Quantum Data Processing Inequalities and their Reverse

by

Shreyas Iyer

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Applied Mathematics (Quantum Information)

Waterloo, Ontario, Canada, 2025

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Any reasonable measure of quantum information must satisfy a data processing inequality, that is, it must not increase under the action of a quantum channel. The same is, therefore, true for measures of distinguishability of quantum states. In this thesis, we study two families of distinguishability measures that are particularly interesting: the Riemannian metric (more precisely, the corresponding semi-norm) and the standard quantum f-divergences (sometimes referred to as just standard f-divergences). However, rather than focusing on the information lost, we ask about the information preserved - namely, a reverse data processing inequality. As is established in this thesis, an exact reverse data processing inequality for all states acted on by a specific channel is not possible for these measures if the output dimension of the quantum channel is no greater than the input dimension (which includes several important channels). Instead, we settle for a reverse data processing inequality on a restricted set of input states, or oftentimes it suffices to only compare the loss of information incurred via two given quantum channels in general. This thesis demonstrates cases of a restricted reverse data processing inequality for these measures and initiates a study of the similarities between the Riemannian metrics and standard quantum f-divergences in this context.

Acknowledgements

This thesis would not have been possible if it were not for everyone who has taken the time to excite and motivate me to pursue my research in quantum information science.

It is my pleasure to thank my supervisor, Graeme Smith, for all of his support and for his role in inspiring me. He became, and forever shall be, the first person to hand me an opportunity to pursue this field when he took me on as a graduate student. His boundless and broad enthusiasm for research has always shone through in our discussions and group meetings. And he has been encouraging of my many (many!) interests over the past two years, allowing me the freedom to decide on my place as a researcher.

I would also like to share my appreciation for the bumbling research environment at the Institute for Quantum Computing. The many talks and enthusiastic researchers have kept me going. There is no shortage of people to discuss maths with! In particular, Richard Cleve has been very great for this, ever since my first year at Waterloo, when he was always available after his QIC 710 lectures to chat about the broader picture of quantum computing. I also thank: the postdocs in my group - Yunkai Wang, Paula Belzig and Peixue Wu - for offering projects that helped to expand my horizons and for their dedication to making advanced topics seem more approachable; and Amolak Kalra for our many chats about connections between quantum error correction and quantum information theory.

This thesis is, in fact, the consequence of a project proposed by Peixue; after attending QIP 2025, he was very supportive of my newfound interest in the fundamental properties of entropic quantities, so he suggested working to extend the ideas from their paper [6]. His suggestions for papers/topics to read up and present on, and his role as a mentor to discuss results with, ultimately prepared me in a relatively short amount of time to produce the interesting results presented in this thesis. Paula and Graeme were also great to discuss these ideas with, and helped me loosen up and focus on playing around with the ideas.

There are some experiences that I find in hindsight to have been irreplaceable. I thank Christoph Hirche, Albert H. Werner, Jan Philip Solovej, Matthias Christandl and Suzanne Anderson, for organising the fantastic masterclass in Copenhagen on 'Entropy Inequalities in Quantum Information Science' that inspired me as an undergraduate student to switch into this field, after I learned about the rich mathematics involved and engaged in many insightful discussions with researchers. I also thank Ophelia Crawford and Alexandra Moylett for being excellent supervisors and friends during my summer internship at Riverlane, where I worked in their fault-tolerance group. Our frequent whiteboard discussion helped me to adjust to my project quickly and it actually broke me out of my shell. The group there was very friendly and full of highly specialised researchers; it was great to learn so much about the ongoing work at Riverlane and to observe their work ethic.

I would also like to express my gratitude to everyone in my life outside of quantum information science.

From my time as a Cambridge undergraduate student, I am grateful to my supportive friend and supervisor (and at one point, my Director of Studies) Johannes Pausch, whose support has been invaluable. He serves as a figure of hard work and dedication, and it is great that we still keep in touch on occasion. I thank my fellow STEM undergraduates at St Catharine's College, whom I could relate to much more than most people. I would especially like to thank my friends Caitanya Durley, Thomas van Dongen, Juan Pottecher, Ed Wheeler, Kazal Oshodi, Kusal Fernando, Joseph Clarke, for always being there.

And from before even my time at Cambridge, I would like to thank Kenneth Jiang. A talented physicist, whom I wish to meet again one day. It is rare to have such a rivalry, that lacks the stress of being in competition, because we both wanted the best for each other. I will never forget his sense of humour or the fact that Kenneth is the one who pushed me to work hard enough that I got into Cambridge. He still drives me forward.

I thank my best friend, Simon Kwan, whom I can always rely on to cheer me up. I don't speak to anyone more than I speak to him. He's been there through all of my tough times. He reminds me that it's never the end of world when something doesn't work out.

Finally, my family. My mother, grandmother and late grandfather are the world to me, and it is with their unwavering love and support in my academics that I have overcome situations that were seemingly insurmountable. My grandfather is fundamentally my idol, and I dream to be like him. My mother, who mostly takes after him, actually reopened a path for me and the scientific research career I aspired for, early on at a time in high school when I was told I lack a talent for the sciences. After all, she is probably the very reason I've wanted to be a scientist for so long. My grandmother shares her love mostly via food, but it is nonetheless a welcome form of affection.

Table of Contents

Author's Declaration **ii**

Abstract **iii**

Acknowledgements **iv**

1 Introduction **1**

1.1 Motivation 1

1.2 Main results and thesis outline 4

2 Preliminaries **7**

2.1 C^*-algebras, Quantum States and Introducing Quantum f-Divergences . . . 7

2.2 Operator Convex Functions and Monotone Riemannian Metrics 10

2.3 Examples of Divergences and Riemannian Metrics 17

2.4 Qubits: Obtaining Explicit Expressions for the Riemannian Metrics 19

3 Properties of Quantum f-Divergences **28**

3.1 Classical f-divergences 28

3.2 Standard f-divergences 31

3.2.1 Monotonicity 34

3.3 Other Quantum f-Divergences 40

3.3.1 Maximal f-Divergences 40

3.3.2 Measured f-Divergences 46
3.3.3 Sandwiched Rényi Divergence 49

4 Reverse-type data processing inequalities for standard f-divergences 53
4.1 An Introduction to Contraction and Expansion Coefficients 53
4.2 Some Subtleties 55
4.3 No Reverse Data Processing over all States 57
4.4 Other Non-Relative Expansion Coefficients can be Positive 60
4.5 The Impact of Classical Output 63
4.6 The Equivalence Between Relative Expansion Coefficients, and the Maximal Metric 71
4.6.1 Cases of Equality between the Divergence and Riemannian Contraction and Expansion Coefficients 71
4.6.2 A Generic Relationship between the Divergence and Riemannian Contraction and Expansion Coefficients 72
4.6.3 Consequences of having Bounded Metrics 74
4.6.4 The Impact of the Maximal Metric 76
4.7 Relative Expansion Coefficients for Quantum Channels with Only Full Rank Output States 83
4.7.1 Primitive Channels and Quantum Markov Chains 87
4.8 Explicit Lower Bounds on the Riemannian Relative Expansion Coefficients 88
4.8.1 Generalised Dephasing Channel 88
4.9 Qubit Calculations 91
4.9.1 A Case of Strict Inequality between the Riemannian and Divergence Coefficients 91
4.9.2 Dephasing Channel and Amplitude Damping Channel 95

5 Conclusion 102

References 104

APPENDICES 109

A Local Symmetry of quantum f-divergences 110

B Lemmas for the DPI of Sandwiched Rényi Divergences 113

C Some Perturbation Theory 114

C.1 Rellich's Theorem 114

D Proof of Lemma 4.2 116

Chapter 1

Introduction

1.1 Motivation

How can we quantify the difference between quantum states acting on the same Hilbert space? This is a fundamental question in quantum information theory. Some widely applicable distinguishability measures can be motivated via the error probabilities of a *state discrimination problem* [21]. A challenger samples some unknown state ω from two known states $\{\rho, \gamma\}$, and the tester performs a POVM to decide between the simple hypotheses:

$$H_0 : \omega = \rho, \quad H_1 : \omega = \gamma$$

Depending on the testing paradigm, one may favour H_0 and set it to the default - this can be considered as *asymmetric hypothesis testing* - or they can be completely unbiased (up to a prior) - and this corresponds to *symmetric hypothesis testing*. The tester may be tasked to distinguish between $\{\rho^{\otimes n}, \gamma^{\otimes n}\}$ for large n, and in this case the probability of failure (or the probability of a Type II error in the asymmetric setting) roughly takes the form 2^{-nD}, for some constant positive *error exponent* D. The error exponent is recognised as a measure of distinguishability between ρ and γ. The different cases are as follows:

- **Symmetric hypothesis testing, single copy** $(n = 1)$. The optimal error probability gives rise to the *trace distance* (Holevo-Helstrom bound).[22, 33, 32]
- **Symmetric hypothesis testing, many copies** $(n \gg 1)$. The asymptotic optimal error exponent gives rise to the *Petz-Rényi divergences* $D_\alpha(\rho\|\gamma)$ via the Chernoff bound, i.e. $D = \min_{0 \leqslant \alpha \leqslant 1} D_\alpha(\rho\|\gamma)$ [3, 45].

- **Asymmetric hypothesis testing, many copies ($n \gg 1$).** The asymptotic optimal Type-II error exponent, for bounded Type-I error, gives rise to the *(Umegaki) quantum relative entropy* $D = D(\rho\|\gamma)$ via the quantum Stein's lemma [38, 27].

Many interesting distinguishability measures are based on these, such as the sandwiched Rényi divergence (see Section 3.3.3) and the quantum Hockey stick divergences [31], that have applications to different hypothesis testing paradigms and resource-theoretic tasks. Here, we are interested in the family of *standard f-divergences*. To understand them, let us first recall the quantum relative entropy for two density operators $\rho, \gamma \in \mathcal{B}(\mathcal{H})$:

$$D(\rho\|\gamma) = \begin{cases} \mathrm{Tr}\left(\rho\big(\log\rho - \log\gamma\big)\right) & \text{if } \mathrm{supp}(\rho) \leqslant \mathrm{supp}(\gamma), \\ \infty & \text{else.} \end{cases} \tag{1.1}$$

The quantum relative entropy, just like its classical counterpart, satisfies many convenient properties that one might hope for from a good measure of distinguishability (despite its lack of symmetry): positivity, continuity (as long as $\mathrm{supp}\,\rho \leqslant \mathrm{supp}\,\gamma$), invariance under isometries or even under tensoring both arguments by the same state, and the property of monotonicity (the 'data processing inequality') that is especially important to this thesis [56, 40, 60, 50, 27]:

$$D(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)) \leqslant D(\rho\|\gamma), \text{ for all states } \rho, \gamma, \text{CPTP maps } \mathcal{N}. \tag{1.2}$$

Any meaningful distinguishability measure is expected to satisfy monotonicity, because it should capture the information lost about the states that have passed through a (potentially noisy) quantum channel. As it turns out, the above additionally applies to positive maps $\mathcal{N}$ that satisfy some reasonable conditions [26, 44].

A key observation is that the above properties do not uniquely characterise the relative entropy, and we can generalise to *quantum f-divergences* that can also satisfy all of them [26, 25]. A quantum f-divergence is a distinguishability measure D_f that reduces to one of *Csiszár's classical f-divergences* (that generalise the classical relative entropy) when restricted to commuting quantum states. The specific subset of standard f-divergences are defined by

$$D_f^{\mathrm{std}}(\rho\|\gamma) = \mathrm{Tr}\,\gamma^{1/2} f(\Delta_{\rho,\gamma})(\gamma^{1/2}), \quad \Delta_{\rho,\gamma}(X) := \rho X \gamma^{-1}.$$

They are interesting because sometimes conditions of the form

$$D_f^{\mathrm{std}}(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)) = D_f^{\mathrm{std}}(\rho\|\gamma) \tag{1.3}$$

can imply that $\mathcal{N}$ can be reversed on the set $\{\rho, \gamma\}$ with a channel depending only on $\mathcal{N}, \gamma$ [48][36]. In particular, this is true if the equality (1.3) holds for the relative entropy alone.

This is important in the context of quantum error correction, where we want to recover states in some code space after they've been acted on by a noisy channel. Recently, it has been discovered that the preservation of the Riemannian semi-norms - distinguishability measures which correspond to the local behaviours of the standard f-divergences (see Section 2.2) - suffice for reversibility. This thesis develops the connections between standard f-divergences and Riemannian semi-norms.

In a way, the theme of this essay is to compare the behaviour of different measures of distinguishability. Specifically, this thesis initiates a discussion of the following meta question:

How do various distinguishability measures compare the amount of distinguishability preserved by two given quantum channels?

The relevant quantities that we'll deal with are the *contraction coefficients* and *relative expansion coefficients.* Previous works have compared quantum channels via the amount of information they preserve, i.e. several frameworks assign a partial order between quantum channels and are related to contraction coefficients [30, 9]. For example, [6] constructed a less-noisy yet non-degradable (amplitude damping) channel using an estimate based on contraction and relative expansion coefficients; moreover, a complete version of such bounds would have implications for long–standing additivity problems in quantum capacities [57].

A large body of previous work [2, 11, 12, 17, 18, 19, 20, 28, 31, 37, 59] has focused on the *strong data processing inequality* (SDPI) for a quantum channel $\mathcal{N}$. Concretely, one asks whether there exists a contraction coefficient $c < 1$ such that

$$D_f(\mathcal{N}(\rho) \,\|\, \mathcal{N}(\gamma)) \leqslant c\, D_f(\rho\|\gamma), \qquad \forall \rho \neq \gamma,$$

for a quantum f-divergence D_f. An analogous question can be asked for the Riemannian semi-norm induced by D_f. When the channel admits a unique fixed point, the SDPI (with the reference state given by this fixed point) is closely connected to the *modified logarithmic Sobolev inequality* (MLSI), which often yields sharper mixing time estimates than spectral gap methods. This connection has been particularly useful for bounding the mixing times of quantum Markov chains, especially in the case of primitive channels (i.e., those with a unique full-rank fixed point) [2, 17, 18, 19, 20, 59].

In contrast, one can investigate a form of *reverse data processing inequality* (RDPT) for a quantum channel $\mathcal{N}$: namely, does there exist a *relative expansion coefficient* $c > 0$ and a quantum channel $\mathcal{M}$ such that

$$D_f(\mathcal{N}(\rho) \,\|\, \mathcal{N}(\gamma)) \;\geqslant\; c\, D_f(\rho\|\gamma),\; \forall \rho \neq \gamma \in \operatorname{Im} \mathcal{M}\,?$$

We will see later that quantum Markov chains based on primitive channels again provide a natural application of this perspective (see Section 4.7.1).

This thesis will explore the quantum data processing inequality (1.2) and the existence of a reverse quantum data processing inequality in the context of (especially) the standard quantum f-divergences.

1.2 Main results and thesis outline

The following is a brief summary of the main results in each chapter.

Chapter 2: Preliminaries
The first section (Section 2.1) briefly explains the origins of standard quantum f-divergences. Formally, the standard (and maximal) quantum f-divergence arises from the GNS representations of quantum states [25], and this section attempts to justify this. The standard quantum f-divergence is defined and recognised as a generalisation of the quantum relative entropy.

The second section (Section 2.2) is a walkthrough on operator convex functions and (monotone) Riemannian metrics. Riemannian metrics are motivated as a standalone measure of distinguishability of two states, and established as a generalisation of the classical Fisher information. Many fundamental properties about Riemannian metrics and operator convex functions are introduced here, as they will be essential in the remaining chapters. Importantly, take note of the integral representations for operator convex functions ((2.6), (2.7)) and the lack of positivity and continuity of the Riemannian metric for general choices of states. The Riemannian metric is associated with the second order local behaviour of quantum f-divergences (for input states with equal support).

Sections 2.3 and 2.4 present results about some special cases of the standard f-divergence and Riemannian metrics. Some of these results propagate throughout the thesis, specifically the maximal metric coincides with the quadratic relative entropy/maximal divergence. Other results offer inspiration, for example Lemma 2.2 provided the option to generalise the qubit calculations in [6] (Section 4.8) and ultimately led to Section 4.7 in a successful attempt to demonstrate the positivity of relative expansion coefficients for full Kraus rank Pauli channels.

Chapter 3: Properties of Quantum f-Divergences
As every quantum f-divergence generalises a classical f-divergence, Section 3.1 introduces the classical f-divergence and it's properties. A note is made about the Radon-Nikodym derivative, which does not have a unique non-commutative generalisation (which contributes to the existence of distinct quantum f-divergences). In Section 3.2, the standard

quantum f-divergence is recognised as the classical f-divergence w.r.t. the Nussbaum-Szkola distribution of the two input states, allowing it to immediately acquire properties and relations satisfied by the classical f-divergence (like, positivity). Furthermore, it turns out that many essential properties commonly associated to the relative entropy, are satisfied by standard quantum f-divergences in general. Because of it's importance in the theme of this thesis, the proof of monotonicity is summarised.

Section 3.3 explores other quantum f-divergences. It talks through some of their properties and the connections between them. While this section is not a key element in the rest of this thesis, Theorem 4.3 applies to all of these quantum f-divergences, i.e. they also lack a reverse data processing inequality over all input states. This may warrant future investigation.

Chapter 4: Reverse-type DPI's for Standard f-divergences
This section consists primarily of my new results. Some of them are obtained by essentially replicating a proof of from another paper (just, in this new context), and acknowledgement is provided accordingly.

Sections 4.1, 4.2 prepare the groundwork on contraction and relative expansion coefficients. They are compared with the (degenerate) classical setting, and the implications of the continuity properties of any particular standard f-divergence are assessed to identify alternative optimisation regions that define $\check{\eta}_f^{\text{std}}$.

Section 4.3 extends the proof technique of [6] to show that there cannot be any reverse data processing inequality over all states w.r.t. the quantum f-divergences or the Riemannian metrics. There is hence a discussion about how these relative expansion coefficients compare to those w.r.t. norms in Section 4.4; this plays a role in Section 4.7, because there is a convenient equivalence between them when the quantum channels have only full rank output states.

With the knowledge that the contraction coefficients for classical channels are degenerate, Section 4.5 explores the extent of degeneracy in the relative expansion coefficients when either of the compared channels $\mathcal{N}, \mathcal{M}$ are quantum-classical (QC) or classical-quantum (CQ). One important result in this section is Theorem 4.5, which says that (by essentially the same proof), the degeneracy in relative expansion coefficients is recovered when $\mathcal{N}, \mathcal{M}$ are both QC channels (quantum channels with a commutative image space). On the other hand, when $\mathcal{M}$ is QC, Theorem 4.4 gives $\check{\eta}_f^{\text{std}}(\mathcal{N}, \mathcal{M}) \cong_{\alpha,\beta} \check{\eta}_\kappa^{\text{Riem}}(\mathcal{N}, \mathcal{M}),$ fixed $0 < \alpha < \beta$ for all bounded $\kappa, \kappa_f \in \mathcal{K}$ when combined with Theorem 4.7.

Section 4.6 studies the apparent connection between $\kappa_f(0^+) = \infty$ and $\check{\eta}_f^{\text{std}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_{\kappa_f}^{\text{Riem}}(\mathcal{N}, \mathcal{M})$. In Section 4.6.2, it is shown (as a new result) that $\check{\eta}_f^{\text{std}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_{\kappa_f}^{\text{Riem}}(\mathcal{N}, \mathcal{M})$ necessarily holds for all $f \in \mathcal{F}$. Whereas equality, as discussed in Section 4.6.1, is special.

Section 4.6.4 discusses how cases of $\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \cong \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$ are constructed. For example, Theorem 4.9 establishes a simple method to construct cases of equivalence using the convenient properties of $\kappa_{\max}$. However, the relative entropy does not fit as an example of this construction. Theorems 4.10 and 4.11 demonstrate (perhaps unfortunately) uniqueness results on specific relationships between the standard f-divergence and Riemannian semi-norm that gave rise to the two known cases of equality.

Section 4.7 shows that for quantum channels with only full rank output states, there is an equivalence $\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \cong \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}) \cong \check{\eta}_2(\mathcal{N}, \mathcal{M})$. And in Section 4.7.1, a reverse convergence theorem for discrete time-inhomogeneous Markov chains is identified using Riemannian relative expansion coefficients, i.e. giving an exponential lower bound on convergence.

Finally, Section 4.8 modifies the calculations by [6] to demonstrate that the Riemannian relative expansion coefficients are indeed positive in general, for the cases they showed for the BKM metric.

Chapter 2

Preliminaries

2.1 C^*-algebras, Quantum States and Introducing Quantum f-Divergences

This brief technical review of the evolution of quantum states in the Heisenberg and Schrödinger pictures is provided to understand the origin of the standard quantum f-divergence. This will also be an opportunity to explain the notation used throughout the thesis.

We can understand quantum states as a physical system uniquely corresponding to specific (expected) values of observables. To understand this, let us introduce the notion of a C^*-algebra [63]:

Definition 2.1 (C^*-algebra)**.**
A C^*-algebra is a Banach space (i.e. a normed vector space) $\mathcal{A}$ over $\mathbb{C}$ equipped with a map $(\cdot)^* : \mathcal{A} \to \mathcal{A}$, that satisfies $\forall A, B \in \mathcal{A}, \forall \alpha, \beta \in \mathbb{C}$:

1 . $\|AB\| \leqslant \|A\| \, \|B\|$

2 . $(A^*)^* = A$

3 . $(\alpha A + \beta B)^* = \bar{\alpha} A^* + \bar{\beta} B^*$

4 . $(AB)^* = B^* A^*$

5 . $\|A^* A\| = \|A\|^2$

where $\|\cdot\|$ *denotes the norm associated to* $\mathcal{A}$*.*

We are interested in the Banach space $\mathcal{B}(\mathcal{H})$ of bounded linear operators $A : \mathcal{H} \to \mathcal{H}$ for a finite dimensional Hilbert space $\mathcal{H}$. If we select $(\cdot)^* : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ to be the adjoint operator (i.e. the Hermitian conjugate), then $\mathcal{B}(\mathcal{H})$ is certainly a C^*-algebra. The observables (the quantities that we can measure) of our physical system are represented by the self-adjoint operators in $\mathcal{B}(\mathcal{H})$, which we denote $\mathcal{B}_{sa}(\mathcal{H})$. We will also use $\mathcal{B}(\mathcal{H})_+$ to denote positive semi-definite operators in $\mathcal{B}_{sa}(\mathcal{H})$. The possible physical states associated with the system are described as positive unital linear functionals $\omega : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$, i.e. $\omega(I) = 1$ and $\forall A \in \mathcal{B}(\mathcal{H}) : \omega(A^*A) \geqslant 0$. For an observable $\mathcal{O} \in \mathcal{B}_{sa}(\mathcal{H})$, we interpret $\omega(\mathcal{O})$ as an expectation value of a measurement of the observable on the system in state ω [63]. We will not be concerned with C^*-algebras in more generality than this, but perhaps it is worth appreciating that for all of the interesting results about quantum f-divergences that we'll meet (such as monotonicity), we will not require much stronger conditions.

