rho\{\rho \geq \gamma'\sigma\} = E_{\gamma'}(\rho\|\sigma), \end{aligned} \quad (\text{B.6})$$

i.e. the function cannot decrease any more. When this happens, the function has reached the value

$$\lim_{\gamma \rightarrow \infty} E_\gamma(\rho\|\sigma) = \text{Tr} \rho(\mathbb{1} - \Pi_\sigma) = 1 - \text{Tr} \rho \Pi_\sigma. \quad (\text{B.7})$$

To see why the first equality holds, observe that, on the one hand,

$$\begin{aligned} E_\gamma(\rho\|\sigma) &= \sup_{0 \leq Q \leq \mathbb{1}} \text{Tr}(\rho - \gamma\sigma)Q \\ &\geq \text{Tr}(\rho - \gamma\sigma)(\mathbb{1} - \Pi_\sigma) = 1 - \text{Tr} \rho \Pi_\sigma, \end{aligned} \quad (\text{B.8})$$

while, on the other,

$$E_\gamma(\rho\|\sigma) = \text{Tr}(\rho - \gamma\sigma)\{\rho \geq \gamma\sigma\} \leq \text{Tr} \rho\{\rho \geq \gamma\sigma\}, \quad (\text{B.9})$$

and the rightmost side tends to $1 - \text{Tr} \rho \Pi_\sigma$ as $\gamma \rightarrow \infty$, due to Lemma B.1. This proves (B.7).

We saw that, when $\gamma < \gamma'$, equality in (B.5) can hold only if $\text{Tr} \sigma\{\rho \geq \gamma'\sigma\} = 0$. Since σ and $\{\rho \geq \gamma'\sigma\}$ are positive semi-definite, this can happen if and only if their supports are orthogonal, entailing in particular that $[\sigma, \{\rho \geq \gamma'\sigma\}] = 0$. From this we deduce that

$$\begin{aligned} 0 &= [\rho - \gamma'\sigma, \{\rho \geq \gamma'\sigma\}] \\ &= [\rho, \{\rho \geq \gamma'\sigma\}] \\ &\xrightarrow{\gamma' \rightarrow \infty} [\rho, \mathbb{1} - \Pi_\sigma] = -[\rho, \Pi_\sigma]. \end{aligned} \quad (\text{B.10})$$

Hence, if $[\rho, \Pi_\sigma] \neq 0$, the function $\gamma \mapsto E_\gamma(\rho\|\sigma)$ must be strictly decreasing on the whole half-positive real line, proving (a). If, on the contrary, $[\rho, \Pi_\sigma] = 0$, then we can decompose $\rho = \rho \Pi_\sigma + \rho(\mathbb{1} - \Pi_\sigma)$, where both addends are (Hermitian and) positive semi-definite. Hence, for an arbitrary $\gamma \geq 0$,

$$\begin{aligned} E_\gamma(\rho\|\sigma) &= \text{Tr}(\rho - \gamma\sigma)_+ \\ &= \text{Tr} \rho(\mathbb{1} - \Pi_\sigma) + \text{Tr}(\rho \Pi_\sigma - \gamma\sigma)_+ \\ &= 1 - \text{Tr} \rho \Pi_\sigma + \text{Tr}(\rho \Pi_\sigma - \gamma\sigma)_+; \end{aligned} \quad (\text{B.11})$$

this is equal to $1 - \text{Tr} \rho \Pi_\sigma$ if and only if $\rho \Pi_\sigma \leq \gamma\sigma$, i.e. if and only if $\gamma \geq \exp[D_{\max}(\rho \Pi_\sigma\|\sigma)]$. This completes the proof of (b).

Claims (c)–(e) follow from easily from (a)–(b), once one remembers Eq. (17), which lets us express the function $\varepsilon \mapsto \tilde{D}_{\max}^\varepsilon(\rho\|\sigma) = \log \inf \left\{ \gamma \geq 0 \mid E_\gamma(\rho\|\sigma) \leq \varepsilon \right\}$ as a generalised inverse of the function $\gamma \mapsto E_\gamma(\rho\|\sigma)$. \square

APPENDIX C

A GENTLER MEASUREMENT LEMMA

Lemma 6. *Let $M \in [0, \mathbb{1}]$ be a measurement operator and ρ be an arbitrary state on the same system. If $\text{Tr} M \rho \geq 1 - \varepsilon$ for some $\varepsilon \in [0, 1]$, then*

$$F\left(\rho, \sqrt{M} \rho \sqrt{M}\right) \geq (1 - \varepsilon)^2, \quad (\text{C.1})$$

$$F\left(\rho, \frac{\sqrt{M} \rho \sqrt{M}}{\text{Tr} M \rho}\right) \geq 1 - \varepsilon, \quad (\text{C.2})$$