In this thesis, the default inner product for the Hilbert space $\mathcal{B}(\mathcal{H})$ is the Hilbert-Schmidt inner product $\langle f, g\rangle_{HS} := \mathrm{Tr}(f^*g)$, $f, g \in \mathcal{B}(\mathcal{H})$, and thus by the Riesz representation theorem [63] and the positivity of ω, we can obtain for all physical states ω a corresponding density operator ρ_ω:

$$\exists! \rho_\omega \in \mathcal{B}_{sa}(\mathcal{H}), \rho_\omega \text{ p.s.d.} : \omega(A) = \mathrm{Tr}(\rho_\omega A)$$

Note that $\omega(I) = 1 \iff \mathrm{Tr}(\rho_\omega) = 1$. We will denote the set of density operators on a Hilbert space $\mathcal{H}$ as $\mathcal{D}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H})_+ : \mathrm{Tr}(\rho_\omega) = 1\}$ and the set of positive definite density operators as $\mathcal{D}^+(\mathcal{H})$. Then, conversely, $\rho_\omega \in \mathcal{D}(\mathcal{H}) \implies \omega(A) = \mathrm{Tr}(\rho_\omega A)$ is a unital, positive linear functional.

From this, we achieve an understanding that the physical system can equally be characterised by the set of density operators $\mathcal{D}(\mathcal{H})$ rather than the dual space $\mathcal{B}^*(\mathcal{H})$. We may either refer to ω or the associated ρ_ω as the *state*. Further, a quantum channel $\hat{\Phi}$ in the Heisenberg picture is a positive unital linear map that maps between observables, i.e. $\hat{\Phi} : \mathcal{B}(\mathcal{H}') \to \mathcal{B}(\mathcal{H})$, so that $\omega \mapsto \omega \circ \hat{\Phi}$. In the Schrödinger picture, we consider the adjoint $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ of $\hat{\Phi}$, so that $\omega \circ \hat{\Phi}(B) = \mathrm{Tr}(\rho_\omega \hat{\Phi}(B)) = \mathrm{Tr}(\Phi(\rho_\omega)B)$. Thus, the density operator evolves as $\rho_\omega \mapsto \Phi(\rho_\omega)$ and Φ is (as usual) a CPTP map.

Observe that for states $\varphi, \omega \in \mathcal{B}(\mathcal{H})^*$:

$$\varphi(A) = \langle A\phi, \phi\rangle_{HS}, \quad \omega(A) = \langle A\Omega, \Omega\rangle_{HS}$$

where $\phi = \rho_\varphi^{1/2}, \Omega = \rho_\omega^{1/2}$. For the sake of this motivation, we assume that these density operators are in $\mathcal{D}^+(\mathcal{H})$. In [47], it was observed that ω and φ could be related via the unique positive operator $\Delta_{\rho,\gamma} := L_\rho R_\gamma^{-1}$ such that:

$$\langle A\Omega, \Delta_{\rho_\varphi,\rho_\omega}(A\Omega)\rangle_{HS} = \varphi(AA^*) \tag{2.1}$$

We will consistently discuss the left and right multiplication operators $L_A : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}), X \mapsto AX$, $R_A : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}), X \mapsto XA$ for positive semi-definite $A \in \mathcal{B}(\mathcal{H})$. Note that A^{-1} denotes the generalised inverse, where only the non-zero eigenvalues are inverted, and L_A, R_A are positive operators, i.e. $\langle X, L_A(X)\rangle_{HS}, \langle X, R_A(X)\rangle_{HS} \geqslant 0$.

$\Delta_{\rho,\gamma}$ is called the *relative modular operator*, and we can define the action of continuous functions $f : (0,\infty) \to \mathbb{R}$ on this operator from the spectral decompositions $\rho = \sum_{a\in\operatorname{spec}(\rho)} aP_a$, $\gamma = \sum_{b\in\operatorname{spec}(\gamma)} bQ_b$, as [29]:

$$f(\Delta_{\rho,\gamma}) = \sum_{\substack{a\in\operatorname{spec}\rho,\\ b\in\operatorname{spec}\gamma\setminus\{0\}}} bf\left(\frac{a}{b}\right) L_{P_a} R_{Q_b} \tag{2.2}$$

If f is positive, $f(\Delta_{\rho,\gamma})$ is also a positive operator.

The intuition left by (2.1) is that if we replaced $\Delta_{\rho_\varphi,\rho_\omega}$ with $f(\Delta_{\rho_\varphi,\rho_\omega})$, for some $f(x) \neq 1, x$, the quantity should contain some mixture of the information between ρ_φ and ρ_ω. [47] proved the monotonicity (or data processing inequality) of the *quasi relative entropy* $D_f^A(\rho\|\gamma) := \langle A\gamma^{1/2}, f(\Delta_{\rho,\gamma})(A\gamma^{1/2})\rangle_{HS}$, i.e. $D_f^{\Phi(A)}(\Phi(\rho)\|\Phi(\gamma)) \leqslant D_f^A(\rho\|\gamma)$ for all CPTP Φ, for operator convex f. This establishes the quasi relative entropy as a geniune measure of distinguishability between density operators ρ, γ. Later, the *standard quantum f-divergence* was defined for the special case that $A = I$: $D_f^{\text{std}}(\rho\|\gamma) := D_f^I(\rho\|\gamma)$, and generalised to non-invertible $\rho, \gamma \in \mathcal{B}_{\text{sa}}(\mathcal{H})$ by effectively imposing a continuity condition [29, Proposition 2.2, Corollary 2.3]:

$$\begin{aligned} D_f^{\text{std}}(\rho\|\gamma) &= \lim_{\varepsilon\to 0}\langle(\gamma+\varepsilon I)^{1/2}, f(\Delta_{\rho+\varepsilon I,\gamma+\varepsilon I})((\gamma+\varepsilon I)^{1/2})\rangle_{HS} && (2.3)\\ &= \langle\gamma^{1/2}, f(\Delta_{\rho,\gamma})(\gamma^{1/2})\rangle_{HS} + f'(\infty)\operatorname{Tr}\rho(1-\gamma^0) && (2.4)\\ &= \sum_{\substack{a\in\operatorname{spec}\rho,\\ b\in\operatorname{spec}\gamma\setminus\{0\}}} bf\left(\frac{a}{b}\right)\operatorname{Tr}P_aQ_b + af'(\infty)\operatorname{Tr}P_aQ_0 && (2.5)\end{aligned}$$

where $f(0) := f(0^+), f'(\infty) := \lim_{x\to\infty}\frac{f(x)}{x}$, and γ^0 denotes the orthogonal projector onto $\operatorname{supp}\gamma$. From the expression (2.5), we can see that there is a one-to-one correspondence between D_f^{std} and f, because $f(x) = \frac{D(xI\|I)}{\dim \mathcal{H}}, x \in (0, +\infty)$ [29, Proposition 2.8]. Also, we have

$$D_f^{\text{std}}(\rho\|\gamma) \equiv D_{\tilde{f}}^{\text{std}}(\gamma\|\rho),$$

for a function $\tilde{f}(x) := xf(x^{-1})$, $x \in (0, \infty)$; $\tilde{f}$ will be called the *transpose* of f.

In the case $f(x) = x\log x$, this defines the (Umegaki) relative entropy $D(\rho\|\gamma) = D_{x\log x}^{\text{std}}(\rho\|\gamma)$ from (1.1). The other quantum f-divergences can be viewed as building on the standard quantum f-divergence. The quantum f-divergences share a lot of convenient properties with the relative entropy, as we will see in later sections. It is because of this that we can learn a lot about the important traits that make the relative entropy interesting through studying other quantum f-divergences; some of the analysis around the titular reverse data processing inequalities will be based on this principle.

2.2 Operator Convex Functions and Monotone Riemannian Metrics

The purpose of this thesis is to analyse the amount of distinguishability that is preserved after a channel has acted on a set of states. While we may prefer to focus on the standard quantum f-divergence, this can be difficult in practice. To overcome the computational difficulties of dealing with the relative entropy, [6] reduced the problem to working with the corresponding *Riemannian metric* instead. Later on, we attempt to understand and extend this reduction. The main objective of this section is to review some essential details from [37, 52, 49] that will assist in developing an appreciation for the Riemannian metric.

While the connection between the standard f-divergence and Riemannian metric is a theme in this thesis, the Riemannian metric should also be appreciated as a standalone measure of distinguishability between quantum states. In the classical setting of comparing two probability distributions, the Fisher information was developed as the unique metric on the probability simplex (i.e. commuting quantum states) that is monotone under the action of classical channels (i.e. stochastic mappings) [43]. The Riemannian metrics defined on density operators recover a notion of Fisher information.

A *metric* is a complex inner product $M_\rho(\cdot,\cdot) : \mathcal{B}(\mathcal{H}) \times \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ parametrised by positive definite $\rho \in \mathcal{D}^+(\mathcal{H}) \subseteq \mathcal{D}(\mathcal{H})$ for which the following conditions hold [49]:

1. $(A, B) \mapsto M_\rho(A, B)$ is sesquilinear for all $\rho \in \mathcal{D}^+(\mathcal{H})$.
2. $M_\rho(A, A) \geqslant 0$ for all $A \in \mathcal{B}(\mathcal{H})$, $\rho \in \mathcal{D}^+(\mathcal{H})$ with equality iff $A = 0$.

3. $\rho \mapsto M_\rho(A, A)$ is continuous for every $A \in \mathcal{B}(\mathcal{H})$.

A *monotone metric* satisfies the extra condition of monotonicity under CPTP maps:

4. $M_{\Phi(\rho)}(\Phi(A), \Phi(A)) \leqslant M_\rho(A, A)$ for every $A \in \mathcal{B}(\mathcal{H}), \rho \in \mathcal{D}^+(\mathcal{H})$, CPTP Φ

While it may seem odd to parametrise an inner product by a density operator, this is only natural to ensure unitary covariance is possible, which is necessary for monotonicity: $M_{U\rho U*}(UAU^*, UAU^*) = M_\rho(A, A)$ for every $A \in \mathcal{B}(\mathcal{H}), \rho \in \mathcal{D}^+(\mathcal{H}), U \in U(d)$. Requiring that the inner product is unitarily invariant, on the other hand, uniquely determines the Hilbert-Schmidt inner product up to a constant multiple: $\langle A, B\rangle_{HS} = \langle UAU^*, UBU^*\rangle_{HS}$ for every $A \in \mathcal{B}(\mathcal{H}), \rho \in \mathcal{D}^+(\mathcal{H}), U \in U(d)$. In fact, the Hilbert-Schmidt inner product is the special case when $\rho = \frac{I}{d}$ (up to a constant factor).

As opposed to the classical setting, there are now multiple Riemannian metrics. In fact, there is an important characterisation of the Riemannian metric in terms of operator monotone decreasing functions [49, 37][24, Theorem 2.4]:

Theorem 2.1. *[49, Theorem 7] There is a one-to-one correspondence between symmetric Riemannian metrics (i.e. $M_\rho(A, B) = M_\rho(B^*, A^*)$) and operator monotone functions $\kappa : (0, \infty) \mapsto (0, \infty)$ satisfying $\kappa(x^{-1}) = x\kappa(x)$.*

We will henceforth denote the symmetric Riemannian metrics by $M^\kappa_\rho(A, B)$ and their corresponding norm as $\|A\|_{\kappa,\rho} := \sqrt{M^\kappa_\rho(A, A)}$. By the polarisation identity, $(A, B) \mapsto M^\kappa_\rho(A, B)$ is uniquely determined by $A \mapsto \|A\|_{\kappa,\rho}$. Usually, we will take $A = \rho - \gamma$, for a density operator γ. As in the case of quantum f-divergences, we want to speak generally about $\rho, \gamma \in \mathcal{D}(\mathcal{H})$, so we do this at the expense of losing positive definiteness and continuity w.r.t. ρ. Specifically, $\|A\|_{\kappa,\rho}$ is a norm on $A \in \operatorname{supp} \rho$, but $\|A\|_{\kappa,\rho} = 0$ for $A \in (\operatorname{supp} \rho)^\perp$, hence it's only a semi-norm in general. We define $T_\rho\mathcal{D}(\mathcal{H}) := \{A \in \mathcal{B}(\mathcal{H}) : \operatorname{Tr} A = 0, A = A^*, \operatorname{supp} A \leqslant \operatorname{supp} \rho\}$ as an appropriate set of *tangent vectors* at ρ.

You have now heard that standard quantum f-divergences are parametrised by operator convex functions and the symmetric Riemannian metrics are parametrised by operator monotone functions. Let's recall what this means:

1. An *operator convex function* is a real function $f : (0, \infty) \to \mathbb{R}$ that satisfies:

$$f(\lambda A + (1 - \lambda)B) \leqslant \lambda f(A) + (1 - \lambda)f(B) \ \forall\lambda \in (0, 1), \forall A, B \in \mathbb{P}_d, \forall d \in \mathbb{N}$$

2. An *operator monotone function* is a positive function $f : (0,\infty) \to (0,\infty)$ that satisfies:

$$f(A) \leqslant f(B)\ \forall A, B \in \mathbb{P}_d, A \leqslant B, \forall d \in \mathbb{N}$$

An *operator monotone decreasing function* κ is a function $\kappa(x) = 1/f(x)$ where f is an operator monotone function.

$\mathbb{P}_d$ denotes the positive semi-definite operators in $\mathcal{B}(\mathbb{C}^d)$. In particular, A, B can be super-operators, such as $L_C, R_C, C \in \mathcal{B}(\mathcal{H})$.

We will be focused on the following classes of functions:

$$\begin{aligned}
&\mathcal{F} := \{f : (0,\infty) \to \mathbb{R},\ \text{operator convex},\ f(1) = 0, f''(1) > 0\},\\
&\mathcal{F}_{\text{sym}} := \{f : (0,\infty) \to \mathbb{R},\ \text{operator convex},\ f(x) = xf(x^{-1}) \text{ for } x > 0,\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad f(1) = f'(1) = 0, f''(1) = 2\},\\
&\mathcal{K} := \{\kappa : (0,\infty) \to \mathbb{R},\ \text{operator convex},\ x\kappa(x) = \kappa(x^{-1}) \text{ for } x > 0, \kappa(1) = 1\}
\end{aligned}$$

From [24], we realise that integral representations for these classes of functions can be obtained by considering that the divided difference functions of operator convex functions are operator monotone, and operator monotone functions have a Löwner integral representation given by a unique finite measure [41]. These integral representations are as follows:

Theorem 2.2 (Integral Representations)**.** *[28, 24]*

(i) If $f : (0,\infty) \to \mathbb{R}$ is an operator convex function, then there exists a unique constant $c \geqslant 0$ and a unique positive measure μ on $[0,\infty)$ with $\int_{[0,\infty)} (1+s)^{-1}\, d\mu(s) < +\infty$ such that

$$f(x) = f(1) + f'(1)(x-1) + c(x-1)^2 + \int_{[0,\infty)} \frac{(x-1)^2}{x+s}\, d\mu(s), \quad x \in (0,\infty). \tag{2.6}$$

(ii) If $\kappa : (0,\infty) \to \mathbb{R}$ is an operator convex function and it satisfies the normalization $\kappa(1) = 1$ and the symmetry condition $x\kappa(x) = \kappa(x^{-1})$ for all $x > 0$, then $\kappa(x) > 0$ for all $x > 0$ (hence $\kappa \in \mathcal{K}$), and there exists a unique probability measure m on $[0,1]$ such that

$$\kappa(x) = \int_{[0,1]} \frac{1+x}{(x+s)(1+sx)} \cdot \frac{(1+s)^2}{2}\, dm(s) \tag{2.7}$$

$$= \int_{[0,1]} \left(\frac{1}{x+s} + \frac{1}{sx+1}\right) \cdot \frac{(1+s)}{2}\, dm(s), \quad x \in (0,\infty). \tag{2.8}$$

These integral representations are incredibly important as they provide a way to decompose problems about quantum f-divergences and symmetrised Riemannian metrics, as we can often work with the individual terms or cases of the integrand separately. Since the functions $\kappa_s(x) = \frac{1+x}{(x+s)(1+sx)} \cdot \frac{(1+s)^2}{2} \in \mathcal{K}$ for $s \in [0,1]$ (take $m(s) = 1$), we can consider $\mathcal{K}$ as a *Choquet simplex* with $\kappa_s, s \in [0,1]$ as the extreme points.

In particular, notice that operator convex functions f are necessarily convex and infinitely differentiable, i.e. $f \in C^\infty((0,\infty))$. This is quite strong if we consider that for classical f-divergences, analogous (and, in fact, stronger) results were obtained assuming only strong convexity at 1 and continuous twice differentiability in the neighborhood of 1 (and bounded lim sup).

Indeed, making the same assumptions, it is possible to obtain that the f-divergences are locally symmetric at second order (see Appendix A). However, in all of our cases of interest, to see that they are exactly second order and that there is a connection between f-divergences and symmetric Riemannian metrics, we rely on a result by [37] that is based on the integral representation of operator convex functions. We will review this result now.

First, we note that there is a natural correspondence between $\mathcal{K}$ and $\mathcal{F}_{sym}$ via their integral representations:

Proposition 2.1. *[28, Proposition 2.2]*
For a function $\kappa : (0,\infty) \to (0,\infty)$ consider the following conditions:

(a) κ is operator convex,

(b) κ is operator monotone decreasing,

(c) $f(x) \equiv (x-1)^2\kappa(x)$ is operator convex.

Then (a) $\Leftarrow$ (b) $\Leftrightarrow$ (c). Moreover, if $x\kappa(x) = \kappa(x^{-1})$ for all $x > 0$ or equivalently $f(x) = xf(x^{-1})$ for all $x > 0$, then the above conditions (a)–(c) are all equivalent.

i.e. there is a correspondence $\kappa \in \mathcal{K} \leftrightarrow f \in \mathcal{F}_{\text{sym}}$ via $g(x) = (x-1)^2\kappa(x)$. [37, 52] demonstrated that the symmetrised Riemannian metric could be understood as the Fisher information w.r.t. a corresponding f-divergence, i.e.

Theorem 2.3 (Fisher Information)**.** *[37, Theorem 2.8]*
For each $f \in \mathcal{F}_{\text{sym}}$ and density matrix $\rho \in \mathcal{D}^+(\mathcal{H})$, we can associate $\kappa(x) = \frac{f + xf(x^{-1})}{f''(1)(x-1)^2} \in \mathcal{K}$ and this satisfies

$$M_\rho^\kappa(A,B) = -\frac{\partial^2}{\partial\alpha\partial\beta} D_f^{\text{std}}(\rho + \alpha A \| \rho + \beta B)\bigg|_{\alpha=\beta=0} = f''(1)\langle A, \Omega_\rho^\kappa(B)\rangle_{HS} = f''(1)\,\mathrm{Tr}\, A\Omega_\rho^\kappa(B)$$

To derive this, they considered another fundamental expression for the standard f-divergence, which we will now show:

Theorem 2.4 (Integral representation of standard f-divergence)**.** *[37, Theorem 2.5] Suppose we are given an operator convex function $f \in \mathcal{F}$, then for all $\rho, \gamma \in \mathcal{D}^+(\mathcal{H})$:*

$$D_f^{\text{std}}(\rho\|\gamma) = \operatorname{Tr}(\rho-\gamma)\left[c\rho^{-1} + \mu(0)\gamma^{-1}\right](\rho-\gamma) \tag{2.9}$$

$$+ \int_0^\infty \operatorname{Tr}\left((\rho-\gamma)\frac{1}{L_\rho + sR_\gamma}(\rho-\gamma)\right) d\mu(s)$$

$$= \operatorname{Tr}\left[(\rho-\gamma)R_\gamma^{-1}\nu(\Delta_{\rho,\gamma})(\rho-\gamma)\right] \tag{2.10}$$

where c, $\mu(0)$ and μ are as in 2.6, and $\nu(x) := \frac{f(x)-f'(1)(x-1)}{(x-1)^2}$.

Proof.
By 2.6, we can write $f(x),\ x \in (0,\infty)$ as:

$$f(x) = f'(1)(x-1) + c(x-1)^2 + \mu(0)\frac{(x-1)^2}{x} + \int_{(0,\infty)} \frac{(x-1)^2}{x+s}\, d\mu(s)$$

Notice that the definition 2.3 of the f-divergence is linear in f, so we break down any f-divergence for $f \in \mathcal{F}$:

$$D_f^{\text{std}}(\rho\|\gamma) = f'(1)D_{(x-1)}^{\text{std}}(\rho\|\gamma) + cD_{(x-1)^2}^{\text{std}}(\rho\|\gamma) + \mu(0)D_{\frac{(x-1)^2}{x}}^{\text{std}}(\rho\|\gamma) + \int_{(0,\infty)} D_{\frac{(x-1)^2}{x+s}}^{\text{std}}(\rho\|\gamma)\, d\mu(s)$$

However, the first term doesn't actually count:

$$D_{(x-1)}^{\text{std}}(\rho\|\gamma) = \operatorname{Tr}\ \gamma^{1/2}(L_\rho R_\gamma^{-1} - 1)\gamma^{1/2} = \operatorname{Tr}\ (\rho-\gamma) = 0$$

And the second term evaluates to:

$$D_{(x-1)^2}^{\text{std}}(\rho\|\gamma) = \operatorname{Tr}\ \gamma^{1/2}(L_\rho R_\gamma^{-1} - 1)^2\gamma^{1/2} = \operatorname{Tr}\ (\rho^2\gamma^{-1} - 2\rho + \gamma) = \operatorname{Tr}\ (\rho-\gamma)^2\gamma^{-1}$$

We next observe that since $\frac{(x-1)^2}{x} = x \cdot (x^{-1}-1)^2$ is the transpose of $(x-1)^2$:

$$D_{\frac{(x-1)^2}{x}}^{\text{std}}(\rho\|\gamma) = D_{(x-1)^2}^{\text{std}}(\gamma\|\rho) = \operatorname{Tr}\ (\rho-\gamma)^2\rho^{-1}$$

And finally:

$$
\begin{aligned}
D^{\text{std}}_{\frac{(x-1)^2}{x+s}}(\rho\|\gamma) &= \langle\, \gamma^{1/2}, (L_\rho R_\gamma^{-1} - 1)^2 (L_\rho R_\gamma^{-1} + s)^{-1} \gamma^{1/2} \rangle_{HS} \\
&= \langle\, (L_\rho R_\gamma^{-1} - 1)\gamma^{1/2}, (L_\rho R_\gamma^{-1} + s)^{-1} (L_\rho R_\gamma^{-1} - 1) \gamma^{1/2} \rangle_{HS} \\
&= \langle\, (\rho - \gamma)\gamma^{-1/2}, (L_\rho R_\gamma^{-1} + s)^{-1} (\rho - \gamma) \gamma^{-1/2} \rangle_{HS} \\
&= \langle\, \rho - \gamma, (L_\rho R_\gamma^{-1} + s)^{-1} R_\gamma^{-1} (\rho - \gamma) \rangle_{HS} \\
&= \langle\, \rho - \gamma, (L_\rho + s R_\gamma)^{-1} (\rho - \gamma) \rangle_{HS}
\end{aligned}
$$

Putting all of these results together gives the result. □

Proof of Theorem 2.3.
By Theorem 2.4, we can write:

$$
\begin{aligned}
D^{\text{std}}_f(\rho + \alpha A \| \rho + \beta B) =& \langle \alpha A - \beta B, R^{-1}_{\rho+\alpha A} \nu(\Delta_{\rho+\alpha A, \rho+\beta B})(\alpha A - \beta B) \rangle_{HS} \\
=& \alpha^2 \langle A, R^{-1}_{\rho+\alpha A} \nu(\Delta_{\rho+\alpha A, \rho+\beta B}) A \rangle_{HS} - \alpha\beta (\langle A, R^{-1}_{\rho+\alpha A} \nu(\Delta_{\rho+\alpha A, \rho+\beta B}) B \rangle_{HS} \\
&+ \langle B, R^{-1}_{\rho+\alpha A} \nu(\Delta_{\rho+\alpha A, \rho+\beta B}) A \rangle_{HS}) + \beta^2 \langle B, R^{-1}_{\rho+\alpha A} \nu(\Delta_{\rho+\alpha A, \rho+\beta B}) B \rangle_{HS}
\end{aligned}
$$

Evaluating gives the result:

$$
\begin{aligned}
\frac{\partial^2}{\partial\alpha\,\partial\beta} D^{\text{std}}_f(\rho + \alpha A \,\|\, \rho + \beta B)\Big|_{\alpha=\beta=0} &= \big[\langle A, R^{-1}_{\rho+\alpha A} \nu\left(\Delta_{\rho+\alpha A, \rho+\beta B}\right) B \rangle_{\text{HS}} \\
&\quad + \langle B, R^{-1}_{\rho+\alpha A} \nu\left(\Delta_{\rho+\alpha A, \rho+\beta B}\right) A \rangle_{\text{HS}}\big]_{\alpha=\beta=0} \\
&= \langle A, R^{-1}_\rho \nu\left(\Delta_{\rho,\rho}\right) B \rangle_{\text{HS}} + \langle B, R^{-1}_\rho \nu\left(\Delta_{\rho,\rho}\right) A \rangle_{\text{HS}} \\
&= \langle A, R^{-1}_\rho (\nu\left(\Delta_{\rho,\rho}\right) + \widetilde{\nu}\left(\Delta_{\rho,\rho}\right)) A \rangle_{\text{HS}} \\
&= f''(1) \langle A, \Omega^\kappa_\rho(B) \rangle_{\text{HS}}
\end{aligned}
$$

where $\widetilde{\nu}(x) = \frac{\widetilde{f}(x) - \widetilde{f}'(1)(x-1)}{(x-1)^2} = \frac{\widetilde{f}(x) + f'(1)(x-1)}{(x-1)^2}$ since $\widetilde{f}'(1) = -f'(1)$, so $\nu + \widetilde{\nu} = f''(1)\kappa$. □

The following result demonstrates a similar connection between the local behaviour of standard f-divergences and symmetric Riemannian metrics. This should be one of the main takeaways from this section, and we will see an important application in Section 4.6.