$$\frac{1}{2} \left\| \rho - \frac{\sqrt{M} \rho \sqrt{M}}{\text{Tr} M \rho} \right\|_1 \leq \sqrt{\varepsilon}, \quad (\text{C.3})$$

$$\frac{1}{2} \left\| \rho - \sqrt{M} \rho \sqrt{M} \right\|_1 \leq \begin{cases} \sqrt{\varepsilon} \left(1 - \frac{3\varepsilon}{4}\right) & \text{if } \varepsilon \leq 2/3, \\ 1/\sqrt{3} & \text{if } \varepsilon > 2/3 \end{cases} \quad (\text{C.4})$$

$$\begin{aligned} \left\| \rho - \sqrt{M} \rho \sqrt{M} \right\|_+ &\leq \sqrt{\varepsilon} \left(1 - \frac{3\varepsilon}{4}\right) + \frac{\varepsilon}{2} \\ &\leq \sqrt{\varepsilon(2 - \varepsilon)}. \end{aligned} \quad (\text{C.5})$$

All of the primary bounds are tight for all $\varepsilon \in [0, 1]$.

Proof. The two numbers $x := \text{Tr} \rho M$ and $y := \text{Tr} \rho \sqrt{M}$ satisfy the following constraints: first, $x \in [1 - \varepsilon, 1]$; second, $x \leq y$, because $M \leq \sqrt{M}$ as the eigenvalues of M are between 0 and 1; third, $y \leq \sqrt{x}$, which follows easily from the Cauchy–Schwarz inequality. These constraints imply that

$$\begin{aligned} \sqrt{F}\left(\rho, \sqrt{M} \rho \sqrt{M}\right) &= \text{Tr} \sqrt{\sqrt{\rho} \sqrt{M} \rho \sqrt{M} \sqrt{\rho}} \\ &= \text{Tr} \sqrt{\rho} \sqrt{M} \sqrt{\rho} \\ &= \text{Tr} \sqrt{M} \rho = y \geq x \geq 1 - \varepsilon \end{aligned} \quad (\text{C.6})$$

and analogously

$$\begin{aligned} \sqrt{F}\left(\rho, \frac{\sqrt{M} \rho \sqrt{M}}{\text{Tr} M \rho}\right) &= \frac{\text{Tr} \sqrt{\sqrt{\rho} \sqrt{M} \rho \sqrt{M} \sqrt{\rho}}}{\sqrt{\text{Tr} M \rho}} \\ &\geq \frac{x}{\sqrt{x}} \geq \sqrt{1 - \varepsilon}. \end{aligned} \quad (\text{C.7})$$

This proves (C.1) and (C.2), which in turn implies (C.3) immediately due to the Fuchs–van de Graaf inequalities (12). We therefore move on to the proof of (C.4) and (C.5).

Defining the two purifications of ρ and $\sqrt{M} \rho \sqrt{M}$ respectively given by $|\alpha\rangle := (\sqrt{\rho} \otimes \mathbb{1})|\phi\rangle$ (normalised) and $|\beta\rangle := \left(\sqrt{\sqrt{M} \rho \sqrt{M}} \otimes \mathbb{1}\right)|\phi\rangle$ (subnormalised), where $|\phi\rangle := \sum_{i=1}^d |ii\rangle$ is the un-normalised maximally entangled state, by data processing we see that

$$\begin{aligned} \frac{1}{2} \left\| \rho - \sqrt{M} \rho \sqrt{M} \right\|_1 &\leq \frac{1}{2} \|\alpha - \beta\|_1 \\ &= \sqrt{\left(\frac{\text{Tr}[\alpha + \beta]}{2}\right)^2 - \text{Tr} \alpha \beta} \\ &= \sqrt{\left(\frac{1+x}{2}\right)^2 - y^2}. \end{aligned} \quad (\text{C.8})$$

Here, in the first line we introduced the notation $\alpha := |\alpha\rangle\langle\alpha|$ and similarly for β , and the second one follows from the general formula

$$\|X\|_1 = \sqrt{2\text{Tr}(X^2) - (\text{Tr} X)^2} \quad (\text{C.9})$$

valid for all rank-two Hermitian operators with non-positive determinant (cf. proof of Lemma 10). All that remains to do is to prove that

$$\begin{aligned} & \max_{x \in [1-\varepsilon, 1], x \leq y \leq \sqrt{x}} \sqrt{\left(\frac{1+x}{2}\right)^2 - y^2} \\ & = f(\varepsilon) := \begin{cases} \sqrt{\varepsilon(1 - \frac{3\varepsilon}{4})} & \text{if } \varepsilon \leq 2/3, \\ 1/\sqrt{3} & \text{if } \varepsilon > 2/3 \end{cases}. \end{aligned} \quad (\text{C.10})$$

The maximum is always achieved when $y = x$. The rest of the calculation is left to the reader.