Lemma 2.1 (Standard Quantum f-divergences are Locally Riemannian)**.**
Let $f \in \mathcal{F}$, then for any density operator ρ and traceless Hermitian operator X, with $\operatorname{supp} X \leqslant \operatorname{supp} \rho$*, acting on the finite d-dimensional Hilbert space $\mathcal{H}$, we define $\rho_\varepsilon := \rho + \varepsilon X \geqslant 0$ for $\varepsilon \in \mathbb{R}$ sufficiently small. Then*

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} D_f^{\mathrm{std}}(\rho_\varepsilon || \rho) \equiv \frac{1}{2} \frac{d^2}{d\varepsilon^2}\Big|_{\varepsilon=0} D_f^{\mathrm{std}}(\rho_\varepsilon || \rho) = \frac{f''(1)}{2} ||X||^2_{\kappa_f,\, \rho} \tag{2.11}$$

where $\kappa_f(x) := \frac{f(x)+\tilde{f}(x)}{f''(1)(x-1)^2}$ for $x \neq 1$.

Proof. We can find an open neighbourhood $I \subseteq \mathbb{R}$ of 0 such that $\rho_\varepsilon \geqslant 0$, $\operatorname{supp} \rho_\varepsilon = \operatorname{supp} \rho \; \forall \varepsilon \in I$.

By Theorem 2.4, we can write:

$$\begin{aligned}
D_f^{\mathrm{std}}(\rho_\varepsilon \| \rho) &= \varepsilon^2 \left\langle X, R_{\rho_\varepsilon}^{-1} \nu(\Delta_{\rho,\rho_\varepsilon})(X) \right\rangle \\
\Longrightarrow \lim_{\varepsilon \to 0} \frac{1}{\varepsilon^2} D_f^{\mathrm{std}}(\rho_\varepsilon \| \rho) &= \left\langle X, R_\rho^{-1} \nu(\Delta_{\rho,\rho})(X) \right\rangle \\
&= \left\langle X, R_\rho^{-1} \left(\frac{\nu + \tilde{\nu}}{2} \right) (\Delta_{\rho,\rho})(X) \right\rangle_{HS} \\
&= \frac{f''(1)}{2} \left\langle X, R_\rho^{-1} \kappa_f(\Delta_{\rho,\rho})(X) \right\rangle_{HS} \\
&\equiv \frac{f''(1)}{2} ||X||^2_{\kappa_f,\, \rho}
\end{aligned}$$

□

2.3 Examples of Divergences and Riemannian Metrics

We will consider five different examples, as in [28].

Example 2.1. *The function* $f(x) = x \log x \in \mathcal{F}$ *gives us the relative entropy:*

$$D(\rho\|\gamma) := D^{\mathrm{std}}_{x\log x}(\rho\|\gamma) = \mathrm{Tr}(\rho \log \rho - \rho \log \gamma), \tag{2.12}$$

and $\kappa_{\mathrm{BKM}}(y) := \kappa_{x \log x}(y) = \frac{\log y}{y-1} \in \mathcal{K}$ *gives*

$$\Omega^{\kappa_{\mathrm{BKM}}}_{\rho}(A) = \int_0^\infty (\rho + rI)^{-1} A (\rho + rI)^{-1} dr. \tag{2.13}$$

The corresponding monotone metric is called the BKM metric. The symmetrisation of the relative entropy corresponding to $f(x) = (x-1)\log x \in \mathcal{F}_{\mathrm{sym}}$ *satisfies* $f(0^+) = f'(\infty) = \infty$, *and thus* $D^{\mathrm{std}}_{(x-1)\log x}(\rho\|\gamma)$ *is not strictly continuous in the first or second argument.*

Example 2.2. *The function* $f(x) = (x-1)^2 \in \mathcal{F}$ *gives us the quadratic relative entropy or* χ^2*-divergence:*

$$D_{\max}(\rho\|\gamma) := D^{\mathrm{std}}_{(x-1)^2}(\rho\|\gamma) = \mathrm{Tr}(\rho-\gamma)^2\gamma^{-1} = \mathrm{Tr}\big(\rho^2\gamma^{-1}\big) - 1, \tag{2.14}$$

and $\kappa_{\max}(y) := \kappa_{(x-1)^2}(y) = \frac{y+1}{2y} \in \mathcal{K}$ *gives*

$$\Omega^{\kappa_{\max}}_{\rho}(A) = \frac{\rho^{-1}A + A\rho^{-1}}{2}. \tag{2.15}$$

The corresponding monotone metric is called the maximal metric. Just like the relative entropy, the symmetrisation $f(x) = \frac{(x-1)^2(x+1)}{2x}$ *satisfies* $f(0^+) = f'(\infty) = \infty$.

An interesting and useful property is that the maximal divergence and the Riemannian semi-norm coincide:

$$D_{\max}(\rho\|\gamma) = \|\rho - \gamma\|^2_{\kappa_{\max},\gamma} \tag{2.16}$$

It's because of this, and $\kappa(x) \leqslant \kappa_{\max}(x)\ \forall x \in (0,\infty)$ *for any* $\kappa(x) \in \mathcal{K}$, *that the maximal metric becomes a primary focus of this thesis.*

Example 2.3. *The function* $f(x) = (x-1)^2 x^{-1/2} \in \mathcal{F}_{\mathrm{sym}}$ *gives us the central power divergence:*

$$D_{(x-1)^2x^{-1/2}}(\rho\|\gamma) = \mathrm{Tr}\big((\rho-\gamma)\rho^{-1/2}(\rho-\gamma)\gamma^{-1/2}\big) \tag{2.17}$$

and $\kappa_c(y) := \kappa_{(x-1)^2 x^{-1/2}}(y) = y^{-1/2}$ *gives*

$$\Omega_\rho^{\kappa_c}(A) = \rho^{-1/2} A \rho^{-1/2}. \tag{2.18}$$

The corresponding monotone metric is the central power metric. Interestingly, this is the only choice of $\kappa \in \mathcal{K}$ *such that both* $\Omega_\rho^{\kappa_c}$ *and* $(\Omega_\rho^{\kappa_c})^{-1}$ *are CP for all* $\rho \in \mathbb{P}_d$.

Example 2.4. *The functions* $f_t(x) = \frac{x - x^t}{t(1-t)} \in \mathcal{F}$ *for any* $t \in (0,1) \cup (1,2]$ *give us the Wigner-Yanase-Dyson (WYD) divergences. This time,* $\kappa_t^{\mathrm{WYD}}(x) \equiv \frac{1}{t(1-t)} \cdot \frac{(1-x^t)(1-x^{1-t})}{(1-x)^2} \in \mathcal{K}$ *is bounded (i.e.* $\kappa_t(0^+) < \infty$*), so any corresponding divergences are bounded. When* $t = 1/2$*, the Wigner-Yanase divergence (WY) is given by*

$$D^{\mathrm{std}}_{4(x-\sqrt{x})}(\rho\|\gamma) = 4(1 - \mathrm{Tr}\big(\rho^{1/2}\gamma^{1/2}\big)) \tag{2.19}$$

and $\kappa_{\mathrm{WY}}(y) := \kappa_{1/2}^{\mathrm{WYD}}(y) = 4\frac{(1-\sqrt{y})^2}{(1-y)^2} = 4(1+\sqrt{y})^{-2}$ *gives*

$$\Omega_\rho^{\kappa_{1/2}^{\mathrm{WYD}}}(A) = 4(\sqrt{L_\rho} + \sqrt{R_\rho})^{-2}(A). \tag{2.20}$$

The corresponding monotone metric is the WY metric.
Note: $\kappa_t^{\mathrm{WYD}} = \kappa_{1-t}^{\mathrm{WYD}}$ *and* $\lim_{t\to 0^+} \kappa_t^{\mathrm{WYD}} = \kappa^{\mathrm{BKM}}$*. As a result,* $\lim_{t\to 0^+} D^{\mathrm{std}}_{f_t}(\rho\|\gamma) = D^{\mathrm{std}}_{(x-1)\log x}(\rho\|\gamma)$.

Example 2.5. *The function* $f(x) = \frac{2(x-1)^2}{x+1} \in \mathcal{F}$ *gives us the minimal divergence*

$$D_{\frac{2(x-1)^2}{x+1}}(\rho\|\gamma) = \langle \rho - \gamma, \frac{2}{L_\rho + R_\gamma}(\rho - \gamma)\rangle_{HS}, \tag{2.21}$$

and $\kappa_{\min}(y) := \kappa_f(y) = \frac{2}{1+y}$ *gives (the symmetric logarithmic derivative)*

$$\Omega_\rho^{\kappa_f}(A) = \frac{2}{L_\rho + R_\gamma}(A) = 2\int_0^\infty \exp(-t\rho) A \exp(-t\rho) dt. \tag{2.22}$$

The corresponding monotone metric is the minimal metric. $\kappa(x) \geqslant \kappa_{\min}(x)\, \forall x \in (0,\infty)$ *for any* $\kappa(x) \in \mathcal{K}$ *(so clearly, it's bounded), which can sometimes be useful, since it can act as an 'edge case' like* $\kappa_{\max}(x)$*. As it turns out,* $\kappa_{\min}(x)$ *is sometimes useful as a counterexample (see Theorem 4.15).*

2.4 Qubits: Obtaining Explicit Expressions for the Riemannian Metrics

When dealing with qubit states and CPTP maps, it's sometimes practical to use the Bloch representation to compute the standard f-divergences and their Riemannian semi-norm. The latter is what we'll work with, and this becomes essential to demonstrate cases where a channel can preserve distinguishability.

Recall that the identity and Pauli matrices

$$\mathbb{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

together form an orthogonal basis for $\mathcal{B}(\mathbb{C}^2)$ w.r.t. the Hilbert-Schmidt inner product. Any traceless Hermitian operator X and density operator ρ can be represented by real vectors:

$$\begin{aligned} X &= \mathbf{y} \cdot \sigma = y_1\sigma_x + y_2\sigma_y + y_3\sigma_z, \quad \mathbf{y} \in \mathbb{R}^3 \\ \rho &= \frac{1}{2}(\mathbb{I}_2 + \mathbf{w} \cdot \sigma) = \frac{1}{2}(\mathbb{I}_2 + w_1\sigma_x + w_2\sigma_y + w_3\sigma_z), \quad \mathbf{w} \in \mathbb{R}^3. \end{aligned} \tag{2.23}$$

where $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ denotes the vector of Pauli matrices. Note that the eigenvalues of ρ are $\frac{1}{2}(1 \pm |\mathbf{w}|)$, so it is a density operator if and only if $|\mathbf{w}| \leqslant 1$. i.e. the set of density operators can be identified with the unit ball/'Bloch sphere' in $\mathbb{R}^3$, where the pure states lie on the surface. This Bloch representation has already been used to study the contraction coefficient for many qubit channels [28], and the relative entropy/BKM relative expansion coefficients [6] (see Section 4.8 for the generalisations). The following basic properties are useful, see [28, Appendix B]:

$$\begin{aligned} &\text{Product rule: } (a\mathbb{I}_2 + \mathbf{w} \cdot \sigma)(b\mathbb{I}_2 + \mathbf{y} \cdot \sigma) = (ab + \mathbf{w} \cdot \mathbf{y})\mathbb{I}_2 + (a\mathbf{y} + b\mathbf{w} + i\mathbf{w} \times \mathbf{y}) \cdot \sigma, \\ &\text{Inverse rule: } (a\mathbb{I}_2 + \mathbf{w} \cdot \sigma)^{-1} = \frac{a\mathbb{I}_2 - \mathbf{w} \cdot \sigma}{a^2 - |\mathbf{w}|^2}, \\ &\text{Square root rule: } (c\mathbb{I}_2 + \mathbf{w} \cdot \sigma)^{\frac{1}{2}} = \sqrt{\frac{\zeta(c, \mathbf{w})}{2}} \left(\mathbb{I}_2 + \frac{\mathbf{w} \cdot \sigma}{\zeta(c, \mathbf{w})} \right) \end{aligned} \tag{2.24}$$

where $\mathbf{w} \times \mathbf{y}$ is the cross product of two vectors and $\zeta(c, \mathbf{w}) := c + \sqrt{c^2 - |\mathbf{w}|^2}$

We will now see how to express the Riemannian norms in terms of Bloch vectors. We will consider the following definitions throughout:

$$\rho := \frac{1}{2}(\mathbb{I}_2 + \mathbf{w} \cdot \sigma), \ \gamma := \frac{1}{2}(\mathbb{I}_2 + \mathbf{u} \cdot \sigma), \ X := \mathbf{y} \cdot \sigma, \ \mathbf{y} := \mathbf{w} - \mathbf{u}, \ \mathbf{w}, \mathbf{u} \in \mathbb{R}^3 \tag{2.25}$$

Explicit calculation for the Extreme Points

Lemma 2.2. *[28] For the traceless Hermitian operator X and density operator $\rho =: \frac{P}{2}$ given by (2.25) and considering the extreme points $\kappa_s(x) = \frac{1+s}{2}(\frac{1}{x+s} + \frac{1}{1+sx}) \in \mathcal{K}$:*

$$\|X\|^2_{\kappa_s,\,\rho} = 2\langle X, \Omega_P^{\kappa_s}(X)\rangle = \frac{4(1+s)^2}{\xi_s(|\mathbf{w}|^2)}\left[|\mathbf{y}|^2 + \frac{4s(\mathbf{w}\cdot\mathbf{y})^2}{(1+s)^2(1-|\mathbf{w}|^2)}\right],$$

where

$$\xi_s(x) := (1+s)^2 - (1-s)^2 x = (1+s)^2(1-x) + 4sx.$$

Proof.
Observe that

$$\langle X, \Omega_P^{\kappa_s}(X)\rangle = \left\langle X, \frac{1+s}{L_P + sR_P}(X)\right\rangle$$

Write $c\mathbb{I}_2 + \mathbf{z}\cdot\sigma = (L_P + sR_P)^{-1}(\mathbf{y}\cdot\sigma)$, then

$$\begin{aligned}\mathbf{y}\cdot\sigma &= (\mathbb{I}_2 + \mathbf{w}\cdot\sigma)(c\mathbb{I}_2 + \mathbf{z}\cdot\sigma) + s(c\mathbb{I}_2 + \mathbf{z}\cdot\sigma)(\mathbb{I}_2 + \mathbf{w}\cdot\sigma)\\ &= (1+s)\,(c + \mathbf{w}\cdot\mathbf{z})\,\mathbb{I}_2 + \left[(1+s)\,(\mathbf{z} + c\mathbf{w}) + i(1-s)\mathbf{w}\times\mathbf{z}\right]\cdot\sigma\end{aligned}$$

$$\begin{aligned}\therefore\ c &= -\mathbf{w}\cdot\mathbf{z},\\ \mathbf{y} &= (1+s)\,(\mathbf{z} + c\mathbf{w}) + i(1-s)\mathbf{w}\times\mathbf{z} = (1+s)\left((\mathbb{I}_2 - \mathbf{w}\mathbf{w}^\top)\mathbf{z}\right) + i(1-s)\mathbf{w}\times\mathbf{z}\end{aligned}$$

Consider $\mathbf{z} = \mathbf{z}_1 + i\mathbf{z}_2$, then comparing real and imaginary parts yields:

$$\begin{aligned}0 &= (1+s)\left[(\mathbb{I}_2 - \mathbf{w}\mathbf{w}^T)\mathbf{z}_2 - \mathbf{z}_1\times\mathbf{w}\right]\\ \mathbf{y} &= (1+s)\left[(\mathbb{I}_2 - \mathbf{w}\mathbf{w}^T)\mathbf{z}_1 + \mathbf{z}_2\times\mathbf{w}\right]\end{aligned}$$

Using $(a\,\mathbb{I}_2 - b\mathbf{w}\mathbf{w}^\top)^{-1} = a^{-1}(\mathbb{I}_2 + \frac{b}{a-b|\mathbf{w}|^2}\mathbf{w}\mathbf{w}^\top)$:

$$\begin{aligned}
\mathbf{z}_2 &= \frac{1-s}{1+s}\left[\mathbb{I}_2 + \frac{\mathbf{w}\mathbf{w}^T}{1-|\mathbf{w}|^2}\right](\mathbf{z}_1 \times \mathbf{w}) = \frac{1-s}{1+s}(\mathbf{z}_1 \times \mathbf{w}). \\
\implies \mathbf{y} &= \frac{1}{1+s}\left[\left((1+s)^2 - (1-s)^2|\mathbf{w}|^2\right)\mathbb{I}_2 - 4s\mathbf{w}\mathbf{w}^T\right]\mathbf{z}_1 \\
&= \frac{1}{1+s}\left[\xi_s(|\mathbf{w}|^2)\mathbb{I}_2 - 4s\mathbf{w}\mathbf{w}^T\right]\mathbf{z}_1 \\
\implies \langle(\mathbf{y}\cdot\sigma), \frac{1+s}{L_P + sR_P}(\mathbf{y}\cdot\sigma)\rangle_{HS} &= \mathrm{Tr}(\mathbf{y}\cdot\sigma)(c\mathbb{I}_2 + \mathbf{z}\cdot\sigma) = 2(\mathbf{y}\cdot\mathbf{z}_1) \\
&= 2(1+s)^2\mathbf{y}\cdot\left[\xi_s(|\mathbf{w}|^2)\mathbb{I}_2 - 4s\mathbf{w}\mathbf{w}^T\right]^{-1}\mathbf{y} \\
&= \frac{2(1+s)^2}{\xi_s(|\mathbf{w}|^2)}\mathbf{y}\cdot\left[\mathbb{I}_2 + \frac{4s}{\xi_s(|\mathbf{w}|^2) - 4s|\mathbf{w}|^2}\mathbf{w}\mathbf{w}^T\right]\mathbf{y} \\
&= \frac{2(1+s)^2}{\xi_s(|\mathbf{w}|^2)}\left[|\mathbf{y}|^2 + \frac{4s(\mathbf{w}\cdot\mathbf{y})^2}{(1+s)^2(1-|\mathbf{w}|^2)}\right]
\end{aligned}$$

This gives the result. □

As special cases, $\kappa_{\max} := \kappa_0, \kappa_{\min} := \kappa_1$, so we have the following results:

Corollary 2.1. *For the traceless Hermitian operator X and density operator ρ given by* (2.25),

$$\|X\|^2_{\kappa_{\max},\,\rho} = D^{\mathrm{std}}_{(x-1)^2}(\gamma\|\rho) = \frac{4|\mathbf{y}|^2}{1-|\mathbf{w}|^2}. \tag{2.26}$$

Corollary 2.2. *For the traceless Hermitian operator X and density operator ρ given by* (2.25),

$$\|X\|^2_{\kappa_{\min},\,\rho} = \frac{|\mathbf{y}|^2(1-|\mathbf{w}|^2) + (\mathbf{w}\cdot\mathbf{y})^2}{1-|\mathbf{w}|^2}. \tag{2.27}$$

Explicit calculation for the BKM metric

It is not practical to compute the relative entropy in the qubit case, so instead [6] worked with the BKM metric (or, more precisely, the corresponding Riemannian semi-norm).

Lemma 2.3. *[6] For the traceless Hermitian operator X and density operator ρ given by (2.25),*

$$\begin{aligned}\|X\|^2_{\rho,x\log x} &= 4|\mathbf{y}|^2\int_1^\infty \frac{u^2+|\mathbf{w}|^2\cos 2\theta}{(u^2-|\mathbf{w}|^2)^2}du \\ &= 2|\mathbf{y}|^2\Big(\frac{1+\cos 2\theta}{1-|\mathbf{w}|^2}+\frac{1-\cos 2\theta}{2|\mathbf{w}|}\ln\frac{1+|\mathbf{w}|}{1-|\mathbf{w}|}\Big).\end{aligned} \tag{2.28}$$

where θ is the angle between $\mathbf{y}$ and $\mathbf{w}$.

Proof. Recall that $X=\mathbf{y}\cdot\sigma$ and $\rho=\frac{1}{2}(\mathbb{I}_2+\mathbf{w}\cdot\sigma)$, use the definition of BKM metric, we have

$$\begin{aligned}\|X\|^2_{\rho,x\log x} &= \int_0^\infty \mathrm{Tr}\Big((\mathbf{y}\cdot\sigma)\big(\frac{1}{2}(\mathbb{I}_2+\mathbf{w}\cdot\sigma)+u\mathbb{I}_2\big)^{-1}(\mathbf{y}\cdot\sigma)\big(\frac{1}{2}(\mathbb{I}_2+\mathbf{w}\cdot\sigma)+u\mathbb{I}_2\big)^{-1}\Big)du \\ &= 4\int_0^\infty \mathrm{Tr}\Big((\mathbf{y}\cdot\sigma)\big((2u+1)\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}(\mathbf{y}\cdot\sigma)\big((2u+1)\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}\Big)du \\ &= 2\int_1^\infty \mathrm{Tr}\Big((\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}(\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}\Big)du.\end{aligned}$$

Then using the Product rule and Inverse rule in (2.24), for any $u>1$, we have

$$(\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}=\frac{(\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2-\mathbf{w}\cdot\sigma\big)}{u^2-|\mathbf{w}|^2}=\frac{-(\mathbf{w}\cdot\mathbf{y})\mathbb{I}_2+(u\mathbf{y}+i\mathbf{w}\times\mathbf{y})\cdot\sigma}{u^2-|\mathbf{w}|^2}$$

thus using the Product rule again,

$$\begin{aligned}&\mathrm{Tr}\Big((\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}(\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}\Big) \\ =&\frac{\mathrm{Tr}\Big(\big(-(\mathbf{w}\cdot\mathbf{y})\mathbb{I}_2+(u\mathbf{y}+i\mathbf{w}\times\mathbf{y})\cdot\sigma\big)^2\Big)}{(u^2-|\mathbf{w}|^2)^2} \\ =&2\frac{|\mathbf{w}\cdot\mathbf{y}|^2+(u\mathbf{y}+i\mathbf{w}\times\mathbf{y})\cdot(u\mathbf{y}+i\mathbf{w}\times\mathbf{y})}{(u^2-|\mathbf{w}|^2)^2} \\ =&2\frac{u^2|\mathbf{y}|^2+|\mathbf{w}\cdot\mathbf{y}|^2-|\mathbf{w}\times\mathbf{y}|^2}{(u^2-|\mathbf{w}|^2)^2}.\end{aligned}$$

Plugging it back to the integral, we have

$$\begin{aligned}\|X\|^2_{\rho,x\log x} &= 2\int_1^\infty \mathrm{Tr}\Big((\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}(\mathbf{y}\cdot\sigma)\big(u\mathbb{I}_2+\mathbf{w}\cdot\sigma\big)^{-1}\Big)du\\ &= 4\int_1^\infty \frac{u^2|\mathbf{y}|^2+|\mathbf{w}\cdot\mathbf{y}|^2-|\mathbf{w}\times\mathbf{y}|^2}{(u^2-|\mathbf{w}|^2)^2}du\\ &= 4|\mathbf{y}|^2\int_1^\infty \frac{u^2+|\mathbf{w}|^2\cos 2\theta}{(u^2-|\mathbf{w}|^2)^2}du.\end{aligned}$$

To compute the above integral, note that for $|\mathbf{w}|<1$, the following calculations hold:

$$\begin{aligned}\int_1^\infty \frac{u^2}{(u^2-|\mathbf{w}|^2)^2}du &= \frac{1}{2}\Big(\frac{1}{1-|\mathbf{w}|^2}-\frac{1}{2|\mathbf{w}|}\ln\frac{1-|\mathbf{w}|}{1+|\mathbf{w}|}\Big),\\ \int_1^\infty \frac{1}{(u^2-|\mathbf{w}|^2)^2}du &= \frac{1}{2|\mathbf{w}|^2}\Big(\frac{1}{1-|\mathbf{w}|^2}+\frac{1}{2|\mathbf{w}|}\ln\frac{1-|\mathbf{w}|}{1+|\mathbf{w}|}\Big).\end{aligned}$$

Therefore, by some simple algebra, we conclude the proof by showing

$$4|\mathbf{y}|^2\int_1^\infty \frac{u^2+|\mathbf{w}|^2\cos 2\theta}{(u^2-|\mathbf{w}|^2)^2}du = 2|\mathbf{y}|^2\left(\frac{1+\cos 2\theta}{1-|\mathbf{w}|^2}+\frac{1-\cos 2\theta}{2|\mathbf{w}|}\ln\frac{1+|\mathbf{w}|}{1-|\mathbf{w}|}\right).$$

□

Explicit calculation for central power metric

Recall, $f_c(x)=\frac{1}{2}(x-1)^2x^{-\frac{1}{2}},\ \kappa_c(x)=\frac{1}{2}x^{-\frac{1}{2}}$.

Lemma 2.4. *[28] For the traceless Hermitian operator X and density operator ρ given by (2.25),*

$$\|X\|^2_{\rho,(x-1)^2x^{-1/2}} = \frac{2|\mathbf{y}|^2}{(1-|\mathbf{w}|^2)(1+\sqrt{1-|\mathbf{w}|^2})}\left(|\mathbf{w}|^2\cos 2\theta+(1+\sqrt{1-|\mathbf{w}|^2})^2\right) \tag{2.29}$$

where θ is the angle between $\mathbf{y}$ and $\mathbf{w}$.

Proof.
Define $P := 2\rho = I+\mathbf{w}\cdot\boldsymbol{\sigma}$. Observe that by (2.24):

$$(\mathbf{y}\cdot\boldsymbol{\sigma})(I+\mathbf{w}\cdot\boldsymbol{\sigma})^{-1/2} = \sqrt{\frac{\zeta(1,|\mathbf{w}|)}{2(1-|\mathbf{w}|^2)}}\left[-\frac{1}{\zeta(1,|\mathbf{w}|)}(\mathbf{w}\cdot\mathbf{y})I+\left(\mathbf{y}+\frac{i}{\zeta(1,|\mathbf{w}|)}(\mathbf{w}\times\mathbf{y})\right)\cdot\sigma\right]$$

$$\begin{aligned}
\operatorname{Tr}\left((\mathbf{y}\cdot\boldsymbol{\sigma})\Omega_P^{x^{-1/2}}(\mathbf{y}\cdot\boldsymbol{\sigma})\right) &= \operatorname{Tr}\left[(\mathbf{y}\cdot\boldsymbol{\sigma})(I+\mathbf{w}\cdot\boldsymbol{\sigma})^{-1/2}\right]^2 \\
&= \frac{\zeta(1,|\mathbf{w}|)}{2(1-|\mathbf{w}|^2)}\left[\frac{1}{\zeta(1,|\mathbf{w}|)^2}(\mathbf{w}\cdot\mathbf{y})^2+|\mathbf{y}|^2-\frac{1}{\zeta(1,|\mathbf{w}|)^2}|\mathbf{w}\times\mathbf{y}|^2\right] \\
&= \frac{\zeta(1,|\mathbf{w}|)}{2(1-|\mathbf{w}|^2)}\left[\frac{1}{\zeta(1,|\mathbf{w}|)^2}|\mathbf{w}|^2|\mathbf{y}|^2(\cos^2\theta-\sin^2\theta)+|\mathbf{y}|^2\right] \\
&= \frac{|\mathbf{y}|^2}{2(1-|\mathbf{w}|^2)\zeta(1,|\mathbf{w}|)}\left(|\mathbf{w}|^2\cos 2\theta+\zeta(1,|\mathbf{w}|)^2\right),
\end{aligned}$$

This gives the result. □

Explicit calculation for the WY metric

Recall, $f_{\mathrm{WY}}(x)=4(1-\sqrt{x})^2,\ \kappa_{\mathrm{WY}}(x)=4(1+\sqrt{x})^{-2}$

Lemma 2.5. *For the density operator ρ,γ given by* (2.25),

$$D_{4(1-\sqrt{x})^2}(\rho,\gamma)=8(1-\frac{1}{2}(\zeta^{1/2}(1,\mathbf{w})\zeta^{1/2}(1,\mathbf{u})+\frac{\mathbf{w}\cdot\mathbf{u}}{\zeta^{1/2}(1,\mathbf{w})\zeta^{1/2}(1,\mathbf{u})})) \tag{2.30}$$

Proof.

$$\begin{aligned}
D_{4(1-\sqrt{x})^2}(\rho,\gamma) &= \langle\gamma^{1/2}, f_{WY}(L_\rho R_\gamma^{-1})(\gamma^{1/2})\rangle = 4\langle\gamma^{1/2},(1-L_\rho^{1/2}R_\gamma^{-1/2})^2(\gamma^{1/2})\rangle \\
&= 4\ \operatorname{Tr}\ \gamma^{1/2}(\gamma^{1/2}-2\rho^{1/2}+\rho\gamma^{-1/2}) = 4\ \operatorname{Tr}\ (\gamma+\rho-2\rho^{1/2}\gamma^{1/2}) \\
&= 8\ (1-\operatorname{Tr}\ \rho^{1/2}\gamma^{1/2}) \\
&= 8\ (1-\frac{1}{4}\zeta^{1/2}(1,\mathbf{w})\zeta^{1/2}(1,\mathbf{u})\ \operatorname{Tr}\ (\mathbb{I}_2+\frac{\mathbf{w}\cdot\sigma}{\zeta(1,\mathbf{w})})(\mathbb{I}_2+\frac{\mathbf{u}\cdot\sigma}{\zeta(1,\mathbf{u})})) \\
&= 8(1-\frac{1}{2}\zeta^{1/2}(1,\mathbf{w})\zeta^{1/2}(1,\mathbf{u})(1+\frac{\mathbf{w}\cdot\mathbf{u}}{\zeta(1,\mathbf{w})\zeta(1,\mathbf{u})})) \\
&= 8(1-\frac{1}{2}(\zeta^{1/2}(1,\mathbf{w})\zeta^{1/2}(1,\mathbf{u})+\frac{\mathbf{w}\cdot\mathbf{u}}{\zeta(1,\mathbf{w})^{1/2}\zeta(1,\mathbf{u})^{1/2}}))
\end{aligned}$$

□

For the formula of the WY Riemannian metric, the following lemma is useful:

Lemma 2.6. *[28] Let $P = \mathbb{I}_2 + \mathbf{w} \cdot \sigma$ with $|\mathbf{w}| < 1$, i.e., $\rho = \frac{P}{2}$ is a mixed state. Then for every $y \in \mathbb{R}^3$:*

$$\frac{1}{L_{P/2} + R_{P/2}}(\mathbf{y} \cdot \sigma) = \frac{2}{L_P + R_P}(\mathbf{y} \cdot \sigma) = -\frac{\mathbf{w} \cdot \mathbf{y}}{1 - |\mathbf{w}|^2}\mathbb{I}_2 + \left(\mathbf{y} + \frac{\mathbf{w} \cdot \mathbf{y}}{1 - |\mathbf{w}|^2}\mathbf{w}\right) \cdot \sigma$$

Proof.
Write $c\mathbb{I}_2 + \mathbf{z} \cdot \sigma = 2(L_P + R_P)^{-1}(\mathbf{y} \cdot \sigma)$, then

$$\begin{aligned}\mathbf{y} \cdot \sigma &= \frac{1}{2}(\mathbb{I}_2 + \mathbf{w} \cdot \sigma)(c\mathbb{I}_2 + \mathbf{z} \cdot \sigma) + \frac{1}{2}(\mathbb{I}_2 + \mathbf{w} \cdot \sigma)(c\mathbb{I}_2 + \mathbf{z} \cdot \sigma) \\ &= (c + \mathbf{w} \cdot \mathbf{z})\,\mathbb{I}_2 + (\mathbf{z} + c\mathbf{w}) \cdot \sigma\end{aligned}$$

$\therefore c = -\mathbf{w} \cdot \mathbf{z}, \quad \mathbf{y} = \mathbf{z} + c\mathbf{w} = (\mathbb{I}_2 - \mathbf{w}\mathbf{w}^\top)\mathbf{z}$

Using $(\mathbb{I}_2 - \mathbf{w}\mathbf{w}^\top)^{-1} = \mathbb{I}_2 + \frac{1}{1-|\mathbf{w}|^2}\mathbf{w}\mathbf{w}^\top$:

$$\mathbf{z} = \mathbf{y} + \frac{\mathbf{w} \cdot \mathbf{y}}{1 - |\mathbf{w}|^2}\mathbf{w}, \; c = -\mathbf{w} \cdot \mathbf{z} = -\frac{\mathbf{w} \cdot \mathbf{y}}{1 - |\mathbf{w}|^2}$$

□

Lemma 2.7. *[28] For the traceless Hermitian operator X and density operator ρ given by (2.25),*

$$\|X\|^2_{\rho,\kappa_{\mathrm{WY}}} = \frac{8}{\zeta(|\mathbf{y}|^2, \mathbf{w})} + \frac{2(3\zeta(1,\mathbf{w})^2 - |\mathbf{w}|^2)}{(\zeta(1,\mathbf{w})^2 - |\mathbf{w}|^2)^2}(\mathbf{w} \cdot \mathbf{y})^2 \tag{2.31}$$

Proof.
Define $P := 2\rho = \mathbb{I}_2 + \mathbf{w} \cdot \boldsymbol{\sigma}$. Observe that by (2.24), we can write $P^{1/2} = \sqrt{\frac{\zeta(1,\mathbf{w})}{2}}(\mathbb{I}_2 + \frac{\mathbf{w} \cdot \sigma}{\zeta(1,\mathbf{w})})$, so by considering (2.6):

$$\frac{2}{\sqrt{L_P} + \sqrt{R_P}}(\mathbf{y} \cdot \boldsymbol{\sigma}) = \sqrt{\frac{2}{\zeta(1,\mathbf{w})}}\left[-\frac{\zeta(1,\mathbf{w})\,\mathbf{w} \cdot \mathbf{y}}{\zeta(1,\mathbf{w})^2 - |\mathbf{w}|^2}\mathbb{I}_2 + \left(\mathbf{y} + \frac{(\mathbf{w} \cdot \mathbf{y})\mathbf{w}}{\zeta(1,\mathbf{w})^2 - |\mathbf{w}|^2}\right) \cdot \boldsymbol{\sigma}\right]$$

with $\zeta(1,\mathbf{w}) = \zeta(1,\mathbf{w})(|\mathbf{w}|) = 1 + \sqrt{1 - |\mathbf{w}|^2}$. Therefore,

$$
\begin{aligned}
&\left\langle \mathbf{y}\cdot\sigma,\ \frac{4}{(\sqrt{L_P}+\sqrt{R_P})^2}\,(\mathbf{y}\cdot\sigma)\right\rangle_{HS} \\
&= \mathrm{Tr}\left[\frac{2}{\sqrt{L_P}+\sqrt{R_P}}(\mathbf{y}\cdot\boldsymbol{\sigma})\right]^2 \\
&= \frac{4}{\zeta(1,\mathbf{w})}\left[\frac{\zeta(1,\mathbf{w})^2(\mathbf{w}\cdot\mathbf{y})^2}{(\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2)^2}+|\mathbf{y}|^2+\frac{2(\mathbf{w}\cdot\mathbf{y})^2}{\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2}+\frac{(\mathbf{w}\cdot\mathbf{y})^2|\mathbf{w}|^2}{(\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2)^2}\right] \\
&= \frac{4}{\zeta(1,\mathbf{w})}\left[|\mathbf{y}|^2+(\mathbf{w}\cdot\mathbf{y})^2\,\frac{3\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2}{(\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2)^2}\right] \\
&= 4\left\langle \mathbf{y},\left[\zeta(1,\mathbf{w})\mathbb{I}_2-(2-\zeta(1,\mathbf{w})^{-1})\mathbf{w}\mathbf{w}^T\right]^{-1}\mathbf{y}\right\rangle_{HS}
\end{aligned}
$$

Since $(a\mathbb{I}_2-b\mathbf{w}\mathbf{w}^T)^{-1}=a^{-1}(\mathbb{I}_2+\frac{b}{a-b|\mathbf{w}|^2}\mathbf{w}\mathbf{w}^T)^{-1}$

$$
\begin{aligned}
&\left[\zeta(1,\mathbf{w})\mathbb{I}_2-(2-\zeta(1,\mathbf{w})^{-1})\mathbf{w}\mathbf{w}^T\right]^{-1}=\frac{1}{\zeta(1,\mathbf{w})}\left[\mathbb{I}_2+\frac{3\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2}{(\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2)^2}\mathbf{w}\mathbf{w}^T\right] \\
&\implies \left\langle \mathbf{y}\cdot\sigma,\ \frac{4}{(\sqrt{L_P}+\sqrt{R_P})^2}\,(\mathbf{y}\cdot\sigma)\right\rangle_{HS}=\frac{4}{\zeta(|\mathbf{y}|^2,\mathbf{w})}+\frac{3\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2}{(\zeta(1,\mathbf{w})^2-|\mathbf{w}|^2)^2}(\mathbf{w}\cdot\mathbf{y})^2
\end{aligned}
$$

□

Any qubit linear map $\mathcal{N}:\mathbb{M}_2\to\mathbb{M}_2$ has a one-to-one correspondence to a 4×4 matrix $\mathcal{T}_{\mathcal{N}}$ in the basis of Pauli operators:

$$
\mathcal{N}(c_0\mathbb{I}_2+c_1\sigma_x+c_2\sigma_y+c_3\sigma_z)=c_0'\mathbb{I}_2+c_1'\sigma_x+c_2'\sigma_y+c_3'\sigma_z,\quad \mathbf{c}'=\mathcal{T}_{\mathcal{N}}\mathbf{c}. \tag{2.32}
$$

If $\mathcal{N}$ is trace-preserving, we must have $c_0=c_0'$ thus $\mathcal{T}_{\mathcal{N}}$ has the form

$$
\mathcal{T}_{\mathcal{N}}=\begin{pmatrix}1&0&0&0\\ t_1&a_{11}&a_{12}&a_{13}\\ t_2&a_{21}&a_{22}&a_{23}\\ t_3&a_{31}&a_{32}&a_{33}\end{pmatrix}. \tag{2.33}
$$

If $\mathcal{N}$ is Hermitian-preserving, it is clear that all the elements of $\mathcal{T}_{\mathcal{N}}$ are real. Denote

$$
T=\begin{pmatrix}a_{11}&a_{12}&a_{13}\\ a_{21}&a_{22}&a_{23}\\ a_{31}&a_{32}&a_{33}\end{pmatrix}\in\mathbb{M}_3(\mathbb{R}),\quad \mathbf{t}=\begin{pmatrix}t_1\\ t_2\\ t_3\end{pmatrix}\in\mathbb{R}^3. \tag{2.34}
$$

For any $\rho = \frac{1}{2}(\mathbb{I}_2 + \mathbf{w} \cdot \sigma)$, $\mathcal{N}(\rho)$ can be represented as

$$\mathcal{N}(\rho) = \frac{1}{2}(\mathbb{I}_2 + (T\mathbf{w} + \mathbf{t}) \cdot \sigma). \tag{2.35}$$

We refer the reader to [8] for a complete analysis on the pair $(T, \mathbf{t})$ such that $\mathcal{N}$ is a quantum channel. Here we only remark that if $\mathcal{N}$ is positive, then $\forall \mathbf{w} \in \mathbb{R}^3$ with $|\mathbf{w}| \leqslant 1$, we have $|T\mathbf{w} + \mathbf{t}| \leqslant 1$.

Given $X = \mathbf{y} \cdot \sigma$ and $\rho = \frac{1}{2}(\mathbb{I}_2 + \mathbf{w} \cdot \sigma)$ with $|\mathbf{w}| \leqslant 1$, we denote

$$\mathbf{y}_{\mathcal{N}} = T\mathbf{y}, \quad \mathbf{w}_{\mathcal{N}} = T\mathbf{w} + \mathbf{t}. \tag{2.36}$$

Using Lemma 2.3, we have

$$g_{\mathcal{N}(\rho)}(\mathcal{N}(X)) = 2|\mathbf{y}_{\mathcal{N}}|^2 \left(\frac{1 + \cos 2\theta_{\mathcal{N}}}{1 - |\mathbf{w}_{\mathcal{N}}|^2} + \frac{1 - \cos 2\theta_{\mathcal{N}}}{2|\mathbf{w}_{\mathcal{N}}|} \ln \frac{1 + |\mathbf{w}_{\mathcal{N}}|}{1 - |\mathbf{w}_{\mathcal{N}}|} \right) \tag{2.37}$$

$$= \frac{4|\mathbf{y}_{\mathcal{N}}|^2}{1 - |\mathbf{w}_{\mathcal{N}}|^2} \left(\cos^2 \theta_{\mathcal{N}} + \sin^2 \theta_{\mathcal{N}} f(|\mathbf{w}_{\mathcal{N}}|) \right), \tag{2.38}$$

where $\theta_{\mathcal{N}}$ is the angle between $\mathbf{y}_{\mathcal{N}}$ and $\mathbf{w}_{\mathcal{N}}$, and the function f is

$$f(x) := \frac{1 - x^2}{2x} \ln \frac{1 + x}{1 - x}, \quad x \in [0, 1]. \tag{2.39}$$

Note that $f(x) > 0$ for any $x \in [0, 1)$ and $f(1) = 0$. When $x \to 1-$,

$$f(x) \sim -(1 - x^2) \ln\big(1 - x^2\big). \tag{2.40}$$

Chapter 3

Properties of Quantum f-Divergences

3.1 Classical f-divergences

To provide some context for the different quantum f-divergences that we'll meet, it will be useful to consider the classical f-divergences, denoted D_f^{cl}. After all, a quantum f-divergence is a functional $D_f : \mathcal{B}(\mathcal{H})_+ \times \mathcal{B}(\mathcal{H})_+ \to (-\infty, \infty]$ that reduces to the classical f-divergence over commuting density operators $\rho = \sum_{x\in\mathcal{X}} p_x |x\rangle\langle x|$, $\gamma = \sum_{x\in\mathcal{X}} q_x |x\rangle\langle x|$ via $D_f(\rho\|\gamma) = D_f^{\text{cl}}(\{p_x\}_x\|\{q_x\}_x)$ [25]. The following results are provided so that they can be referred to in the later preliminary sections, but our interests are primarily in the quantum setting, so the proofs will be left out. For more details, see [53], which has been a very convenient resource and is the basis for this section.

Definition 3.1. *(Classical f-divergence). Let $f : (0,\infty) \to \mathbb{R}$ be a convex function with $f(1) = 0$. Let P and Q be two probability distributions on a measurable space $(\mathcal{X}, \mathcal{E})$. If $P \ll Q$, then the f-divergence is defined as*

$$D_f^{\text{cl}}(P\|Q) := \mathbb{E}^Q\left[f\left(\frac{dP}{dQ}\right)\right]$$

where $\frac{dP}{dQ}$ is a Radon–Nikodym derivative and $f(0) := f(0+)$. More generally, let $f'(\infty) := \lim_{x\to 0^+} xf(1/x)$. Suppose that $Q(dx) = q(x)\mu(dx)$ and $P(dx) = p(x)\mu(dx)$ for some common dominating measure μ, then we have

$$D_f^{\text{cl}}(P\|Q) = \int_{q>0} q(x)f\left(\frac{p(x)}{q(x)}\right)d\mu + f'(\infty)P[q=0]$$

with the agreement that if $P[q = 0] = 0$, the last term is taken to be zero regardless of the value of $f'(\infty)$ (which could be infinite).

Remark. For the discrete case, with $Q(x)$ and $P(x)$ being the respective pmfs, we can also write

$$D_f^{\mathrm{cl}}(P\|Q) = \sum_x Q(x) f\left(\frac{P(x)}{Q(x)}\right) \tag{3.1}$$

with the understanding that

- $f(0) := f(0^+)$,
- $0f\left(\frac{0}{0}\right) := 0$, and
- $0f\left(\frac{a}{0}\right) := \lim_{x\downarrow 0} xf\left(\frac{a}{x}\right) = af'(\infty) \quad$ for $a > 0$.

Recall the Radon-Nikodym theorem and thus the Radon-Nikodym derivative [58]:

Theorem 3.1 (Radon-Nikodym Theorem)**.** *The probability measure Q is equivalent to the probability measure P if and only if there exists a P-a.s. positive random variable Z such that*

$$Q(A) = \mathbb{E}^P\left(Z\,\mathbb{1}_A\right) = \int_A Z\,dP.$$

for each $A \in \mathcal{F}$.

The random variable Z is called the density, or the Radon–Nikodym derivative, of Q with respect to P, and is often denoted

$$Z = \frac{dQ}{dP}.$$

Also, P has a density with respect to Q given by

$$\frac{dP}{dQ} = \frac{1}{Z}.$$

The above Radon-Nikodym derivative relates two probability distributions in a similar way to how we observed in Section 2.1 that the relative modular operator relates two quantum states; actually they do precisely the same [25]. In the non-commutative setting, however, there arises an ambiguity in defining the Radon-Nikodym derivative. The relative modular operator we met in Section 2.1 is one way, corresponding to the standard f-divergence, but alternative choices can give us other variations of the quantum f-divergence. While considering choices of the Radon-Nikodym derivative can be helpful, it does not strictly give functionals that satisfy monotonicity. Monotonicity is fundamental for valid distinguishability measures; we are generally interested in finding *monotone distinguishability measures* even if they don't strictly arise from such a choice.