The case of the generalised trace distance is analogous. We have

$$\begin{aligned} \left\| \rho - \sqrt{M}\rho\sqrt{M} \right\|_+ & \leq \frac{1}{2} \|\alpha - \beta\|_1 + \frac{1}{2} |\text{Tr}(\alpha - \beta)| \\ & = \sqrt{\left(\frac{1+x}{2}\right)^2 - y^2} + \frac{1}{2}(1-x) \end{aligned} \quad (\text{C.11})$$

using now the data processing inequality for $\|\cdot\|_+$ [25, Prop. 3.8]. An explicit maximisation over the feasible range of x and y yields the stated result.

To see that the bounds are tight, it suffices to consider the single-qubit state $\rho = |0\rangle\langle 0|$ and $M = |\phi_\varepsilon\rangle\langle\phi_\varepsilon|$, where $|\phi_\varepsilon\rangle := \sqrt{1-\varepsilon}|0\rangle + \sqrt{\varepsilon}|1\rangle$. Then on the one hand we have that $\text{Tr} \rho M = 1 - \varepsilon$, while on the other

$$F(\rho, \sqrt{M}\rho\sqrt{M}) = \text{Tr} \rho\sqrt{M} = |\langle 0|\phi_\varepsilon\rangle|^2 = 1 - \varepsilon, \quad (\text{C.12})$$

and moreover

$$\begin{aligned} \frac{1}{2} \left\| \rho - \sqrt{M}\rho\sqrt{M} \right\|_1 & = \frac{1}{2} \| |0\rangle\langle 0| - (1-\varepsilon)|\phi_\varepsilon\rangle\langle\phi_\varepsilon| \|_1 \\ & = \sqrt{\varepsilon \left(1 - \frac{3\varepsilon}{4}\right)}, \\ \left\| \rho - \sqrt{M}\rho\sqrt{M} \right\|_+ & = \sqrt{\varepsilon \left(1 - \frac{3\varepsilon}{4}\right)} + \frac{\varepsilon}{2}. \end{aligned} \quad (\text{C.13})$$

When $\varepsilon > 2/3$, to achieve the tighter bound for $\frac{1}{2}\|\cdot\|_1$ it suffices to choose $\rho = |0\rangle\langle 0|$ and $M = |\phi_{2/3}\rangle\langle\phi_{2/3}|$. This completes the proof. \square

APPENDIX D

HILBERT PROJECTIVE METRIC

As a curiosity for interested readers who made it all the way to Appendix D, we can show that the weak/strong converse duality studied in this work extends beyond the smooth max-relative entropy to the smooth Hilbert projective metric D_Ω , defined as [96], [97]

$$\begin{aligned} D_\Omega^{\varepsilon, \Delta}(\rho\|\sigma) & := \inf_{\rho' \in \mathcal{B}_\Delta^\varepsilon(\rho)} D_\Omega(\rho'\|\sigma), \\ D_\Omega(\rho\|\sigma) & := D_{\max}(\rho\|\sigma) + D_{\max}(\sigma\|\rho), \end{aligned} \quad (\text{D.1})$$

where Δ stands for one of $(T, =)$, (T, \leq) , $(P, =)$, (P, \leq) .

The connection relies on a technical lemma that connects $D_\Omega^{\varepsilon, \Delta}$ with $D_{\max}^{\varepsilon, \Delta}$. One can easily notice that D_Ω is typically much larger than D_{\max} , and indeed $D_\Omega(\rho\|\sigma) = \infty$ unless ρ and σ both have the same support. Perhaps surprisingly, smoothing makes this difference (asymptotically) negligible. The finding below was originally shown to hold in the asymptotic i.i.d. setting [98], although it is not difficult to extract the following one-shot statement from the proof in [98].

Lemma D.1. *For all $\varepsilon \in (0, 1)$ and all $\eta \in (0, \varepsilon)$, it holds that*

$$\begin{aligned} D_\Omega^{\varepsilon+\eta, T, =}(\rho\|\sigma) - \log \frac{1}{\eta} & \leq D_{\max}^{\varepsilon, T, =}(\rho\|\sigma) \leq D_\Omega^{\varepsilon, T, =}(\rho\|\sigma), \\ D_\Omega^{\varepsilon+\eta, P, =}(\rho\|\sigma) - \log \frac{1}{\eta^2} & \leq D_{\max}^{\varepsilon, P, =}(\rho\|\sigma) \leq D_\Omega^{\varepsilon, P, =}(\rho\|\sigma). \end{aligned} \quad (\text{D.2})$$