In accordance with the theme of this section, classical f-divergences are made useful by satisfying many of the properties of the relative entropy (i.e. KL-divergence):

Theorem 3.2 (Properties)**.** *[53]*
The following hold:

1. $D^{\text{cl}}_{f_1+f_2}(P\|Q) = D^{\text{cl}}_{f_1}(P\|Q) + D^{\text{cl}}_{f_2}(P\|Q)$.
2. $D^{\text{cl}}_{f}(P\|P) = 0$.
3. $D^{\text{cl}}_{f}(P\|Q) = 0$ *for all* $P \neq Q$ *iff* $f(x) = c(x-1)$ *for some* c*. For any other* f*, we have*
$$D^{\text{cl}}_{f}(P\|Q) = f(0) + f'(\infty) > 0 \quad \text{for } P \perp Q.$$
4. *If* $P_{X,Y} = P_X P_{Y|X}$ *and* $Q_{X,Y} = P_X Q_{Y|X}$*, then the function* $D^{\text{cl}}_{f}(P_{Y|X=x}\|Q_{Y|X=x})$ *is* $\mathcal{X}$*-measurable and*
$$D^{\text{cl}}_{f}(P_{X,Y}\|Q_{X,Y}) = \int_{\mathcal{X}} dP_X(x) D^{\text{cl}}_{f}(P_{Y|X=x}\|Q_{Y|X=x}) =: D^{\text{cl}}_{f}(P_{Y|X}\|Q_{Y|X}|P_X). \tag{3.2}$$
5. *If* $P_{X,Y} = P_X P_{Y|X}$ *and* $Q_{X,Y} = Q_X P_{Y|X}$*, then*
$$D^{\text{cl}}_{f}(P_{X,Y}\|Q_{X,Y}) = D^{\text{cl}}_{f}(P_X\|Q_X). \tag{3.3}$$
6. *Let* $f_1(x) = f(x) + c(x-1)$*, then*
$$D^{\text{cl}}_{f_1}(P\|Q) = D^{\text{cl}}_{f}(P\|Q) \quad \forall P, Q.$$
7. *Non-negativity:* $D_f(P\|Q) \geqslant 0$*. If* f *is strictly convex at 1, then* $D^{\text{cl}}_{f}(P\|Q) = 0$ *if and only if* $P = Q$*.*
8. *Joint convexity:* $(P, Q) \mapsto D^{\text{cl}}_{f}(P\|Q)$ *is a jointly convex function. Consequently,* $P \mapsto D_f(P\|Q)$ *and* $Q \mapsto D^{\text{cl}}_{f}(P\|Q)$ *are also convex.*
9. *Monotonicity*
$$D^{\text{cl}}_{f}(P_{X,Y}\|Q_{X,Y}) \geqslant D^{\text{cl}}_{f}(P_X\|Q_X). \tag{3.4}$$

Since we want positivity for the quantum f-divergences (and thus it must be true for commuting input density operators), we are therefore interested in choices for f that are strictly convex at 1. In fact, this is the reason for the $f''(1) > 0$ condition in $\mathcal{F}$ and $\mathcal{F}_{\text{sym}}$.

The above monotonicity property is equivalent to monotonicity under an arbitrary classical channel (stochastic map). The monotonicity of a quantum f-divergence is a bit different, not only because dealing with quantum channels is necessarily stronger, but also because we will sometimes consider more restricted classes of functions f than we can in the classical setting ($i.e.\mathcal{F}$).

3.2 Standard f-divergences

We will now review some of the results from [29][25], which derived many important properties of the quantum f-divergence. In particular, we will meet their proof of the monotonicity of quantum f-divergences.

Let us recall the expression for the standard f-divergence from 2.1. If $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$ with spectral decompositions

$$\rho = \sum_{a \in \text{spec}(\rho)} a P_a, \quad \gamma = \sum_{b \in \text{spec}(\gamma)} b Q_b \tag{3.5}$$

then we have

$$\begin{aligned} D_f^{\text{std}}(\rho\|\gamma) &= \lim_{\varepsilon \to 0} \langle (\gamma + \varepsilon I)^{1/2}, f(\Delta_{\rho+\varepsilon I, \gamma+\varepsilon I})((\gamma+\varepsilon I)^{1/2}) \rangle_{HS} \\ &= \langle \gamma^{1/2}, f(\Delta_{\rho,\gamma})(\gamma^{1/2}) \rangle_{HS} + f'(\infty) \operatorname{Tr} \rho(1-\gamma^0) && (2.3) \\ &= \sum_{\substack{a \in \text{spec}\, \rho, \\ b \in \text{spec}\, \gamma \setminus \{0\}}} b f\left(\frac{a}{b}\right) \operatorname{Tr} P_a Q_b + a f'(\infty) \operatorname{Tr} P_a Q_0 && (2.5) \end{aligned}$$

We can compare this with the classical f-divergence as follows:

Proposition 3.1 (Nussbaum-Szkoła)**.** *[29, Lemma 2.9] If $\rho, \gamma \in \mathcal{D}(\mathcal{H})$ have the spectral decompositions given in 3.5, then:*

$$D_f^{\text{std}}(\rho\|\gamma) = D_f^{\text{cl}}(\{P(a,b)\}_{a,b} \| \{Q(a,b)\}_{a,b}) \tag{3.6}$$

where $P(a,b) := a \operatorname{Tr}(P_a Q_b)$, $Q(a,b) := b \operatorname{Tr}(P_a Q_b)$ are probability distributions.

Proof. Substitute $P(a,b), Q(a,b)$ into (3.1) and compare. $\square$

The term *Nussbaum-Szkoła distribution* for a pair of states ρ, γ is specifically used for the case that P_a, Q_b are chosen as rank-1 projectors, and they are recognised as a rather versatile tool. Many inequalities established for classical f-divergences for general pairs of distributions can be directly extended to standard f-divergences in this way (see [1]). In particular, they have been used to establish the Chernoff bound and Hoeffding bound (which correspond to error exponents of symmetric and anti-symmetric hypothesis testing respectively).

However, the expression 2.5 can also be used to establish other important properties of the standard f-divergence (simply evaluate both sides):

Corollary 3.1.
Let $A, A_1, A_2, B, B_1, B_2 \in \mathcal{B}(\mathcal{H})_+$ and $f : [0,\infty) \to (0,\infty)$. We have the following:

(i) For every $\lambda \in [0, +\infty)$,
$$D_f^{\text{std}}(\lambda A \| \lambda B) = \lambda D_f^{\text{std}}(A\|B).$$

(ii) If $A_1^0, B_1^0 \perp A_2^0, B_2^0$ (i.e. only the pairs A_1, B_1 and A_2, B_2 can share supports), then
$$D_f^{\text{std}}(A_1 + A_2 \| B_1 + B_2) = D_f^{\text{std}}(A_1\|B_1) + D_f^{\text{std}}(A_2\|B_2).$$

(iii) If $V : \mathcal{H} \to \mathcal{K}$ is a linear or anti-linear isometry, then
$$D_f^{\text{std}}(VAV^* \| VBV^*) = D_f^{\text{std}}(A\|B).$$
In particular, D_f^{std} is transpose invariant.

(iv) If x is a unit vector in some other Hilbert space $\mathcal{K}$, then
$$D_f^{\text{std}}(A \otimes |x\rangle\langle x| \| B \otimes |x\rangle\langle x|) = D_f^{\text{std}}(A\|B).$$

Finally, it is important to not take for granted the continuity of standard f-divergences. The following result tells us precisely when we have continuity.

Proposition 3.2.
Given any $A_k, B_k \in \mathcal{B}(\mathcal{H})_+$ for all $k \in \mathbb{N}$, and assume that $\lim_{k\to\infty} A_k = A$, $\lim_{k\to\infty} B_k = B$. Then

1. *If* $f'(\infty) < \infty$, $A_k = A$ *for all* $k \in \mathbb{N}$,

$$\lim_{k\to\infty} D_f^{\mathrm{std}}(A\|B_k) = D_f^{\mathrm{std}}(A\|B).$$

2. *If* $f(0) < \infty$, $B_k = B$ *for all* $k \in \mathbb{N}$,

$$\lim_{k\to\infty} D_f^{\mathrm{std}}(A_k\|B) = D_f^{\mathrm{std}}(A\|B).$$

3. *If* $f(0), f'(\infty) < \infty$,

$$\lim_{k\to\infty} D_f^{\mathrm{std}}(A_k\|B_k) = D_f^{\mathrm{std}}(A\|B).$$

Proof.
According to the appropriate case, note that $S(A_k\|B_k)$ is finite for any k. Then by the definition of the standard f-divergence, we can choose a sequence $\varepsilon_k, \varepsilon'_k > 0$, $k \in \mathbb{N}$, such that $\lim_{k\to\infty} \varepsilon_k = 0$, and for all $k \in \mathbb{N}$,

$$D_f^{\mathrm{std}}(A_k + \varepsilon_k I\|B_k + \varepsilon'_k I) - \frac{1}{k} < D_f^{\mathrm{std}}(A_k\|B_k) < D_f^{\mathrm{std}}(A + \varepsilon_k I\|B_k + \varepsilon'_k I) + \frac{1}{k}.$$

Let $\widetilde{A}_k := A_k + \varepsilon_k I, \widetilde{B}_k := B_k + \varepsilon_k I$, which is strictly positive for any $k \in \mathbb{N}$. Obviously, $\lim_{k\to\infty} \widetilde{A}_k = A, \lim_{k\to\infty} \widetilde{B}_k = B$, and the assertion will follow if we can show that

$$\lim_{k\to\infty} D_f^{\mathrm{std}}(\widetilde{A}_k\|\widetilde{B}_k) = D_f^{\mathrm{std}}(A\|B).$$

Let

$$A = \sum_{a\in\mathrm{spec}(A)} aP_a, \quad B = \sum_{b\in\mathrm{spec}(B)} bQ_b, \quad \widetilde{A}_k = \sum_{a_k\in\mathrm{spec}(\widetilde{A}_k)} a_k P_{a_k}^{(k)}, \quad \widetilde{B}_k = \sum_{b_k\in\mathrm{spec}(\widetilde{B}_k)} b_k Q_{b_k}^{(k)}$$

be the spectral decompositions of the respective operators. Then

$$D_f^{\mathrm{std}}(\widetilde{A}_k\|\widetilde{B}_k) = \sum_{a\in\mathrm{spec}(\widetilde{A}_k)} \sum_{c\in\mathrm{spec}(\widetilde{B}_k)} b_k f\left(\frac{a_k}{b_k}\right) \mathrm{Tr}(P_{a_k}^{(k)} Q_{b_k}^{(k)}).$$

Then consider Rellich's Theorem Version I (Theorem C.1) to get the result:

1. If $b > 0$: $b_k f(\frac{a_k}{b_k}) \to bf(\frac{a}{b})$
2. If $b = 0$: $b_k f(\frac{a_k}{b_k}) \to af'(\infty) =: 0f(\frac{a}{0})$
3. $\sum_{a_k\to a} P_{a_k}^{(k)} \to P_a, \sum_{b_k\to b} P_{b_k}^{(k)} \to P_b$

□

3.2.1 Monotonicity

We recall the proof of monotonicity of the standard f-divergence; for the details, see [26]. The point is to have an awareness of the role played by the relevant lemmas and the power of decomposing the operator convex function f. Recall, for reference in the following lemmas, that for a positive semi-definite (PSD) operator A, A^0 denotes the projector onto supp A.

This proof applies to *substochastic maps* Φ, i.e. $\hat{\Phi}(Y^*)\hat{\Phi}(Y) \leqslant \hat{\Phi}(Y^*Y)$ for all $Y \in \mathcal{B}(\mathcal{H}_2)$. In fact, this implies that Φ is positive and trace-decreasing. In the special case that Φ is also trace-preserving, it is called *stochastic*, and a CPTP map is necessarily stochastic [26]. We will work primarily with CPTP maps in the later sections, so this result is actually stronger than for our purposes; working in greater generality is mostly for mathematical interest.

In this proof [26] used an alternative integral representation to (2.6), instead expanding around $x = 0$:

Theorem 3.3. *A continuous real-valued function f on $[0, \infty)$ is operator convex iff it can be expressed as*

$$f(x) = f(0) + ax + bx^2 + \int_{(0,\infty)} \left(\frac{x}{1+t} - \frac{x}{x+t} \right) d\mu(t), \quad x \in [0, \infty)$$

for unique $a \in \mathbb{R}$, $b \geqslant 0$, non-negative measure μ on $(0, \infty)$ such that

$$b = \lim_{x\to\infty} \frac{f(x)}{x^2}, \quad a = f(1) - f(0) - b, \quad \int_{(0,\infty)} \frac{d\mu(t)}{(1+t)^2} < \infty$$

Proposition 3.3. *A continuous real-valued function f on $[0, \infty)$ is operator convex and has*

$$f'(\infty) = \lim_{x\to\infty} \frac{f(x)}{x} < \infty,$$

iff it can be expressed as:

$$f(x) = f(0) + f'(\infty)x - \int_{(0,\infty)} \frac{x}{x+t} \, d\mu(t), \quad x \in [0, \infty)$$

for positive measure μ on $(0, \infty)$ s.t.

$$\int_{(0,\infty)} \frac{d\mu(t)}{1+t} < \infty$$

Let us list the other results involved in the proof:

Lemma 3.1. *[26, Lemma 3.2]*
Let $B \in \mathcal{B}(\mathcal{H}_1)_+$ *and let* $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ *be a positive map s.t.*

$$\Phi^*(\Phi(B)^0) \leqslant I_1$$

Then

$$\operatorname{Tr} \Phi(B) \leqslant \operatorname{Tr} B$$

(all of the above are implied when Φ *is trace-decreasing)*

And TFAE:

(i) $\operatorname{Tr} \Phi(B) = \operatorname{Tr} B$

(ii) $\forall$ *functions* f *on* $\operatorname{spec}(B)$ *s.t.* $f(0) = 0$ *if* $0 \in \operatorname{span}(B)$*:*

$$f(B)\Phi^*(\Phi(B)^0) = \Phi^*(\Phi(B)^0)f(B) = f(B)$$

(iii) $\operatorname{supp} B \leqslant \operatorname{supp} P_{\{1\}}(\Phi^*(\Phi(B)^0))$*, where* $P_{\{1\}}(X)$ *denotes the spectral projection on the 1-eigenspace of* X

(iv) Φ *is trace-preserving on* $B^0\mathcal{B}(\mathcal{H}_1)B^0$
($\therefore$ $A \in \mathcal{B}(\mathcal{H}_1)_+$ *s.t.* $\operatorname{supp} A \leqslant \operatorname{supp} B$ *satisfies* $\operatorname{Tr} \Phi(A) = \operatorname{Tr} A$*)*

(v) For the map Φ^*_B*, we have:*

$$\Phi^*_B(\Phi(B)) = B$$

Lemma 3.2. *[26, Lemma 3.5]*

Let $A, B \in \mathcal{B}(\mathcal{H}_1)_+$ *and* $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ *be a positive map. Then*

$$\Phi(B^0AB^0)\Phi(B)^{-1}\Phi\left(B^0AB^0\right) \leqslant \Phi(B^0AB^{-1}AB^0)$$

(implied by Φ *a Schwarz map, since a Schwarz map is necessarily positive)*

In particular, if $\operatorname{supp} A \leqslant \operatorname{supp} B$ *then:*

$$\Phi(A)\Phi(B)^{-1}\Phi(A) \leqslant \Phi(AB^{-1}A)$$

If, moreover, Φ is also trace-decreasing, then:

$$S_{f_2}(\Phi(A)\|\Phi(B)) = \mathrm{Tr}\,\Phi(A)^2\Phi(B)^{-1} \leqslant \mathrm{Tr}\,A^2B^{-1} = S_{f_2}(A\|B)$$

(where S_{f_2} is 2-Renyi)

Lemma 3.3. *[26, Lemma 4.1, Lemma 4.2]*
For a $B \in \mathcal{B}(\mathcal{H}_1)_+$ and a substochastic map $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$, let

$$V : \mathcal{B}(\mathcal{H}_2) \to \mathcal{B}(\mathcal{H}_1),\ X \mapsto \Phi^*(X\Phi(B)^{-1/2})B^{1/2}$$

then $V = R_{B^{1/2}} \circ \Phi^* \circ R_{\Phi(B)^{-1/2}}$ is a contraction*, and:*

$$V^*(L_A R_{B^{-1}})V \leqslant L_{\Phi(A)} R_{\Phi(B)^{-1}}$$

Further,

$$V(\Phi(B^{1/2})) = B^{1/2} \quad \textit{iff} \quad \mathrm{Tr}\,\Phi(B) = \mathrm{Tr}\,B \quad \textit{(via Lemma 3.1)}$$

Theorem 3.4 (Monotonicity of Standard f-divergences)**.**
Let $A, B \in \mathcal{B}(\mathcal{H}_1)_+$ and $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ be a substochastic map s.t.

$$\mathrm{Tr}\,\Phi(B) = \mathrm{Tr}\,B,$$

and let f be an operator convex function on $[0, +\infty)$.

Assume further, either:

$$\mathrm{Tr}\,\Phi(A) = \mathrm{Tr}\,A \quad \textit{or} \quad f'(\infty) \geqslant 0$$

Then:

$$D_f^{\mathrm{std}}(\Phi(A)\|\Phi(B)) \leqslant D_f^{\mathrm{std}}(A\|B)$$

Proof.
PART 1: When f is continuous at 0

Step 1: Break down the problem and assume $\mathrm{supp}\,A \leqslant \mathrm{supp}\,B$.
Using Theorem 3.3,

$$f(x) = f(0) + ax + bx^2 + \int_{(0,\infty)} \frac{x}{1+t} + \varphi_t(x)d\mu(t), \quad \varphi_t(x) := -\frac{x}{x+t}$$

Let $\Delta := L_A R_{B^{-1}}$, $\tilde{\Delta} := L_{\Phi(A)} R_{\Phi(B)^{-1}}$ then,

$$D_f^{\mathrm{std}}(A\|B) = f(0)\mathrm{Tr}\, B + a\mathrm{Tr}\, AB^0 + b\mathrm{Tr}\, A^2 B^{-1} \tag{3.7}$$

$$+ \int_{(0,\infty)} \left(\frac{\mathrm{Tr}\, AB^0}{1+t} + D_{\varphi_t}^{\mathrm{std}}(A\|B) \right) d\mu(t) + f'(\infty)\mathrm{Tr}\, A(I - B^0) \tag{3.8}$$

But $\mathrm{Tr}\, B = \mathrm{Tr}\, \Phi(B)$ and since $b \geqslant 0$,

$$b\mathrm{Tr}\, A^2 B^{-1} \geqslant b\mathrm{Tr}\, \Phi(A)^2 \Phi(B)^{-1} \qquad (\because \text{ Lemmas 3.3 and 3.1 })$$

$\mathrm{Tr}\, \Phi(A) = \mathrm{Tr}\, A \qquad (\because \mathrm{supp}\, A \leqslant \mathrm{supp}\, B$ and by Lemma 3.1)

$$\Rightarrow \mathrm{Tr}\, AB^0 = \mathrm{Tr}\, A = \mathrm{Tr}\, \Phi(A) = \mathrm{Tr}\, \Phi(A)\Phi(B)^0$$
$$(\because \mathrm{supp}\, A \leqslant \mathrm{supp}\, B \Rightarrow \mathrm{supp}\, \Phi(A) \leqslant \mathrm{supp}\, \Phi(B))$$
$$\Rightarrow \mathrm{Tr}\, A(I - B^0) = \mathrm{Tr}\, \Phi(A)(I - \Phi(B)^0) = 0$$

φ_t is operator convex, operator monotone decreasing, $\varphi_t(0) = 0$

$\therefore V^* \varphi_t(\Delta) V \geqslant \varphi_t(V^* \Delta V) \geqslant \varphi_t(\tilde{\Delta})$ for V in Lemma 3.3

$$\begin{aligned} \therefore \quad D_{\varphi_t}^{\mathrm{std}}(A\|B) &= \langle B^{1/2}, \varphi_t(\Delta)(B^{1/2}) \rangle_{HS} \\ &= \langle V(\Phi(B)^{1/2}), \varphi_t(\Delta) V(\Phi(B)^{1/2}) \rangle_{\mathrm{HS}} \quad \text{(via Lemma 3.3)} \\ &\geqslant \langle \Phi(B)^{1/2}, \varphi_t(\tilde{\Delta}) \Phi(B)^{1/2} \rangle_{\mathrm{HS}} = D_{\varphi_t}^{\mathrm{std}}(\Phi(A)\|\Phi(B)) \end{aligned}$$

$\therefore D_f^{\mathrm{std}}(\Phi(A)\|\Phi(B)) \leqslant D_f^{\mathrm{std}}(A\|B)$, since we've shown that all the terms are monotone!

Step 2: Reduce all cases to the previous case (for most part)

Case 1 $(\omega(f) = \infty)$:

1. If $\mathrm{supp}\, A \nleqslant \mathrm{supp}\, B$: $D_f^{\mathrm{std}}(A\|B) = \infty$, so monotonicity holds (refer to (3.7))
2. If $\mathrm{supp}\, A \leqslant \mathrm{supp}\, B$: previous step applies.

Case 2 $(\mathrm{Tr}\, \Phi(A) = \mathrm{Tr}\, A)$:

Let $B_\varepsilon := B + \varepsilon A$ so that $\mathrm{supp}\, A \leqslant \mathrm{supp}\, B_\varepsilon$ and $\mathrm{Tr}\, \Phi(B_\varepsilon) = \mathrm{Tr}\, B_\varepsilon$ by linearity

∴ previous argument applies:

$$D_f^{\mathrm{std}}(\Phi(A)\|\Phi(B_\varepsilon)) \leqslant D_f^{\mathrm{std}}(A\|B_\varepsilon)$$

$f'(\infty) < \infty$ WLOG (otherwise, previous case applies)

∴ 2nd-argument continuity of f-divergence yields monotonicity as $\varepsilon \to 0$

Case 3 ($0 \leqslant \omega(f) < \infty$):

By Proposition 3.3,

$$f(x) = f(0) + f'(\infty)x + \int_{(0,\infty)} \varphi_t(x)\, d\mu(t)$$

$\therefore D_f^{\mathrm{std}}(A\|B) = f(0)\mathrm{Tr}\, B + f'(\infty)\mathrm{Tr}\, AB^0 + \int_{(0,\infty)} D_{\varphi_t}^{\mathrm{std}}(A\|B)\, d\mu(t) + f'(\infty)\mathrm{Tr}\, A(I - B^0)$

$$= f(0)\mathrm{Tr}\, B + f'(\infty)\mathrm{Tr}\, A + \int_{(0,\infty)} D_{\varphi_t}^{\mathrm{std}}(A\|B)\, d\mu(t)$$

and since $\mathrm{Tr}\, \Phi(A) \leqslant \mathrm{Tr}\, A$, all terms satisfy monotonicity!

PART 2: Separate the discontinuity of f at 0
Consider $\tilde{f}_\alpha = -x^\alpha$, $x \in [0,\infty)$, $\alpha \in (0,1)$.

These functions are operator convex, continuous at 0, and $\tilde{f}'_\alpha(\infty) = 0\ \forall \alpha \in (0,1)$.
$\therefore -\mathrm{Tr}\, \Phi(A)^\alpha \Phi(B)^{1-\alpha} = D_{\tilde{f}_\alpha}^{\mathrm{std}}(\Phi(A)\|\Phi(B)) \leqslant D_{\tilde{f}_\alpha}^{\mathrm{std}}(A\|B) = -\mathrm{Tr}\, A^\alpha B^{1-\alpha}$

∴ Taking $\alpha \downarrow 0$:

$$\mathrm{Tr}\, \Phi(A)^0 \Phi(B) \geqslant \mathrm{Tr}\, A^0 B$$

$\therefore S_{\mathbb{1}_{\{0\}}}(\Phi(A)\|\Phi(B)) = \mathrm{Tr}\, \Phi(B) - \mathrm{Tr}\, \Phi(A)^0\Phi(B) \leqslant \mathrm{Tr}\, B - \mathrm{Tr}\, A^0 B = S_{\mathbb{1}_{\{0\}}}(A\|B)$

Let $\delta = f(0) - f(0^+)$, then using $\tilde{f} := f - \delta \mathbb{1}_{\{0\}}$ (which is continuous by construction):

$$\begin{aligned} D_f^{\text{std}}(\Phi(A)\|\Phi(B)) &= D_{\tilde{f}}^{\text{std}}(\Phi(A)\|\Phi(B)) + \delta D_{\mathbb{1}_{\{0\}}}^{\text{std}}(\Phi(A)\|\Phi(B)) \\ &\leqslant D_{\tilde{f}}^{\text{std}}(A\|B) + \delta D_{\mathbb{1}_{\{0\}}}^{\text{std}}(A\|B) \\ &= D_f^{\text{std}}(A\|B) \quad \text{(by linearity)} \end{aligned}$$

□

3.3 Other Quantum f-Divergences

3.3.1 Maximal f-Divergences

In Section 2.1, we motivated the standard f-divergence via the relative modular operator $\Delta_{\rho,\gamma} = L_\rho R_\gamma^{-1}$, which was chosen as the Radon-Nikodym derivative (this terminology was introduced in Section 3.1) for $\rho, \gamma \in \mathcal{D}^+(\mathcal{H})$. Alternatively, we can choose $\widetilde{\Delta}_{\rho,\gamma} := R_{\sigma^{-1/2}\rho\sigma^{-1/2}}$, which is called the *commutant Radon-Nikodymn derivative* [42, 25], and this defines the following quantum f-divergence:

Definition 3.2 (Maximal f-divergence)**.** *[25]*
Let $f \in \mathcal{F}$ be an operator convex function with $f(1) = 0, f''(1) > 0$. Let $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$ be two positive semi-definite operators s.t. $\operatorname{supp}\rho = \operatorname{supp}\gamma$*. The maximal f-divergence is defined by*

$$\widehat{D}_f(\rho\|\gamma) := \langle \gamma^{1/2}, f(\gamma^{-1/2}\rho\gamma^{-1/2})\gamma^{1/2}\rangle_{HS} = \operatorname{Tr}\, \gamma^{1/2} f(\gamma^{-1/2}\rho\gamma^{-1/2})\gamma^{1/2} \tag{3.9}$$

For general $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$,

$$\widehat{D}_f(\rho\|\gamma) := \lim_{\varepsilon\to 0^+} \widehat{D}_f(\rho + \varepsilon I\|\gamma + \varepsilon I) \tag{3.10}$$

From this definition, it is not difficult to see that maximal f-divergences satisfy the many convenient properties of standard f-divergences from Corollary 3.1.

Proposition 3.4.
Let $A, A_1, A_2, B, B_1, B_2 \in \mathcal{B}(\mathcal{H})_+$ and $f \in \mathcal{F}$. We have the following:

(i) For every $\lambda \in [0, +\infty)$,
$$\widehat{D}_f(\lambda A\|\lambda B) = \lambda \widehat{D}_f(A\|B).$$

(ii) If $A_1^0, B_1^0 \perp A_2^0, B_2^0$ (i.e. only the pairs A_1, B_1 and A_2, B_2 can share supports), then
$$\widehat{D}_f(A_1 + A_2\|B_1 + B_2) = \widehat{D}_f(A_1\|B_1) + \widehat{D}_f(A_2\|B_2).$$

(iii) If $V : \mathcal{H} \to \mathcal{K}$ is a linear or anti-linear isometry, then
$$\widehat{D}_f(VAV^*\|VBV^*) = \widehat{D}_f(A\|B).$$
In particular, $\widehat{D}_f$ is transpose invariant.

(iv) If x is a unit vector in some other Hilbert space $\mathcal{K}$, then
$$\widehat{D}_f(A \otimes |x\rangle\langle x|\|B \otimes |x\rangle\langle x|) = \widehat{D}_f(A\|B).$$

Operator Perspectives and Monotonicity

To appreciate the maximal f-divergence as a valid quantum f-divergence, we will recall the key details of a proof of its monotonicity w.r.t. positive trace-preserving operators by [25]. All of the following definitions come from [25], as they attempted to generalise previous works that only considered positive definite density operators and functions f s.t. $f(0^+) < \infty$ like [42].

To understand this proof, we first introduce the *operator perspective* P_f of a function $f \in \mathcal{F}$, which is the two-variable operator function contained within the trace of (3.9). i.e. for two positive semi-definite operators $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$ with the same support, it is defined as:

$$P_f : (\rho, \gamma) \mapsto \gamma^{1/2} f(\gamma^{-1/2} \rho \gamma^{-1/2}) \gamma^{1/2}.$$

Note that $P_f(\rho, \gamma) \equiv P_{\tilde{f}}(\gamma, \rho)$, where $\tilde{f}(x) := x f(x^{-1}) \in \mathcal{F}$ is the transpose of f.

We can define this operator perspective more generally. We observe the following:

Proposition 3.5. *[25, Proposition 3.29]*
Provided some $f \in \mathcal{F}$, $\widehat{D}_f(\rho\|\gamma) = \infty$ iff one of the following conditions hold:

1. $f'(\infty) = \infty$ *and* $\operatorname{supp} \rho \nleqslant \operatorname{supp} \gamma$
2. $f(0^+) = \infty$ *and* $\operatorname{supp} \sigma \nleqslant \operatorname{supp} \rho$

In all other cases $\widehat{D}_f(\rho\|\gamma)$ is finite.

Proof.
Consider condition 1. There exists a pure state $|\psi\rangle\langle\psi| \leqslant \operatorname{supp} \rho$ such that $\gamma |\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| \gamma = 0$ by assumption.

By definition, $D_f(\rho\|\gamma) := \lim_{\varepsilon \to 0^+} \operatorname{Tr} \, P_f(\rho + \varepsilon I, \gamma + \varepsilon I)$. Since the perspective $P_f(\rho + \varepsilon I, \gamma + \varepsilon I)$ is positive semi-definite for all $\varepsilon > 0$, by the definition of trace, we have the following lower bound:

$$\begin{aligned} D_f(\rho\|\gamma) &\geqslant \lim_{\varepsilon \to 0^+} \langle\psi| P_f(\rho + \varepsilon I, \gamma + \varepsilon I) |\psi\rangle \\ &= \lim_{\varepsilon \to 0^+} \langle\psi| (\gamma + \varepsilon I)^{1/2} f((\gamma + \varepsilon I)^{-1/2} (\rho + \varepsilon I)(\gamma + \varepsilon I)^{-1/2}) (\gamma + \varepsilon I)^{1/2} |\psi\rangle \end{aligned}$$

Finally, by considering $\gamma |\psi\rangle\langle\psi| = |\psi\rangle\langle\psi| \gamma = 0$ and Jensen's inequality:

$$\begin{aligned} D_f(\rho\|\gamma) &\geqslant \lim_{\varepsilon\to 0^+} \varepsilon \langle\psi| f((\gamma+\varepsilon I)^{-1/2}(\rho+\varepsilon I)(\gamma+\varepsilon I)^{-1/2}) |\psi\rangle \\ &\geqslant \lim_{\varepsilon\to 0^+} \varepsilon f(\langle\psi| (\gamma+\varepsilon I)^{-1/2}(\rho+\varepsilon I)(\gamma+\varepsilon I)^{-1/2} |\psi\rangle) \\ &= \lim_{\varepsilon\to 0^+} \varepsilon f(\varepsilon^{-1}(\langle\psi| \rho |\psi\rangle + 1)) \\ &= (\langle\psi| \rho |\psi\rangle + 1) f'(\infty) = \infty \end{aligned}$$

Starting instead from condition 2, the result is immediate: simply consider $f \to \tilde{f}$, then $\tilde{f}'(\infty) = f(0^+) = \infty$ and $D_f(\rho\|\gamma) \equiv \lim_{\varepsilon\to 0^+} \operatorname{Tr}\, P_{\tilde{f}}(\gamma+\varepsilon I, \rho+\varepsilon I)$, so this reduces to the previous case. □

The operator perspective is now extended to provide a convenient expression for the maximal divergence:

Proposition 3.6. *[25, Corollary 3.28]*
Provided some $f \in \mathcal{F}$ such that $\widehat{D}_f(\rho\|\gamma) < \infty$, define the operator perspective for $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$ (not necessarily $\operatorname{supp}\rho = \operatorname{supp}\gamma$*) by:*

$$P_f(\rho, \gamma) := \lim_{n\to\infty} P_f(\rho + K_n, \gamma + K_n)$$

Where $K_n \in \mathcal{B}(\mathcal{H})_+$ is any sequence s.t. $\rho + K_n, \gamma + K_n > 0$ for every n, and $K_n \to 0$. Then the maximal f-divergence is given by:

$$\widehat{D}_f(\rho\|\gamma) = \operatorname{Tr}\, P_f(\rho, \gamma)$$

We will need the following *Choi-Davis-Jensen inequality*, that can be used to fully characterise operator convex functions [14, 10, 16, 4]:

Lemma 3.4 (Choi-Davis-Jensen inequality)**.**
Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ be a unital positive linear map, $A \in \mathcal{B}(\mathcal{H})_{sa}$, and f an operator convex function defined on an interval containing $\operatorname{spec}\rho$, *then:*

$$f(\Phi(A)) \leqslant \Phi(f(A))$$

Now that we have all of the ingredients, we will complete the proof:

Theorem 3.5 (Monotonicity of Maximal f-divergences)**.** *[25, 27, 42]*
Provided an $f \in \mathcal{F}$ and a trace-preserving positive linear map $\Phi : \mathcal{B}(\mathcal{H}_1) \to \mathcal{B}(\mathcal{H}_2)$ $\rho, \gamma \in \mathcal{B}(\mathcal{H}_1)_+$, then for every $\rho, \gamma \in \mathcal{B}(\mathcal{H}_1)_+$:

$$\widehat{D}_f(\Phi(\rho)\|\Phi(\gamma)) \leqslant \widehat{D}_f(\rho\|\gamma)$$

Proof.
If $\widehat{D}_f(\rho\|\gamma) = \infty$, the claim holds trivially. Therefore, wlog, we suppose $\widehat{D}_f(\rho\|\gamma) < \infty$. In this case, $\widehat{D}_f(\rho\|\gamma) = \text{Tr } P_f(\rho, \gamma)$ by Proposition 3.6.

To construct appropriate unital maps to apply Lemma 3.4, we first restrict the domain of the output states of Φ to ensure that there are full-rank output states, that can be transformed into the identity. We do this by slightly modifying Φ as a map onto $\Phi(I)^0\mathcal{B}(\mathcal{K})\Phi(I)^0 = \mathcal{B}(\Phi(I)^0\mathcal{K})$, so we can wlog set $\Phi(I)^0 = I$.

Let $\rho_n := \rho + n^{-1}I, \gamma_n := \gamma + n^{-1}I$, so that $\operatorname{supp}\rho_n = \operatorname{supp}\gamma_n = \operatorname{supp} I$. Then we define a positive unital map Φ_{σ_n}:

$$\Phi_{\sigma_n} : A \mapsto \Phi(\gamma_n)^{-1/2}\Phi(\Phi(\gamma_n)^{1/2}A\Phi(\gamma_n)^{1/2})\Phi(\gamma_n)^{-1/2}.$$

By Lemma 3.4,

$$f(\Phi_{\gamma_n}(\gamma_n^{-1/2}\rho_n\gamma_n^{-1/2})) \leqslant \Phi(f(\gamma_n^{-1/2}\rho_n\gamma_n^{-1/2}))$$

Which can be expresses as follows:

$$\begin{aligned} P_f(\Phi(\rho_n), \Phi(\gamma_n)) &= \Phi(\gamma_n)^{1/2}f(\Phi(\gamma_n)^{-1/2}\Phi(\rho_n)\Phi(\gamma_n)^{-1/2})\Phi(\gamma_n)^{1/2} \\ &\leqslant \Phi(\gamma_n^{1/2}f(\gamma_n^{-1/2}\rho_n\gamma_n^{-1/2})\gamma_n^{1/2}) \\ &= \Phi(P_f(\rho_n, \gamma_n)) \end{aligned}$$

Finally, we take the trace of both sides. Since Φ is trace-preserving:

$$\widehat{D}_f(\Phi(\rho_n), \Phi(\gamma_n)) \leqslant \widehat{D}_f(\rho_n, \gamma_n)$$

And by taking $n \to \infty$, we obtain the result. □

The Maximality of the Maximal f-Divergence

There is an interesting interpretation of the maximal f-divergence first observed by [42], as the solution of the following optimisation problem over classical f-divergences [42, 25]:

$$\begin{aligned} \widehat{D}_f(\rho\|\gamma) := \inf\{D_f^{\text{cl}}(\mathbf{p}\|\mathbf{q}) : \mathbf{p}, \mathbf{q} \in \mathcal{B}(\mathcal{H}')_+ \text{ are commuting}, \dim \mathcal{H}' < \infty \text{ and} \\ \Phi(\mathbf{p}) = \rho, \Phi(\mathbf{q}) = \gamma \text{ for some CPTP map } \Phi : \mathcal{B}(\mathcal{K}) \to \mathcal{B}(\mathcal{H})\} \end{aligned}$$

It turns out that as a consequence, the maximal f-divergence is the largest of all monotone quantum f-divergences:

Theorem 3.6 (Maximality of Maximal f-divergence)**.** *[25, Proposition 4.1] Suppose we are provided a monotone quantum f-divergence D_f, some operator convex function $f \in \mathcal{F}$, then for every $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$:*

$$D_f(\rho \| \gamma) \leqslant \widehat{D}_f(\rho \| \gamma)$$

Proof.
Both of these f-divergences are defined by considering $\rho + \varepsilon I, \gamma + \varepsilon I$ and taking $\varepsilon \to 0^+$. wlog, we assume ρ, γ are positive definite.

We take a look at the spectral decomposition of $\gamma^{-1/2} \rho \gamma^{-1/2}$:

$$\gamma^{-1/2} \rho \gamma^{-1/2} = \sum_{i=1}^{k} \lambda_i P_i,$$

where P_i are orthogonal projections, $\sum_{i=1}^{k} P_i = I$, and $\lambda_i \geqslant 0$.

Observe the following relations:

1. $\rho = \sum_{i=1}^{k} \lambda_i \gamma^{1/2} P_i \gamma^{1/2}$
2. $\gamma = \sum_{i=1}^{k} \gamma^{1/2} P_i \gamma^{1/2}$
3. $\widehat{D}_f(\rho \| \gamma) = \operatorname{Tr} \gamma^{1/2} f(\gamma^{-1/2} \rho \gamma^{-1/2}) \gamma^{1/2} = \sum_i \operatorname{Tr} \gamma P_i \, f((\lambda_i \operatorname{Tr} \gamma P_i)(\operatorname{Tr} \gamma P_i)^{-1})$

Let $\mathbf{e}_i$ be a orthonormal basis for (the commutative algebra) $\mathbb{C}^k$:

$$\mathbf{p} := \sum_i (\lambda_i \operatorname{Tr} \gamma P_i) \mathbf{e}_i, \quad \mathbf{q} := \sum_i (\operatorname{Tr} \gamma P_i) \mathbf{e}_i$$

Then $\widehat{D}_f(\rho \| \gamma) = D_f^{\mathrm{cl}}(\mathbf{p} \| \mathbf{q})$. Thus, to apply the new definition of the maximal f-divergence and prove this result, we need to find a CPTP map Φ such that $\Phi(\mathbf{p}) = \rho, \Phi(\mathbf{q}) = \gamma$ and use the monotonicity of D_f.

The following choice for $\Phi : \mathbb{C}^k \to \mathcal{B}(\mathcal{H})$ satisfies these requirements:

$$\Phi(\sum_i x_i \mathbf{e}_i) := \sum_i x_i \frac{\gamma^{1/2} P_i \gamma^{1/2}}{\operatorname{Tr} \gamma P_i}$$

As a result,

$$D_f(\rho\|\gamma) = D_f(\Phi(\mathbf{p})\|\Phi(\mathbf{q})) \leqslant D_f(\mathbf{p}\|\mathbf{q}) = D_f^{\mathrm{cl}}(\mathbf{p}\|\mathbf{q}) = \widehat{D}_f(\rho\|\gamma)$$

□

Unlike the standard f-divergence, which we saw each had different local second order behaviours up to symmetry (Lemma 2.1), the local behaviour of the maximal f-divergences are all the same: they are given by the maximal Riemannian metric, corresponding to $\kappa_{\max}$. As noted by [42], this local behaviour is quite fundamental via the following similarity to the optimisation problem for the maximal f-divergence:

$$\|X\|^2_{\kappa_{\max},\,\rho} = \operatorname{Tr} X^2\rho^{-1} = \inf\{\sum_{i=1}^{\dim\mathcal{H}'}\frac{v_i^2}{p_i} : \mathbf{p},\mathbf{v}\in\mathcal{B}(\mathcal{H}')_+ \text{ are commuting}, \dim\mathcal{H}'<\infty \text{ and } \Phi(\mathbf{p})=\rho, \Phi(\mathbf{v})=X \text{ for some CPTP map } \Phi:\mathcal{B}(\mathcal{K})\to\mathcal{B}(\mathcal{H})\}$$

Originally, this observation of the maximal f-divergences' local behaviour was made by [51]:

Theorem 3.7 (Maximal f-divergences are Locally Maximal Riemannian)**.** *Let $f\in\mathcal{F}$, then for any density operator ρ and traceless Hermitian operator X, with* $\operatorname{supp}X\leqslant\operatorname{supp}\rho$, *acting on the finite d-dimesnional Hilbert space $\mathcal{H}$, we define $\rho+\varepsilon X\geqslant 0$ for $\varepsilon\in\mathbb{R}$ sufficiently small. Then*

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\widehat{D}_f(\rho_\varepsilon\|\rho) = \frac{1}{2}\frac{d^2}{d\varepsilon^2}\Big|_{\varepsilon=0}\widehat{D}_f(\rho_\varepsilon\|\rho) = \frac{f''(1)}{2}\|X\|^2_{\kappa_{\max},\rho}$$

where $\kappa_{\max}(x) := \frac{x+1}{2x}$.

Proof.
We can find an open neighbourhood $I\subseteq\mathbb{R}$ of 0 such that $\rho_\varepsilon\geqslant 0$, $\operatorname{supp}\rho_\varepsilon = \operatorname{supp}\rho\ \forall\varepsilon\in I$.

By the Taylor expansion of $f(x)$ (which is infinitely differentiable at $x=1$ by assumption), since $\rho^{-1/2}\rho_\varepsilon\rho^{-1/2}-I = \varepsilon\rho^{-1/2}X\rho^{-1/2}$:

$$\begin{aligned}\widehat{D}_f(\rho_\varepsilon\|\rho) &= \operatorname{Tr}\rho^{1/2}f(\rho^{-1/2}\rho_\varepsilon\rho^{-1/2})\rho^{1/2}\\ &= f'(1)\operatorname{Tr}\rho^{1/2}(\rho^{-1/2}\rho_\varepsilon\rho^{-1/2}-I)\rho^{1/2} + \frac{f''(1)}{2}\operatorname{Tr}\rho^{1/2}(\rho^{-1/2}\rho_\varepsilon\rho^{-1/2}-I)^2\rho^{1/2} + o(\varepsilon^2)\\ &= \frac{\varepsilon^2}{2}f''(1)\operatorname{Tr}X\rho^{-1}X + o(\varepsilon^2)\end{aligned}$$

Which gives the result, as $\|X\|^2_{\kappa_{\max},\,\rho} = \operatorname{Tr}X^2\rho^{-1}$. □

3.3.2 Measured f-Divergences

The *measured (or minimal) f-divergence* for a convex function f is, like the maximal f-divergence, defined via an optimisation over values of the classical f-divergence:

Definition 3.3 (Measured f-Divergence)**.** *[25]*
Suppose we are given a convex function f. Let $\{M_x\}$ denote a (variable) POVM on $\mathcal{H}$. For any $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$, set $p_x := \mathrm{Tr}[M_x \rho]$, $q_x := \mathrm{Tr}[M_x \gamma]$. The measured f-divergence is

$$D_f^{\mathrm{meas}}(\rho\|\gamma) := \sup_{\{M_x\}} \sum_x q_x\, f\left(\frac{p_x}{q_x}\right) \equiv \sup_{\{M_x\}} D_f^{\mathrm{cl}}(\mathbf{p}\|\mathbf{q}).$$

From this definition, we can deduce that the following convenient properties of the standard f-divergences hold for the measured f-divergences (Corollary 3.1):

Proposition 3.7.
Let $A, A_1, A_2, B, B_1, B_2 \in \mathcal{B}(\mathcal{H})_+$ and $f : (0, \infty) \to \mathbb{R}$ be a convex function. We have the following:

(i) For every $\lambda \in [0, +\infty)$,

$$D_f^{\mathrm{meas}}(\lambda A\|\lambda B) = \lambda D_f^{\mathrm{meas}}(A\|B).$$

(ii) If $A_1^0, B_1^0 \perp A_2^0, B_2^0$ (i.e. only the pairs A_1, B_1 and A_2, B_2 can share supports), then

$$D_f^{\mathrm{meas}}(A_1 + A_2\|B_1 + B_2) = D_f^{\mathrm{meas}}(A_1\|B_1) + D_f^{\mathrm{meas}}(A_2\|B_2).$$

(iii) If $V : \mathcal{H} \to \mathcal{K}$ is a linear or anti-linear isometry, then

$$D_f^{\mathrm{meas}}(VAV^*\|VBV^*) = D_f^{\mathrm{meas}}(A\|B).$$

In particular, D_f^{meas} is transpose invariant.