Proof. The upper bound is obvious from the definition. For the lower bound, consider trace distance first and let ρ' be any feasible state for $D_{\max}^{\varepsilon, T, =}(\rho\|\sigma)$, that is, one such that $\frac{1}{2}\|\rho - \rho'\|_1 \leq \varepsilon$ and $\rho' \leq \lambda\sigma$. Define now

$$\rho'' := (1-\eta)\rho' + \eta\sigma. \quad (\text{D.3})$$

Noticing that

$$\frac{1}{2}\|\rho'' - \rho\|_1 \leq \frac{1}{2}\|\rho'' - \rho'\|_1 + \frac{1}{2}\|\rho' - \rho\|_1 \leq \eta + \varepsilon, \quad (\text{D.4})$$

we have that ρ'' is a feasible state for $D_\Omega^{\varepsilon+\eta, T, =}(\rho\|\sigma)$. We then bound

$$\rho'' \leq (1-\eta)\lambda\sigma + \eta\sigma \leq (1-\eta)\lambda\sigma + \eta\lambda\sigma = \lambda\sigma, \quad (\text{D.5})$$

where we simply used that $\lambda \geq 1$, and

$$\sigma \leq \sigma + \frac{1-\eta}{\eta}\rho' = \frac{1}{\eta}\rho''. \quad (\text{D.6})$$

This altogether gives

$$\begin{aligned} D_\Omega(\rho''\|\sigma) & = D_{\max}(\rho''\|\sigma) + D_{\max}(\sigma\|\rho'') \\ & \leq \log \lambda + \log \frac{1}{\eta}. \end{aligned} \quad (\text{D.7})$$

Optimising over all feasible ρ' concludes the proof of the trace distance case.

The statement for the purified distance is obtained similarly. Given a feasible state ρ' with $P(\rho, \rho') \leq \varepsilon$ we now construct

$$\rho'' := (1-\eta^2)\rho' + \eta^2\sigma, \quad (\text{D.8})$$

which satisfies that

$$P(\rho'', \rho) \leq P(\rho'', \rho') + \varepsilon \quad (\text{D.9})$$

by the triangle inequality. Now, we can use the operator monotonicity of the square root to get

$$F(\rho'', \rho') \geq (1-\eta^2)F(\rho', \rho') = 1 - \eta^2 \quad (\text{D.10})$$

which gives $P(\rho'', \rho') \leq \eta$. The rest of the proof is analogous. \square

Combined with Theorem 12, we obtain a type of weak/strong converse duality relation that relates $D_H^{1-\varepsilon}$ directly with D_Ω^ε .

Corollary D.2. For all $\varepsilon \in (0, 1)$, all $\eta \in (0, 1 - \sqrt{\varepsilon})$, and all $\mu \in (0, \varepsilon)$ it holds that

$$\begin{aligned}
 D_{\Omega}^{\sqrt{\varepsilon}+\eta, T,=}(\rho\|\sigma) &= \log \frac{1}{\eta} + \log \frac{1}{\varepsilon} \\
 &\leq D_H^{1-\varepsilon}(\rho\|\sigma) \\
 &\leq D_{\Omega}^{\varepsilon-\mu, T,=}(\rho\|\sigma) + \log \frac{1}{\mu}, \\
 D_{\Omega}^{\sqrt{\varepsilon}+\eta, P,=}(\rho\|\sigma) &= \log \frac{1}{\eta^2} + \log \frac{1}{\varepsilon} \\
 &\leq D_H^{1-\varepsilon}(\rho\|\sigma) \\
 &\leq D_{\Omega}^{\sqrt{\varepsilon-\mu}, P,=}(\rho\|\sigma) + \log \frac{1}{\mu^2}.
 \end{aligned} \tag{D.11}$$

The result of Lemma D.1 furthermore implies that not only do the smooth max-relative entropy and the smooth Hilbert projective metric have the same first-order asymptotics [98], also their second-order asymptotic expansion is the same — for $D_{\Omega}^{\varepsilon, P,=}$, it is given by the result of [13].

ACKNOWLEDGMENT

We acknowledge helpful discussions with Mario Berta, Marco Tomamichel, and Mark M. Wilde.

REFERENCES

[1] R. Renner, “Security of Quantum Key Distribution,” Ph.D. dissertation, ETH Zurich, 2005.

[2] L. Wang and R. Renner, “One-Shot Classical-Quantum Capacity and Hypothesis Testing,” *Phys. Rev. Lett.*, vol. 108, p. 200501, 2012.

[3] F. Buscemi and N. Datta, “The Quantum Capacity of Channels With Arbitrarily Correlated Noise,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 1447–1460, 2010.

[4] D. Leung and W. Matthews, “On the Power of PPT-Preserving and Non-Signalling Codes,” *IEEE Trans. Inf. Theory*, vol. 61, pp. 4486–4499, 2015.

[5] A. Anshu, R. Jain, and N. A. Warsi, “Building Blocks for Communication Over Noisy Quantum Networks,” *IEEE Trans. Inf. Theory*, vol. 65, pp. 1287–1306, 2019.

[6] H.-C. Cheng, “Simple and Tighter Derivation of Achievability for Classical Communication Over Quantum Channels,” *PRX Quantum*, vol. 4, p. 040330, 2023.