(iv) If x is a unit vector in some other Hilbert space $\mathcal{K}$, then

$$D_f^{\mathrm{meas}}(A \otimes |x\rangle\langle x|\|B \otimes |x\rangle\langle x|) = D_f^{\mathrm{meas}}(A\|B).$$

Proof. *(i),(ii),(iv)* are trivial. Here we will only show *(ii)*.
Choose some variable POVM $\{M_x\}$.
We can choose orthogonal projectors $\{\Pi^{(i)}\}_{i=0}^2$ s.t. $\Pi^{(0)} + \Pi^{(1)} + \Pi^{(2)} = I$, $\Pi^{(i)} A_i \Pi^{(i)} = A_i$, $\Pi^{(i)} B_i \Pi^{(i)} = B_i$, i.e. $A_i^0, B_i^0 \leqslant \Pi^{(i)}$.
We define $M_x^{(i)} := \Pi^{(i)} M_x \Pi^{(i)}$, so that $\{M_x^{(i)}\}_{x,i}$ is our new POVM.
Let $p_x^i := \mathrm{Tr}[M_x^{(i)} A_i]$, $q_x^i := \mathrm{Tr}[M_x^{(i)} B_i]$, $i = 0, 1$ (all other expectation values are 0), then we consider $p_x = p_x^0 + p_x^1$, $q_x = q_x^0 + q_x^1$:

$$D_f^{\mathrm{meas}}(A_0 + A_1 \| B_0 + B_2) = \sup_{\{M_x\}} \sum_x q_x f\left(\frac{p_x}{q_x}\right) = \sup_{\{M_x\}} \sum_x (q_x^0 + q_x^1) f\left(\frac{p_x^0 + p_x^1}{q_x^0 + q_x^1}\right)$$

But by applying Jensen's inequality to f:

$$\begin{aligned}
(q_x^0 + q_x^1) f\left(\frac{p_x^0 + p_x^1}{q_x^0 + q_x^1}\right) &= (q_x^0 + q_x^1) f\left(\frac{q_x^0}{q_x^0 + q_x^1} \cdot \frac{p_x^0}{q_x^0} + \frac{q_x^1}{q_x^0 + q_x^1} \cdot \frac{p_x^1}{q_x^1}\right) \\
&\leqslant (q_x^0 + q_x^1) \left[\frac{q_x^0}{q_x^0 + q_x^1} \cdot f\left(\frac{p_x^0}{q_x^0}\right) + \frac{q_x^1}{q_x^0 + q_x^1} \cdot f\left(\frac{p_x^1}{q_x^1}\right)\right] \\
&= q_x^0 f\left(\frac{p_x^0}{q_x^0}\right) + q_x^1 f\left(\frac{p_x^1}{q_x^1}\right) \\
\Rightarrow D_f^{\mathrm{meas}}(A_0 + A_1 \| B_0 + B_2) &\leqslant \sup_{\{M_x\}} \sum_x \sum_{i=0}^1 q_x^i f\left(\frac{p_x^i}{q_x^i}\right) \\
&\leqslant \sum_{i=0}^1 \sup_{\{M_x^{(i)}\}} \sum_x q_x^{(i)} f\left(\frac{p_x^{(i)}}{q_x^{(i)}}\right) \\
&= D_f^{\mathrm{meas}}(A_0 \| B_0) + D_f^{\mathrm{meas}}(A_1 \| B_1)
\end{aligned}$$

By the orthogonal supports $A_1^0, B_1^0 \perp A_2^0, B_2^0$, any POVMs associated to $D_f^{\mathrm{meas}}(A_0 \| B_0)$, $D_f^{\mathrm{meas}}(A_1 \| B_1)$ can be combined into an overall POVM:

$$D_f^{\mathrm{meas}}(A_0 \| B_0) + D_f^{\mathrm{meas}}(A_1 \| B_1) \leqslant D_f^{\mathrm{meas}}(A_0 + A_1 \| B_0 + B_1)$$

Thus we have our result.

□

Since the measurement $\mathcal{M}$ given by $\{M_x\}_{x \in \mathcal{X}}$ ($\mathcal{X}$ the finite set of possible outcomes) can be written as a CPTP map $\mathcal{M} : \mathcal{B}(\mathcal{H}) \to \mathcal{C}^{\mathcal{X}}, A \mapsto \sum_x (\mathrm{Tr}\, A M_x) \mathbf{e}_x$ for an orthonormal basis $\mathbf{e}_x$, this is in fact the smallest of all monotone quantum f-divergences (alluded to in [25]):

Theorem 3.8 (Minimality of Measured f-divergence)**.**
Suppose we are provided a monotone quantum f-divergence D_f, some convex function $f : (0, \infty) \to \mathbb{R}$, then for every $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$:

$$D_f^{\text{meas}}(\rho \| \gamma) \leqslant D_f(\rho \| \gamma)$$

Proof. For any measurement $\mathcal{M}$ given by $\{M_x\}$, by monotonicity:

$$D_f^{\text{cl}}(\mathbf{p} \| \mathbf{q}) = D_f(\mathcal{M}(\rho) \| \mathcal{M}(\gamma)) \leqslant D_f(\rho \| \gamma)$$

Taking the supremum over all choices of $\mathcal{M}$ gives the result. □

For completeness, we also include a proof of monotonicity of the measured f-divergence, to see that it is indeed a monotone quantum f-divergence (similar to a proof by [23] in the von Neumann algebra setting):

Theorem 3.9 (Monotonicity of measured f-divergences)**.**
Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H}')$ be a positive, trace-decreasing map. Then for any density operators $\rho, \gamma \in \mathcal{D}(\mathcal{H})$ (where $\mathcal{H}$ is a finite-dimensional Hilbert space):

$$D_f^{meas}(\Phi(\rho) \| \Phi(\gamma)) \leqslant D_f^{meas}(\rho \| \gamma)$$

Proof.
Φ is positive, trace non-increasing $\Leftrightarrow$ $\hat{\Phi}$ positive, $\hat{\Phi}(I) \leqslant I$

Let $\{M_x\}_{x=1}^k$ be a variable POVM, and for states $\rho, \gamma \in \mathcal{D}(\mathcal{H})$, consider

$$p_x := \text{Tr}[M_x \Phi(\rho)] \equiv \text{Tr}[\hat{\Phi}(M_x)\rho], \quad q_x := \text{Tr}[M_x \Phi(\gamma)] \equiv \text{Tr}[\hat{\Phi}(M_x)\gamma].$$

We can construct another POVM $\{M'_x\}_{x=1}^{k+1}$ via:

$$M'_x := \hat{\Phi}(M_x) \geqslant 0, \quad M'_{k+1} := I - \sum_x \hat{\Phi}(M_x) \geqslant 0$$

Define

$$p'_x := \text{Tr}[M'_x \rho], \quad q'_x := \text{Tr}[M'_x \gamma]$$

Note: $p'_x = p_x, q'_x = q_x, x = 1, \ldots, k$

$$\begin{aligned} \therefore D_f^{\text{meas}}(\Phi(\rho) \| \Phi(\gamma)) := \sup_{\{M_x\}} \sum_{x=1}^k q_x f\left(\frac{p_x}{q_x}\right) &\leqslant \sup_{\{M_x\}} \sum_{x=1}^k q_x f\left(\frac{p_x}{q_x}\right) + q'_{k+1} f\left(\frac{p'_{k+1}}{q'_{k+1}}\right) \\ &\leqslant \sup_{\{M'_x\}} \sum_x \text{Tr}[M'_x \gamma] f\left(\frac{\text{Tr}[M'_x \rho]}{\text{Tr}[M'_x \gamma]}\right) = D_f^{\text{meas}}(\rho \| \gamma) \end{aligned}$$

Thus we have our result. □

3.3.3 Sandwiched Rényi Divergence

In this section, the sandwiched Rényi divergence is defined for positive operators $\rho, \gamma \in \mathcal{B}(\mathcal{H})_+$ and $\alpha > 0$ as:

$$\widetilde{D}_\alpha(\rho||\gamma) = \begin{cases} \alpha' \log \left\| \left(\gamma^{-\frac{1}{2\alpha'}} \frac{\rho}{\operatorname{Tr}\rho} \gamma^{-\frac{1}{2\alpha'}} \right) \right\|_\alpha & \alpha \in (0,1), \text{ or } \alpha \in (1,\infty) \text{ and } \operatorname{supp}(\rho) \leqslant \operatorname{supp}(\gamma) \\ \infty & \alpha \in (1,\infty) \text{ and } \operatorname{supp}(\rho) \nleqslant \operatorname{supp}(\gamma) \\ (\operatorname{Tr}\rho)^{-1} \operatorname{Tr}\rho\,(\log\rho - \log\gamma) & \alpha = 1 \\ \log \|\gamma^{-1/2}\rho\gamma^{-1/2}\|_\infty & \alpha = \infty \end{cases}$$

where $\alpha' := \frac{\alpha}{\alpha-1} = (1-\frac{1}{\alpha})^{-1}$ is the Hölder conjugate of α and $\|X\|_\alpha := (\operatorname{Tr}|X|^\alpha)^{\frac{1}{\alpha}}$. For convenience, we also define a superoperator $\Gamma_\gamma(X) := \gamma^{\frac{1}{2}} X \gamma^{\frac{1}{2}}$ for operators X, positive γ. Implicitly, $\Gamma_\gamma^{-\frac{1}{\alpha}}(X) := \gamma^{-\frac{1}{2\alpha}} X \gamma^{-\frac{1}{2\alpha}}$, $\alpha > 0$. And we define the sandwiched α-norm (which is only a norm for $\alpha \geqslant 1$) as:

$$\|X\|_{\alpha,\gamma} := \|\Gamma_\gamma^{-\frac{1}{\alpha'}}(X)\|_\alpha = \|\gamma^{-\frac{1}{2\alpha'}} X \gamma^{-\frac{1}{2\alpha'}}\|_\alpha$$

We'll of course only be interested in $\operatorname{Tr}\rho = \operatorname{Tr}\gamma = 1$.

From this definition, we can deduce that the Sandwiched Rényi divergence shares the following convenient properties with the examples of quantum f-divergences that we have met in previous sections:

Proposition 3.8.
Let $A, A_1, A_2, B, B_1, B_2 \in \mathcal{B}(\mathcal{H})_+$ and $f \in \mathcal{F}$. We have the following for all $\alpha > 0$:

(i) For every $\lambda \in [0, +\infty)$,

$$\widetilde{D}_\alpha(\lambda A \| \lambda B) = \lambda \widetilde{D}_\alpha(A\|B).$$

(ii) If $A_1^0, B_1^0 \perp A_2^0, B_2^0$ (i.e. only the pairs A_1, B_1 and A_2, B_2 can share supports), then

$$\widetilde{D}_\alpha(A_1 + A_2 \| B_1 + B_2) = \widetilde{D}_\alpha(A_1\|B_1) + \widetilde{D}_\alpha(A_2\|B_2).$$

(iii) If $V : \mathcal{H} \to \mathcal{K}$ is a linear or anti-linear isometry, then

$$\widetilde{D}_\alpha(VAV^* \| VBV^*) = \widetilde{D}_\alpha(A\|B).$$

In particular, $\widetilde{D}_\alpha$ is transpose invariant.

(iv) If x is a unit vector in some other Hilbert space $\mathcal{K}$, then

$$\widetilde{D}_\alpha(A \otimes |x\rangle\langle x| \| B \otimes |x\rangle\langle x|) = \widetilde{D}_\alpha(A\|B).$$

In order to be a useful measure of distinguishability between quantum states, the sandwiched Rényi divergence must satisfy positivity, and monotonicity under CPTP maps. These were properties that we saw to be true for the large class of standard f-divergences, but unlike the Petz Rényi relative entropy, this is not a standard f-divergence; note that the fidelity between quantum states corresponds to $\alpha = \frac{1}{2}$, i.e. $D_{\frac{1}{2}}(\rho||\gamma) = -2\log \mathrm{Tr}|\sqrt{\rho}\sqrt{\gamma}| = -\log F(\rho,\gamma)$. In fact, [15] showed that the sandwiched Rényi divergence is monotone under CPTP maps for $\alpha \geqslant \frac{1}{2}$, and [5] showed positivity for all $\alpha \in (0,1) \cup (1,\infty)$ and that $\alpha \mapsto \widetilde{D}_\alpha(\rho||\gamma), \alpha > 1$ is increasing. [5] provided an alternative proof of monotonicity for $\alpha > 1$ that exploited the fact that the statement of monotonicity can be re-expressed as an inequality of norms (i.e. sandwich α-norms) and a version of the Russo-Dye Theorem (Appendix B) could be used to establish the inequality for $\alpha = 1, \infty$ (which then generalized to $1 < \alpha < \infty$ by the Riesz-Thorin Theorem - Appendix B). [44] then observed that under a more general application of the Russo-Dye Theorem, monotonicity could be extended under the same proof for positive trace non-increasing maps, without requiring the Schwarz property, and we'll revisit this soon. Recently, monotonicity under positive trace-preserving maps for $\alpha \geqslant \frac{1}{2}$ has also been established [23, 34, 35, 7]. Since monotonicity comes with the implication of a loss of distinguishability (and therefore a permanent loss of information), this establishes a notion that positive maps are Markovian [44].

Before we proceed with our proof, let's first remark on the importance of the sandwiched Rényi divergence for the cases that monotonicity is known to hold. [61] proved the following bound on the success probability for any rate R scheme for classical communication over n uses of a quantum channel $\mathcal{N}$ $\forall\alpha \in [1,2)$, noting the importance of the fact $\widetilde{D}_\alpha(\rho||\gamma) \geqslant D_\alpha(\rho||\gamma) := \frac{1}{\alpha-1}\log \mathrm{Tr}(\rho^\alpha\gamma^{1-\alpha}) \forall\alpha \geqslant 1$ via the Araki-Lieb-Thirring inequality (Appendix B), in establishing a stronger bound than using the Petz Rényi divergence:

$$P_{\text{succ}} \leqslant 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{n}\tilde{\chi}_\alpha(\mathcal{N}^n)\right)}, \quad \tilde{\chi}_\alpha(\mathcal{N}) := \max_{\{p_x,\rho_x\}} \min_{\gamma_Q} \widetilde{D}_\alpha\left(\rho_{XQ} \middle\| \rho_X \otimes \gamma_Q\right)$$

($\sum_x p_X(x)|x\rangle\langle x|_X \otimes (\rho_x)_Q$)

Recently [39] gave an operational meaning to the sandwiched Rényi divergence in the case $\alpha \in (\frac{1}{2}, 1)$ by noticing that it gives the exact strong converse exponents of certain quantum tasks like privacy amplification.

Theorem 3.10 (Monotonicity of sandwiched Rényi divergence)**.** *[44, 5]*
Let $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be a positive trace non-increasing map and $\alpha \in (1, \infty)$. Then for any positive $\rho, \gamma \in \mathcal{B}(\mathcal{H})$ (where $\mathcal{H}$ is a finite-dimensional Hilbert space):

$$\widetilde{D}_\alpha(\Phi(\rho)\|\Phi(\gamma)) \leqslant \widetilde{D}_\alpha(\rho\|\gamma)$$

Proof. In this case, $\alpha > 1$, $\operatorname{supp}\rho \nleqslant \operatorname{supp}\gamma \Rightarrow \widetilde{D}_\alpha(\rho\|\gamma) = \infty$, so the inequality holds. Suppose instead $\operatorname{supp}\rho \leqslant \operatorname{supp}\gamma$. By the positivity of Φ, $\operatorname{supp}\Phi(\rho) \leqslant \operatorname{supp}\Phi(\gamma)$, and in this proof the inverses are actually generalised inverses.

Observe $\widetilde{D}_\alpha(\rho\|\gamma) \geqslant \widetilde{D}_\alpha(\Phi(\rho)\|\Phi(\gamma)) \Leftrightarrow \|\Gamma_\gamma^{-1}(\rho)\|_{\alpha,\gamma} \geqslant \|\Gamma_{\Phi(\gamma)}^{-1}(\Phi(\rho))\|_{\alpha,\Phi(\gamma)}$ as $\alpha' \log$ is increasing for $\alpha > 1$.

We define for any linear map $\Phi : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ the induced operator norm w.r.t. the norms $\|\cdot\|_{p,\gamma}$, $\|\cdot\|_{p,\Phi(\gamma)}$ on the input and output spaces respectively, as:

$$\|\Psi\|_{(p,\gamma)\to(p,\Phi(\gamma))} := \sup_{X \in \mathcal{B}(\mathcal{H}),\, \operatorname{supp}X \leqslant \operatorname{supp}\gamma} \frac{\|\Psi(X)\|_{p,\Phi(\gamma)}}{\|X\|_{p,\gamma}}$$

Then $\|\Gamma_{\Phi(\gamma)}^{-1} \circ \Phi(\rho)\|_{\alpha,\Phi(\gamma)} \leqslant \|\Gamma_{\Phi(\gamma)}^{-1} \circ \Phi \circ \Gamma_\gamma\|_{(\alpha,\gamma)\to(\alpha,\Phi(\gamma))} \|\Gamma_\gamma^{-1}(\rho)\|_{\alpha,\gamma}$

$\Psi = \Gamma_{\Phi(\gamma)}^{-1} \circ \Phi \circ \Gamma_\gamma$ is positive on the restriction to states with $\operatorname{supp}\rho \leqslant \operatorname{supp}\gamma$. In order to use the Riesz-Thorin theorem (Appendix B), we use the positive definite perturbation $\gamma_\varepsilon = \gamma + \varepsilon\mathbb{I}$, $\varepsilon > 0$, instead of γ. Note: $\operatorname{supp}(\gamma_\varepsilon) = \mathcal{H}$.

Let $\mathcal{H}' = \operatorname{supp}(\Phi(\gamma_\varepsilon)) = \operatorname{supp}(\Phi(\mathbb{I}))$. Then $\Psi = \Gamma_{\Phi(\gamma_\varepsilon)}^{-1} \circ \Phi \circ \Gamma_{\gamma_\varepsilon} : \mathcal{B}(\mathcal{H}) \mapsto \mathcal{B}(\mathcal{H})$ is a positive map. (as composing with $\Gamma_{\gamma_\varepsilon}$, $\Gamma_{\Phi(\gamma_\varepsilon)}^{-1}$ does not affect the positivity in this case).

Now, considering the Riesz-Thorin theorem (Appendix B),

$\|\Gamma_{\Phi(\gamma_\varepsilon)}^{-1} \circ \Phi \circ \Gamma_{\gamma_\varepsilon}\|_{(p_\theta,\gamma_\varepsilon)\to(p_\theta,\Phi(\gamma_\varepsilon))} \leqslant \|\Gamma_{\Phi(\gamma_\varepsilon)}^{-1} \circ \Phi \circ \Gamma_{\gamma_\varepsilon}\|_{(p_0,\gamma_\varepsilon)\to(p_0,\Phi(\gamma_\varepsilon))}^{1-\theta} \|\Gamma_{\Phi(\gamma_\varepsilon)}^{-1} \circ \Phi \circ \Gamma_{\gamma_\varepsilon}\|_{(p_1,\gamma_\varepsilon)\to(p_1,\Phi(\gamma_\varepsilon))}^{\theta}$

We set $p_0 = 1$, $p_1 = \infty$, $\theta = \alpha' = \frac{\alpha}{\alpha-1}$ so that $p_\theta = \alpha$.

The Russo-Dye theorem (Appendix B) is then employed on $\hat{\Phi}$ to evaluate the RHS. $\hat{\Phi}$ is a positive map and $0 \leqslant \hat{\Phi}(\mathbb{I}) \leqslant \mathbb{I}$ since Φ is positive, trace non-increasing.

This yields:

$$\|\Gamma_{\Phi(\gamma_\varepsilon)}^{-1} \circ \Phi \circ \Gamma_{\gamma_\varepsilon}\|_{(1,\gamma_\varepsilon)\to(1,\Phi(\gamma_\varepsilon))} = \|\Phi\|_{1\to1} = \|\hat{\Phi}\|_{\infty\to\infty} \overset{R-D}{=} \|\hat{\Phi}(\mathbb{I})\|_\infty \leqslant \|\mathbb{I}\|_\infty = 1$$

$$\|\Gamma_{\Phi(\gamma_\varepsilon)}^{-1} \circ \Phi \circ \Gamma_{\gamma_\varepsilon}\|_{(\infty,\gamma_\varepsilon)\to(\infty,\Phi(\gamma_\varepsilon))} \overset{R-D}{=} \|\Gamma_{\Phi(\gamma_\varepsilon)}^{-1} \circ \Phi \circ \Gamma_{\gamma_\varepsilon}(\mathbb{I})\|_\infty = \|\mathbb{I}\|_\infty = 1$$

Then take $\varepsilon \to 0$, and note that the sandwiched Rényi divergence is continuous within the case: $\alpha \in (1, \infty)$ and $\mathrm{supp}(\rho) \leqslant \mathrm{supp}(\gamma)$. This completes the proof.

$\square$

Chapter 4

Reverse-type data processing inequalities for standard f-divergences

4.1 An Introduction to Contraction and Expansion Coefficients

A lot of prior analysis, until [6], was based on the contraction coefficients of a classical or quantum channel, which measure the extent of uniform distinguishability loss induced by the channel. For a classical channel, we consider the set of probability vectors $P_d = \{x \in \mathbb{R}^d_{\geqslant 0} : \sum_i x_i = 1\}$ and sum-zero vectors $S_d = \{v \in \mathbb{R}^d : \sum_i v_i = 0\}$. We can then define for a channel $A : P_d \to P_{d'}$ the *divergence contraction coefficient* w.r.t. the classical f-divergence $D_f^{\mathrm{cl}}(x\|y)$, $x, y \in P_d$, and the *Riemannian contraction coefficient* w.r.t. the χ^2-distribution/Riemannian semi-norm $\phi(x,v)^{\frac{1}{2}} := \sqrt{\sum_i \frac{v_i^2}{x_i}}, x \in P_d, v \in S_d$:

$$\hat{\eta}_f^{\mathrm{cl}}(A) := \sup_{x,y \in P_n} \frac{D_f^{\mathrm{cl}}(Ax\|Ay)}{D_f^{\mathrm{cl}}(x\|y)}, \quad \hat{\eta}^{\mathrm{Riem,cl}}(A) := \sup_{x \in P_n, v \in S} \frac{\phi(Ax, Av)}{\phi(x,v)}.$$

These quantities become considerably rich when the scenario becomes quantum. We have already seen that the Riemannian metric is no longer unique and the semi-norms are

$$\|X\|_{\kappa,\rho} = \sqrt{\langle X, R_\rho^{-1}\kappa(L_\rho R_\rho^{-1})(X)\rangle_{\mathrm{HS}}}$$

for $\kappa \in \mathcal{K}$, $\rho \in \mathcal{D}_d$, $X \in T_\rho \mathcal{D}_d$; when ρ, X are commute, i.e., for some orthonormal basis $\{\psi_i\}_i$, we have $\rho = \sum_i x_i |\psi\rangle\langle\psi_i|$, $X = \sum_i v_i |\psi\rangle\langle\psi_i|$, and all Riemannian semi-norms reduce to $\|X\|^2_{\kappa,\rho} = \mathrm{Tr} X^2 \rho^{-1} = \phi(x,v)$. And, of course, any quantum f-divergence D_f reduces to its corresponding classical f-divergence D_f^{cl} when restricted to commuting input states. The divergence contraction coefficient and Riemannian contraction coefficient respectively for a quantum channel $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ (typically finite-dimensional Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B), \mathcal{D}_d := \mathcal{D}(H_A)$, $\dim \mathcal{H}_A =: d$, are:

$$\hat{\eta}_f(\mathcal{N}) := \sup_{\rho \neq \gamma \in \mathcal{D}_d,\, \mathrm{supp}\,\rho = \mathrm{supp}\,\gamma} \frac{D_f(\mathcal{N}(\rho) \| \mathcal{N}(\gamma))}{D_f(\rho \| \gamma)},$$

$$\hat{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N}) := \sup_{\rho \in \mathcal{D}_d,\, X \in T_\rho \mathcal{D}_d} \frac{\|\mathcal{N}(X)\|^2_{\kappa,\, \mathcal{N}(\rho)}}{\|X\|^2_{\kappa,\rho}}$$

This thesis hopes to develop some of the ideas from [28, 37, 6] to establish a theory of *relative expansion coefficients* that generalise the contraction coefficients and assess uniformly how the output state distinguishability compares between two quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$:

$$\check{\eta}_f(\mathcal{N}, \mathcal{M}) := \inf_{\substack{\rho,\, \gamma \in \mathcal{D}_d \\ \mathrm{supp}\,\rho = \mathrm{supp}\,\gamma}} \frac{D_f(\mathcal{N}(\rho) \| \mathcal{N}(\gamma))}{D_f(\mathcal{M}(\rho) \| \mathcal{M}(\gamma))}$$

$$\check{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N}, \mathcal{M}) := \inf_{\rho \in \mathcal{D}_d,\, X \in T_\rho \mathcal{D}_d} \frac{\|\mathcal{N}(X)\|^2_{\kappa,\, \mathcal{N}(\rho)}}{\|\mathcal{M}(X)\|^2_{\kappa,\, \mathcal{M}(\rho)}}$$

In particular, $\check{\eta}_f(\mathcal{N}) := \check{\eta}_f(\mathcal{N}, \mathrm{id}_{\mathcal{B}(\mathcal{H})})$ and $\check{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N}) := \check{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N}, \mathrm{id}_{\mathcal{B}(\mathcal{H})})$. Generally, if $\mathcal{N} = \mathcal{D} \circ \mathcal{M}$ for some quantum channel $\mathcal{D}$, $\check{\eta}_f(\mathcal{D} \circ \mathcal{M}, \mathcal{M}), \check{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{D} \circ \mathcal{M}, \mathcal{M})$ measure the extent to which the channel $\mathcal{D}$ uniformly preserves the distinguishability between arbitrary states $\mathrm{supp}\,\rho = \mathrm{supp}\,\gamma, \rho, \gamma \in \mathrm{Im}\,\mathcal{D}$. This can be seen, for example, from the definition:

$$D_f(\mathcal{D}(\rho) \| \mathcal{D}(\gamma)) \geqslant \check{\eta}_f(\mathcal{D} \circ \mathcal{M}, \mathcal{M}) D_f(\rho \| \gamma), \quad \forall\, \rho, \gamma \in \mathrm{Im}\,\mathcal{M},\ \mathrm{supp}\,\rho = \mathrm{supp}\,\gamma.$$

Recall that it is the monotonicity of $D_f(\rho \| \gamma)$ and $\|\rho - \gamma\|^2_{\kappa,\rho}$ that make them appropriate as measures of distinguishability between two states $\rho, \gamma \in \mathcal{D}_d$, $\mathrm{supp}\,\rho = \mathrm{supp}\,\gamma$. That is, for all quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$:

$$D_f(\mathcal{N}(\rho) \| \mathcal{N}(\gamma)) \leqslant D_f(\rho \| \gamma), \quad \|\mathcal{N}(X)\|^2_{\kappa, \mathcal{N}(\rho)} \leqslant \|X\|^2_{\kappa,\rho}$$

for all $f \in \mathcal{F}$, $\kappa \in \mathcal{K}$, $\rho, \gamma \in \mathcal{D}_d$, $\operatorname{supp}\rho = \operatorname{supp}\gamma$, $X \in T_\rho \mathcal{D}_d$.