[7] J. M. Renes and R. Renner, “One-Shot Classical Data Compression with Quantum Side Information and the Distillation of Common Randomness or Secret Keys,” *IEEE Trans. Inf. Theory*, vol. 58, pp. 1985–1991, 2012.

[8] F. G. S. L. Brandão and M. B. Plenio, “A Reversible Theory of Entanglement and its Relation to the Second Law,” *Commun. Math. Phys.*, vol. 295, pp. 829–851, 2010.

[9] Z.-W. Liu, K. Bu, and R. Takagi, “One-Shot Operational Quantum Resource Theory,” *Phys. Rev. Lett.*, vol. 123, p. 020401, 2019.

[10] B. Regula, K. Bu, R. Takagi, and Z.-W. Liu, “Benchmarking one-shot distillation in general quantum resource theories,” *Phys. Rev. A*, vol. 101, p. 062315, 2020.

[11] M. Berta, M. Christandl, and R. Renner, “The Quantum Reverse Shannon Theorem Based on One-Shot Information Theory,” *Commun. Math. Phys.*, vol. 306, p. 579, 2011.

[12] K. Fang, X. Wang, M. Tomamichel, and M. Berta, “Quantum Channel Simulation and the Channel’s Smooth Max-Information,” *IEEE Trans. Inf. Theory*, vol. 66, pp. 2129–2140, 2020.

[13] M. Tomamichel and M. Hayashi, “A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks,” *IEEE Trans. Inf. Theory*, vol. 59, pp. 7693–7710, 2013.

[14] Y.-C. Shen, L. Gao, and H.-C. Cheng, “Optimal Second-Order Rates for Quantum Soft Covering and Privacy Amplification,” *IEEE Trans. Inf. Theory*, vol. 70, pp. 5077–5091, 2024.

[15] F. Dupuis, M. Berta, J. Wullschlegler, and R. Renner, “One-Shot Decoupling,” *Commun. Math. Phys.*, vol. 328, pp. 251–284, 2014.

[16] A. Anshu, V. K. Devabathini, and R. Jain, “Quantum Communication Using Coherent Rejection Sampling,” *Phys. Rev. Lett.*, vol. 119, p. 120506, 2017.

[17] N. Datta, “Min- and Max-Relative Entropies and a New Entanglement Monotone,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 2816–2826, 2009.

[18] N. Datta, M. Mosonyi, M. H. Hsieh, and F. G. S. L. Brandão, “A Smooth Entropy Approach to Quantum Hypothesis Testing and the Classical Capacity of Quantum Channels,” *IEEE Trans. Inf. Theory*, vol. 59, pp. 8014–8026, 2013.

[19] F. Dupuis, L. Krämer, P. Faist, J. M. Renes, and R. Renner, “Generalized entropies,” in *XVIIIth International Congress on Mathematical Physics*. World Scientific, 2012, pp. 134–153.

[20] A. Anshu, M. Berta, R. Jain, and M. Tomamichel, “A minimax approach to one-shot entropy inequalities,” *J. Math. Phys.*, vol. 60, p. 122201, 2019.

[21] X. Wang and M. M. Wilde, “Resource theory of asymmetric distinguishability,” *Phys. Rev. Research*, vol. 1, p. 033170, 2019.

[22] M. Tomamichel, R. Colbeck, and R. Renner, “A Fully Quantum Asymptotic Equipartition Property,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 5840–5847, 2009.

[23] F. G. S. L. Brandão and M. B. Plenio, “A Generalization of Quantum Stein’s Lemma,” *Commun. Math. Phys.*, vol. 295, pp. 791–828, 2010.

[24] N. Datta, “Max-relative entropy of entanglement, alias log robustness,” *Int. J. Quantum Inform.*, vol. 07, pp. 475–491, 2009.

[25] M. Tomamichel, *Quantum Information Processing with Finite Resources*. Springer, 2016.

[26] L. Lami, “A Solution of the Generalized Quantum Stein’s Lemma,” *IEEE Trans. Inf. Theory*, vol. 71, pp. 4454–4484, 2025.

[27] L. Lami, M. Berta, and B. Regula, “Asymptotic quantification of entanglement with a single copy,” 2024, preprint arXiv:2408.07067.

[28] N. Datta and F. Leditzky, “Second-Order Asymptotics for Source Coding, Dense Coding, and Pure-State Entanglement Conversions,” *IEEE Trans. Inf. Theory*, vol. 61, pp. 582–608, 2015.

[29] N. Datta and R. Renner, “Smooth Entropies and the Quantum Information Spectrum,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 2807–2815, 2009.

[30] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, “On quantum Rényi entropies: A new generalization and some properties,” *J. Math. Phys.*, vol. 54, p. 122203, 2013.

[31] C. W. Helstrom, *Quantum detection and estimation theory*. New York, USA: Academic Press, 1976.

[32] C. Fuchs and J. van de Graaf, “Cryptographic distinguishability measures for quantum-mechanical states,” *IEEE Trans. Inf. Theory*, vol. 45, pp. 1216–1227, 1999.

[33] F. Hiai and D. Petz, “The proper formula for relative entropy and its asymptotics in quantum probability,” *Commun. Math. Phys.*, vol. 143, pp. 99–114, 1991.

[34] T. Ogawa and H. Nagaoka, “Strong converse and Stein’s lemma in quantum hypothesis testing,” *IEEE Trans. Inf. Theory*, vol. 46, pp. 2428–2433, 2000.

[35] T. Nuradha, Z. Goldfeld, and M. M. Wilde, “Quantum Pufferfish Privacy: A Flexible Privacy Framework for Quantum Systems,” *IEEE Trans. Inf. Theory*, vol. 70, pp. 5731–5762, 2024.

[36] H. Nagaoka and M. Hayashi, “An Information-Spectrum Approach to Classical and Quantum Hypothesis Testing for Simple Hypotheses,” *IEEE Trans. Inf. Theory*, vol. 53, pp. 534–549, 2007.

[37] M. Mosonyi and T. Ogawa, “Quantum Hypothesis Testing and the Operational Interpretation of the Quantum Rényi Relative Entropies,” *Commun. Math. Phys.*, vol. 334, pp. 1617–1648, 2015.

[38] N. Sharma and N. A. Warsi, “Fundamental Bound on the Reliability of Quantum Information Transmission,” *Phys. Rev. Lett.*, vol. 110, p. 080501, 2013.

[39] J. Liu, P. Cuff, and S. Verdú, “ E_{γ} -Resolvability,” *IEEE Trans. Inf. Theory*, vol. 63, pp. 2629–2658, 2017.

[40] W. Yang, R. F. Schaefer, and H. V. Poor, “Wiretap Channels: Nonasymptotic Fundamental Limits,” *IEEE Trans. Inf. Theory*, vol. 65, pp. 4069–4093, 2019.

[41] C. Hirche, C. Rouzé, and D. S. França, “Quantum Differential Privacy: An Information Theory Perspective,” *IEEE Trans. Inf. Theory*, vol. 69, pp. 5771–5787, 2023.

[42] C. Hirche and M. Tomamichel, “Quantum Rényi and f-divergences from integral representations,” *Commun. Math. Phys.*, vol. 405, no. 9, p. 208, 2024.

[43] C. Dwork, G. N. Rothblum, and S. Vadhan, “Boosting and Differential Privacy,” in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, 2010, pp. 51–60.

[44] C. Dwork and A. Roth, “The Algorithmic Foundations of Differential Privacy,” *Found. Trends Theor. Comput. Sci.*, vol. 9, pp. 211–407, 2014.

- [45] L. Zhou and M. Ying, “Differential Privacy in Quantum Computation,” in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 249–262.
- [46] J. M. Renes, “On Privacy Amplification, Lossy Compression, and Their Duality to Channel Coding,” *IEEE Trans. Inf. Theory*, vol. 64, pp. 7792–7801, 2018.
- [47] D. Abdelhadi and J. M. Renes, “On the Second-Order Asymptotics of the Partially Smoothed Conditional Min-Entropy & Application to Quantum Compression,” *IEEE J. Sel. Areas Inf. Theory*, vol. 1, pp. 416–423, 2020.
- [48] R. Renner and S. Wolf, “Smooth Renyi entropy and applications,” in *2004 IEEE International Symposium on Information Theory (ISIT)*, 2004, p. 233.
- [49] M. Tomamichel, R. Colbeck, and R. Renner, “Duality Between Smooth Min- and Max-Entropies,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 4674–4681, 2010.
- [50] R. Bhatia, *Positive Definite Matrices*. Princeton University Press, 2007.
- [51] M. Hayashi, “Security analysis of ϵ -almost dual universal₂ hash functions: Smoothing of min entropy versus smoothing of Rényi entropy of order 2,” *IEEE Trans. Inf. Theory*, vol. 62, pp. 3451–3476, 2016.
- [52] M. J. Donald, “On the relative entropy,” *Commun. Math. Phys.*, vol. 105, pp. 13–34, 1986.
- [53] M. Berta, O. Fawzi, and M. Tomamichel, “On variational expressions for quantum relative entropies,” *Lett. Math. Phys.*, vol. 107, pp. 2239–2265, 2017.
- [54] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.
- [55] K. M. R. Audenaert, M. Mosonyi, and F. Verstraete, “Quantum state discrimination bounds for finite sample size,” *J. Math. Phys.*, vol. 53, p. 122205, 2012.
- [56] G. Gour, “Induced Quantum Divergence: A New Lens on Communication and Source Coding,” 2025, preprint arXiv:2502.13669.
- [57] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel Coding Rate in the Finite Blocklength Regime,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 2307–2359, 2010.
- [58] Y. Polyanskiy, “Channel coding: non-asymptotic fundamental limits,” Ph.D. dissertation, Princeton University, 2010.
- [59] F. Buscemi and G. Gour, “Quantum relative Lorenz curves,” *Phys. Rev. A*, vol. 95, p. 012110, 2017.
- [60] X. Mu, L. Pomatto, P. Strack, and O. Tamuz, “From Blackwell dominance in large samples to Rényi divergences and back again,” *Econometrica*, vol. 89, pp. 475–506, 2021.
- [61] D. Blackwell, “Equivalent Comparisons of Experiments,” *Ann. Math. Stat.*, vol. 24, pp. 265–272, 1953.
- [62] F. Buscemi, “Comparison of Quantum Statistical Models: Equivalent Conditions for Sufficiency,” *Commun. Math. Phys.*, vol. 310, pp. 625–647, 2012.
- [63] J. M. Renes, “Relative submajorization and its use in quantum resource theories,” *J. Math. Phys.*, vol. 57, p. 122202, 2016.
- [64] E. B. Davies, “Quantum stochastic processes,” *Commun. Math. Phys.*, vol. 15, pp. 277–304, 1969.
- [65] A. Winter, “Coding theorem and strong converse for quantum channels,” *IEEE Trans. Inf. Theory*, vol. 45, pp. 2481–2485, 1999.
- [66] M. M. Wilde, *Quantum Information Theory*, 2nd ed. Cambridge: Cambridge University Press, 2017.
- [67] T. Ogawa and H. Nagaoka, “Making Good Codes for Classical-Quantum Channel Coding via Quantum Hypothesis Testing,” *IEEE Trans. Inf. Theory*, vol. 53, pp. 2261–2266, 2007.
- [68] L. Drescher and O. Fawzi, “On simultaneous min-entropy smoothing,” in *2013 IEEE International Symposium on Information Theory (ISIT)*, 2013, pp. 161–165.
- [69] K. Li, Y. Yao, and M. Hayashi, “Tight Exponential Analysis for Smoothing the Max-Relative Entropy and for Quantum Privacy Amplification,” *IEEE Trans. Inf. Theory*, vol. 69, pp. 1680–1694, 2023.
- [70] K. Li and Y. Yao, “Operational Interpretation of the Sandwiched Rényi Divergence of Order 1/2 to 1 as Strong Converse Exponents,” *Commun. Math. Phys.*, vol. 405, p. 22, 2024.
- [71] N. Ramakrishnan, M. Tomamichel, and M. Berta, “Moderate Deviation Expansion for Fully Quantum Tasks,” *IEEE Trans. Inf. Theory*, vol. 69, pp. 5041–5059, 2023.
- [72] D. Petz, “Quasi-entropies for finite quantum systems,” *Rep. Math. Phys.*, vol. 23, pp. 57–65, 1986.
- [73] M. M. Wilde, A. Winter, and D. Yang, “Strong Converse for the Classical Capacity of Entanglement-Breaking and Hadamard Channels via a Sandwiched Rényi Relative Entropy,” *Commun. Math. Phys.*, vol. 331, pp. 593–622, 2014.
- [74] M. Hayashi and M. Tomamichel, “Correlation detection and an operational interpretation of the Rényi mutual information,” *J. Math. Phys.*, vol. 57, p. 102201, 2016.
- [75] R. L. Frank and E. H. Lieb, “Monotonicity of a relative Rényi entropy,” *J. Math. Phys.*, vol. 54, 2013.
- [76] S. Beigi, “Sandwiched Rényi divergence satisfies data processing inequality,” *J. Math. Phys.*, vol. 54, p. 122202, 2013.
- [77] T. Nuradha and M. M. Wilde, “Fidelity-Based Smooth Min-Relative Entropy: Properties and Applications,” *IEEE Trans. Inf. Theory*, vol. 70, pp. 4170–4196, 2024.
- [78] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acín, and F. Verstraete, “Discriminating States: The Quantum Chernoff Bound,” *Phys. Rev. Lett.*, vol. 98, p. 160501, 2007.
- [79] H. Nagaoka, “The Converse Part of Theorem for Quantum Hoeffding Bound,” 2006, preprint arXiv:quant-ph/0611289.
- [80] R. Salzmann and N. Datta, “Total insecurity of communication via strong converse for quantum privacy amplification,” 2022, preprint arXiv:2202.11090.
- [81] J. M. Renes, “Tight Lower Bound on the Error Exponent of Classical-Quantum Channels,” *IEEE Trans. Inf. Theory*, vol. 71, pp. 530–538, 2025.
- [82] K. Li and D. Yang, “Reliability Function of Classical-Quantum Channels,” *Phys. Rev. Lett.*, vol. 134, p. 010802, 2025.
- [83] H.-C. Cheng and P.-C. Liu, “Error Exponents for Quantum Packing Problems via An Operator Layer Cake Theorem,” 2025, preprint arXiv:2507.06232.
- [84] M. Mosonyi and T. Ogawa, “Strong Converse Exponent for Classical-Quantum Channel Coding,” *Commun. Math. Phys.*, vol. 355, pp. 373–426, 2017.
- [85] T. S. Han, *Information-Spectrum Methods in Information Theory*, ser. Stochastic Modelling and Applied Probability, B. Rozovskii and M. Yor, Eds. Berlin, Heidelberg: Springer, 2003, vol. 50.
- [86] S. Beigi and A. Gohari, “Quantum Achievability Proof via Collision Relative Entropy,” *IEEE Trans. Inf. Theory*, vol. 60, pp. 7980–7986, 2014.
- [87] S. Khabbazi Oskoueï, S. Mancini, and M. M. Wilde, “Union bound for quantum information processing,” *Proc. R. Soc. A: Math. Phys. Eng. Sci.*, vol. 475, p. 20180612, 2019.
- [88] R. Jain, J. Radhakrishnan, and P. Sen, “Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states,” in *The 43rd Annual IEEE Symposium on Foundations of Computer Science*, 2002, pp. 429–438.
- [89] R. Jain and A. Nayak, “Short Proofs of the Quantum Substate Theorem,” *IEEE Trans. Inf. Theory*, vol. 58, pp. 3664–3669, 2012.
- [90] P. E. Frenkel, “Integral formula for quantum relative entropy implies data processing inequality,” *Quantum*, vol. 7, p. 1102, 2023.
- [91] A. Jenčová, “Recoverability of quantum channels via hypothesis testing,” *Lett. Math. Phys.*, vol. 114, no. 1, p. 31, 2024.
- [92] M. Berta, L. Lami, and M. Tomamichel, “Continuity of Entropies via Integral Representations,” *IEEE Transactions on Information Theory*, vol. 71, pp. 1896–1908, 2025.
- [93] S. Beigi, C. Hirche, and M. Tomamichel, “Some properties and applications of the new quantum f -divergences,” 2025, preprint arXiv:2501.03799.
- [94] T. Kato, *Perturbation theory for linear operators*, ser. Die Grundlehren der mathematischen Wissenschaften, Band 132. Springer-Verlag New York, Inc., New York, 1966.
- [95] R. Bhatia, *Matrix Analysis*. Springer, 1996.
- [96] P. J. Bushell, “Hilbert’s metric and positive contraction mappings in a Banach space,” *Arch. Rat. Mech. Anal.*, vol. 52, pp. 330–338, 1973.
- [97] D. Reeb, M. J. Kastoryano, and M. M. Wolf, “Hilbert’s projective metric in quantum information theory,” *J. Math. Phys.*, vol. 52, p. 082201, 2011.
- [98] B. Regula, L. Lami, and M. M. Wilde, “Overcoming entropic limitations on asymptotic state transformations through probabilistic protocols,” *Phys. Rev. A*, vol. 107, p. 042401, 2023.

Bartosz Regula received the MMath and PhD degrees in mathematics from the University of Nottingham in 2015 and 2018, respectively. Subsequently, he was a Presidential Postdoctoral Fellow at Nanyang Technological University, Singapore, and then a JSPS Postdoctoral Fellow at the University of Tokyo. He is now a RIKEN Hakubi Team Leader at the RIKEN Center for Quantum Computing in Wako, Japan. His research interests include quantum resource

theories, quantum information theory, and filling out biographies in journal submissions.

Ludovico Lami received the B.Sc. and M.Sc. degrees in physics from the University of Pisa, Italy, in 2012 and 2014, respectively, the Diploma degree in physics from Scuola Normale Superiore, Pisa, Italy, in 2015, and the Ph.D. degree from the Autonomous University of Barcelona, Spain, in 2017. He was an Alexander von Humboldt Research Fellow with the University of Ulm, Germany, from 2020 to 2022, and an Assistant Professor at the University of Amsterdam and QuSoft from 2022 to 2024. He is currently an Assistant Professor (RTdB) with the Scuola Normale Superiore. His research interests include quantum information, entanglement theory, quantum Shannon theory, and foundational aspects of quantum physics

Nilanjana Datta received the Ph.D. degree in mathematical physics from ETH Zurich, Switzerland, in 1996. From 1997 to 2000, she held post-doctoral positions in the Dublin Institute of Advanced Studies, C.N.R.S. Marseille, and EPFL, Lausanne. In 2001, she joined the University of Cambridge, U.K., as a Lecturer in mathematics in Pembroke College, and an Affiliated Lecturer of the Statistical Laboratory, Centre for Mathematical Sciences. She is currently the Professor of Quantum Information Theory in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, and a Fellow of Pembroke College. Her scientific interests include quantum information theory and mathematical physics.