So $\check{\eta}_f(\mathcal{D} \circ \mathcal{M}, \mathcal{M}), \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{D} \circ \mathcal{M}, \mathcal{M}), \hat{\eta}_f(\mathcal{N}), \hat{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}) \in [0,1]$. The first result that we'll show in this section is that for a quantum channel $\mathcal{N} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$, $\dim \mathcal{H}' \leqslant \dim \mathcal{H}$, $\check{\eta}_f(\mathcal{N})$ reduces to an indicator function for the unitary channel $\mathcal{N}(\rho) = U\rho U^\dagger$, $U \in U(d)$.

It was shown in [12, 11] that for classical channels $A : P_d \to P_{d'}$, the contraction coefficients become redundant for operator convex functions $f \in \mathcal{F}$, i.e. $\hat{\eta}_f^{\mathrm{cl}}(A) = \hat{\eta}^{\mathrm{Riem,cl}}(A)\ \forall f \in \mathcal{F}$. This proof naturally applies to give $\check{\eta}_f(\mathcal{N}, \mathcal{M}) = \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$ for quantum-classical channels for all $f \in \mathcal{F}$ so we will see how the proof method works later in Section 4.5. Generally for quantum channels, the coefficients $\hat{\eta}_f^{\mathrm{std}}(\mathcal{N}), \check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M})$ corresponding to the standard f-divergence are the most related to the Riemannian coefficients $\hat{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}), \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$. We usually only have

$$\hat{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}) \leqslant \hat{\eta}_{f_{\mathrm{sym}}}^{\mathrm{std}}(\mathcal{N}) \leqslant \hat{\eta}_f^{\mathrm{std}}(\mathcal{N})$$

for $f \in \mathcal{F}, f_{\mathrm{sym}}(\mathcal{N}) = \frac{f+\tilde{f}}{f''(1)} \in \mathcal{F}_{\mathrm{sym}}, \kappa_f(x) = \frac{f_{\mathrm{sym}}(x)}{(x-1)^2} \in \mathcal{H}$. We'll meet some choices for $f \in \mathcal{F}$ that were shown (or implied) to yield $\hat{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}) = \hat{\eta}_f^{\mathrm{std}}(\mathcal{N})$ for all CPTP $\mathcal{N}$ [28, 6] but [28] also gave some cases where $\hat{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}) < \hat{\eta}_f^{\mathrm{std}}(\mathcal{N})$ for some $f \in \mathcal{F}_{\mathrm{sym}}$, qubit channel $\mathcal{N}$ (see Theorem 4.15). By a similar proof, we also have (see Theorem 4.6)

$$\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_{f_{\mathrm{sym}}}^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}). \tag{4.1}$$

4.2 Some Subtleties

With how the coefficients were introduced, it is natural to be concerned about indeterminate forms, i.e. $\frac{0}{0}$ or $\frac{\infty}{\infty}$. Implicitly, to avoid these issues, we define the coefficients for quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B), \mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ as:

$$\begin{aligned}\check{\eta}_f(\mathcal{N}, \mathcal{M}) &= \sup\{c > 0 : D_f(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\} \geqslant c D_f(\mathcal{M}(\rho)\|\mathcal{M}(\gamma) \\ &\qquad\qquad \forall \rho, \gamma \in \mathcal{D}_d,\ \operatorname{supp}\rho = \operatorname{supp}\gamma\}, \\ \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}) &= \sup\{c > 0 : \|\mathcal{N}(X)\|^2_{\kappa, \mathcal{N}(\rho)} \geqslant c \|\mathcal{M}(X)\|^2_{\kappa, \mathcal{M}(\rho)}\ \forall \rho \in \mathcal{D}_d, X \in T_\rho \mathcal{D}_d\}\end{aligned}$$

In effect, we are taking '$\frac{0}{0} = \frac{\infty}{\infty} = 1$'.

The following alternative definitions for the relative divergence expansion coefficients will sometimes be important, so we will mention them here:

Theorem 4.1. *Let $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B), \mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ be quantum channels and let $f \in \mathcal{F}$, $f(0^+) < \infty$. Then, again using $\mathcal{D}_d := \mathcal{D}(\mathcal{H}_A)$:*

$$\check{\eta}_f(\mathcal{N}, \mathcal{M}) = \inf_{\substack{\rho \neq \gamma \in \mathcal{D}_d \\ \operatorname{supp}\rho \leqslant \operatorname{supp}\gamma}} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho) \| \mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho) \| \mathcal{M}(\gamma))}.$$

Proof. By considering the continuity of standard f-divergences (Proposition 3.2) and the continuity of $\mathcal{N}, \mathcal{M}$: $D_f^{\mathrm{std}}(\mathcal{N}(\rho)||\mathcal{N}(\gamma))$ and $D_f^{\mathrm{std}}(\mathcal{M}(\rho)||\mathcal{M}(\gamma))$ are continuous w.r.t. ρ. First observe that because of the larger optimisation region:

$$\inf_{\substack{\rho \neq \gamma \in \mathcal{D}_d \\ \operatorname{supp}\rho \leqslant \operatorname{supp}\gamma}} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho)||\mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho)||\mathcal{M}(\gamma))} \leqslant \inf_{\substack{\rho \neq \gamma \in \mathcal{D}_d \\ \operatorname{supp}\rho = \operatorname{supp}\gamma}} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho)||\mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho)||\mathcal{M}(\gamma))}$$

Given $\rho \leqslant \gamma$, let $\rho_\varepsilon := (1-\varepsilon)\rho + \varepsilon\gamma$, $\varepsilon \in (0,1)$, so that $\operatorname{supp}\rho_\varepsilon = \operatorname{supp}\gamma$. Then we also have:

$$\begin{aligned}\inf_{\substack{\rho \neq \gamma \in \mathcal{D}_d \\ \operatorname{supp}\rho \leqslant \operatorname{supp}\gamma}} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho) \| \mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho) \| \mathcal{M}(\gamma))} &= \inf_{\substack{\rho \neq \gamma \in \mathcal{D}_d \\ \operatorname{supp}\rho \leqslant \operatorname{supp}\gamma}} \lim_{\varepsilon \to 0^+} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho_\varepsilon) \| \mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho_\varepsilon) \| \mathcal{M}(\gamma))} \\ &\geqslant \inf_{\substack{\rho \neq \gamma \in \mathcal{D}_d \\ \operatorname{supp}\rho = \operatorname{supp}\gamma}} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho) \| \mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho) \| \mathcal{M}(\gamma))}.\end{aligned}$$

□

By a very similar argument, for bounded standard f-divergences:

Theorem 4.2. *Let $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B), \mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ be quantum channels and let $f \in \mathcal{F}$, $f(0^+), f'(\infty) < \infty$. Then, again using $\mathcal{D}_d := \mathcal{D}(\mathcal{H}_A)$:*

$$\check{\eta}_f(\mathcal{N}, \mathcal{M}) = \inf_{\rho \neq \gamma \in \mathcal{D}_d} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho) \| \mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho) \| \mathcal{M}(\gamma))}.$$

4.3 No Reverse Data Processing over all States

The role of this section is to review and extend the result in [6] that the non-relative divergence expansion coefficient of the relative entropy for a channel $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\dim \mathcal{H}_A \leqslant \dim \mathcal{H}_B$ is zero for all non-unitary channels; in fact all of the non-relative expansion coefficients from Section 4.1 turn out to be the same. The trick is to find a family of states $\rho_\varepsilon \in \mathcal{D}(\mathcal{H}_A)$, continuously parametrized by $\varepsilon \in \mathbb{R}$, $|\varepsilon| << 1$, such that $D_f(\rho_0 \| \rho_\varepsilon) = \Theta(\varepsilon)$, $\|\rho_\varepsilon - \rho_0\|^2_{\kappa_f,\, \rho_\varepsilon} = \Theta(\varepsilon)$ while $D(\mathcal{N}(\rho_0)||\mathcal{N}(\rho_\varepsilon)) = o(\varepsilon)$, $\|\mathcal{N}(\rho_\varepsilon - \rho_0)\|^2_{\kappa_f,\mathcal{N}(\rho_\varepsilon)} = o(\varepsilon)$ in the non-trivial cases where $\mathcal{N}$ is neither a unitary channel nor a replacer channel. To do this, we use the following two results:

Lemma 4.1. *[13, Theorem 3.1] If a quantum channel $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is purity-preserving, i.e. it maps any pure state into a pure state, then $\mathcal{N}$ must be either an isometric embedding $\mathcal{N}(\rho) = V\rho V^\dagger$, $V^\dagger V = \mathbb{1}_A$, or a replacer channel $\mathcal{N}(\rho) = \operatorname{Tr}\rho\, |\varphi\rangle\langle\varphi|$ for some state $|\varphi\rangle$.*

Lemma 4.2. *[6, Theorem 3.1] If a quantum channel $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\dim \mathcal{H}_B =: d_B \leqslant \dim \mathcal{H}_A =: d_A$, is not purity-preserving, then there exists an orthogonal projection $P_A \in \mathcal{B}(\mathcal{H}_A)$, $\operatorname{rk} P_A \leqslant d_A - 1$, and a pure state $|\psi\rangle$ such that*

$$\operatorname{supp}\mathcal{N}(|\psi\rangle\langle\psi|) \leqslant \operatorname{supp}\mathcal{N}(P_A)$$

Proof. See Appendix D. ☐

First, we shall essentially modify the strategy of the proof of [6] for our subtly different notion of relative expansion coefficient, while spelling out the sufficient conditions in more detail. In fact, this result applies to all of the quantum f-divergences that we have met:

Theorem 4.3 (No Divergence-based Reverse DPI on All States)**.**
Suppose we are given a quantum channel $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, with $\dim \mathcal{H}_B =: d_B \leqslant \dim \mathcal{H}_A =: d_A$. For any monotone quantum f-divergence D_f, for operator convex $f \in \mathcal{F}$, that extends the classical f-divergence D_f^{cl}:

$$\begin{aligned}\check{\eta}_f^{\mathrm{std}}(\mathcal{N}) &= \mathbb{1}\{\mathcal{N} \textit{ is a unitary channel}\} \\ &= \begin{cases} 1 & \textit{if } d := d_A = d_B \textit{ and } \mathcal{N}(\rho) = U\rho U^\dagger,\, U \in U(d) \\ 0 & \textit{otherwise} \end{cases}\end{aligned}$$

Proof.
Case 1 ($\mathcal{N}(\rho) = U\rho U^\dagger$ for some unitary U):

For any $\rho \neq \gamma \in \mathcal{D}(\mathcal{H}_A)$, $D_f(\mathcal{N}(\rho)||\mathcal{N}(\gamma)) = D_f(\rho||\gamma) > 0$ by unitary invariance of D_f.

As a result, $\check{\eta}_f(\mathcal{N}) = 1$.

Case 2 ($\mathcal{N}(\rho) = \operatorname{Tr}\rho\, |\varphi\rangle\langle\varphi|$ for some pure state $|\varphi\rangle\langle\varphi| \in \mathcal{D}(\mathcal{H}_B)$):

For any $\rho \neq \gamma \in \mathcal{D}(\mathcal{H}_A)$, $D_f(\mathcal{N}(\rho)||\mathcal{N}(\gamma)) = D_f(|\varphi\rangle\langle\varphi| \parallel |\varphi\rangle\langle\varphi|) = 0$ and $D_f(\rho||\gamma) > 0$.

As a result, $\check{\eta}_f(\mathcal{N}) = 0$.

Case 3 ($\mathcal{N}$ is not purity-preserving):
By Lemma 4.2, we can choose states

$$\bar{\rho} = \frac{P_A}{\mathrm{rk} P_A}, \quad \rho_\varepsilon = (1 - o(\varepsilon))\bar{\rho} + o(\varepsilon)\, |\psi\rangle\langle\psi|\,, \quad \gamma_\varepsilon = (1-\varepsilon)\bar{\rho} + \varepsilon\, |\psi\rangle\langle\psi| \in \mathcal{D}(\mathcal{H}_A),$$

for some orthogonal projection $P_A \in \mathcal{B}(\mathcal{H}_A)$ and pure state $|\psi\rangle\langle\psi| \in \mathcal{D}(\mathcal{H}_A)$ satisfying $|\psi\rangle\langle\psi| \perp \bar{\rho}$ but $\operatorname{supp}\mathcal{N}(|\psi\rangle\langle\psi|) \leqslant \operatorname{supp}\mathcal{N}(\bar{\rho})$. Here, $\varepsilon \in (0,1)$ is small.

Via the common eigendecomposition, let $p_\varepsilon = (\frac{1-o(\varepsilon)}{\mathrm{rk}\, P_A}, ..., \frac{1-o(\varepsilon)}{\mathrm{rk}\, P_A}, o(\varepsilon))$, $q_\varepsilon = (\frac{1-\varepsilon}{\mathrm{rk}\, P_A}, ..., \frac{1-\varepsilon}{\mathrm{rk}\, P_A}, \varepsilon)$, then:

$$\begin{aligned} D_f(\rho_\varepsilon \| \gamma_\varepsilon) = D_f^{\mathrm{cl}}(p_\varepsilon \| q_\varepsilon) &= \frac{1-\varepsilon}{\mathrm{rk}\, P_A}\, f\left(\frac{\frac{1-o(\varepsilon)}{\mathrm{rk}\, P_A}}{\frac{1-\varepsilon}{\mathrm{rk}\, P_A}}\right) \cdot \mathrm{rk} P_A + \varepsilon\, f\left(\frac{o(\varepsilon)}{\varepsilon}\right) \\ &= (1-\varepsilon) f\left(\frac{1-o(\varepsilon)}{1-\varepsilon}\right) + \varepsilon f(o(1)) \end{aligned}$$

Observe that by the strict convexity at 1:

$$\begin{aligned} \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} D_f(\rho_\varepsilon \| \gamma_\varepsilon) &= \lim_{\varepsilon \to 0^+} \frac{1-\varepsilon}{\varepsilon} f\left(\frac{1-o(\varepsilon)}{1-\varepsilon}\right) + \lim_{\varepsilon \to 0^+} f(o(1)) \\ &= f'(1) + f(0^+) = f'(1) - \frac{f(1) - f(0^+)}{1-0} > 0, \end{aligned}$$

Thus, we have demonstrated the required first order behaviour. Now, to deduce the second order behaviour of $D_f(\mathcal{N}(\rho_\varepsilon) \| \mathcal{N}(\gamma_\varepsilon))$, we consider the maximal f-divergence.

By assumption, D_f is a monotone quantum f-divergence, which means by the maximality of the maximal f-divergence (Theorem 3.6) that

$$D_f(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon)) \leqslant \widehat{D}_f(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon))$$

Since $\mathcal{N}(|\psi\rangle\langle\psi|) \in T_{\mathcal{N}(\rho)}\mathcal{D}(\mathcal{H}_B)$, $\operatorname{supp}\mathcal{N}(\rho_\varepsilon) = \operatorname{supp}\mathcal{N}(\gamma_\varepsilon)$, and we can write the following expression (recall Definition 3.2):

$$\widehat{D}_f(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon)) = \operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{1/2} f(\operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{-1/2}\operatorname{Tr}\mathcal{N}(\rho_\varepsilon)\operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{-1/2})\operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{1/2}$$

Observing that $\mathcal{N}(\rho_\varepsilon) = \mathcal{N}(\gamma_\varepsilon) + \varepsilon\mathcal{N}(|\psi\rangle\langle\psi| - \bar{\rho}) + o(\varepsilon)$ and using the Taylor expansion of $f(x)$ about $x = 1$:

$$\begin{aligned}
\widehat{D}_f(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon)) &= \operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{1/2} f(\mathcal{N}(\gamma_\varepsilon)^{-1/2}\mathcal{N}(\rho_\varepsilon)\mathcal{N}(\gamma_\varepsilon)^{-1/2})\mathcal{N}(\gamma_\varepsilon)^{1/2} \\
&= f'(1)\operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{1/2}(\mathcal{N}(\gamma_\varepsilon)^{-1/2}\mathcal{N}(\rho_\varepsilon)\mathcal{N}(\gamma_\varepsilon)^{-1/2} - I)\mathcal{N}(\gamma_\varepsilon)^{1/2} \\
&\quad + \frac{f''(1)}{2}\operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{1/2}(\mathcal{N}(\gamma_\varepsilon)^{-1/2}\mathcal{N}(\rho_\varepsilon)\mathcal{N}(\gamma_\varepsilon)^{-1/2} - I)^2\mathcal{N}(\gamma_\varepsilon)^{1/2} \\
&\quad + o(\varepsilon^2) \\
&= \frac{f''(1)}{2}\operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{1/2}(\mathcal{N}(\gamma_\varepsilon)^{-1/2}\mathcal{N}(\rho_\varepsilon)\mathcal{N}(\gamma_\varepsilon)^{-1/2} - I)^2\mathcal{N}(\gamma_\varepsilon)^{1/2} \\
&\quad + o(\varepsilon^2) \\
&= \frac{\varepsilon^2}{2} f''(1)\operatorname{Tr}\mathcal{N}(\gamma_\varepsilon)^{1/2}(\mathcal{N}(\gamma_\varepsilon)^{-1/2}\mathcal{N}(|\psi\rangle\langle\psi| - \bar{\rho})\mathcal{N}(\gamma_\varepsilon)^{-1/2})^2\mathcal{N}(\gamma_\varepsilon)^{1/2} \\
&\quad + o(\varepsilon^2) \\
&= o(\varepsilon)
\end{aligned}$$

The final equality certainly holds, because $\mathcal{N}(|\psi\rangle\langle\psi| - \bar{\rho}) \in T_{\mathcal{N}(\gamma_\varepsilon)}\mathcal{D}(\mathcal{H}_B)$.
Therefore, $D_f(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon)) \leqslant \widehat{D}_f(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon)) = o(\varepsilon)$.

As a result,

$$\begin{aligned}
0 \leqslant \check{\eta}_f(\mathcal{N}) := \inf_{\substack{\rho\neq\gamma\in\mathcal{D}(\mathcal{H}_A),\\ \operatorname{supp}\rho=\operatorname{supp}\gamma}} \frac{D_f\big(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\big)}{D_f(\rho\|\gamma)} &\leqslant \lim_{\varepsilon\to 0^+}\frac{D_f\big(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon)\big)}{D_f(\rho_\varepsilon\|\gamma_\varepsilon)} \\
&= \lim_{\varepsilon\to 0^+}\frac{\frac{D_f\left(\mathcal{N}(\rho_\varepsilon)\|\mathcal{N}(\gamma_\varepsilon)\right)}{\varepsilon}}{\frac{D_f(\rho_\varepsilon\|\gamma_\varepsilon)}{\varepsilon}} = 0
\end{aligned}$$

Thus, $\check{\eta}_f(\mathcal{N}) = 0$. □

Corollary 4.1. *All of the monotone quantum f-divergences we have met (standard, maximal, measured, Riemannian semi-norms) have a trivial non-relative expansion coefficient for $f \in \mathcal{F}$. Similarly, Petz and sandwiched Rényi divergences also have a trivial non-relative expansion coefficient.*

We interpret this theorem as telling us that non-unitary channels whose input space dimension is no larger than the output space dimension do not uniformly preserve information about the distinguishability of all states, w.r.t. the quantum f-divergences and Riemannian semi-norms. Especially when one compares this behavior with expansion coefficients based on norms attributed to $\mathcal{B}(\mathcal{H}_A)$ and $\mathcal{B}(\mathcal{H}_B)$ (see Section 4.4), it is interesting that we observe this for injective quantum channels (that map states uniquely into other states). One of the new results in this thesis is that for the large class of primitive (necessarily $\mathcal{H}_A = \mathcal{H}_B$), injective quantum channels, and n sufficiently large (see Theorem 4.14),

$$\check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}^{n+1}, \mathcal{N}^n) = \inf_{\rho \in \mathcal{N}^n(\mathcal{D}_d), X \in T_\rho \mathcal{N}^n(\mathcal{D}_d)} \frac{\|\mathcal{N}(X)\|^2_{\kappa,\mathcal{N}(\rho)}}{\|X\|^2_{\kappa,\rho}} > 0$$

i.e. a reverse DPI holds. This means that the strict positivity of the Riemannian expansion coefficient (and thus, sometimes the corresponding standard f-divergence expansion coefficient, when there is equivalence) is sensitive to the set of input states of $\mathcal{N}$. In these cases, information about the distinguishability of states is uniformly preserved over a restricted set.

4.4 Other Non-Relative Expansion Coefficients can be Positive

While not a primary focus in this thesis, another important measure of distinguishability is to take the norm of the difference between two states. We can consider $\mathcal{N} : (\mathcal{B}(\mathcal{H}_A), \|\cdot\|_A) \to (\mathcal{B}(\mathcal{H}_B), \|\cdot\|_B)$, $\mathcal{M} : (\mathcal{B}(\mathcal{H}_A), \|\cdot\|_A) \to (\mathcal{B}(\mathcal{H}'_B), \|\cdot\|_{B'})$, bearing in mind that any norm we assign to a finite-dimensional Banach space is equivalent to any other. This allows us to define the following contraction coefficient and relative expansion coefficient (again, $\mathcal{D}_d := \mathcal{D}(\mathcal{H}_A)$):

$$\hat{\eta}^{\|\cdot\|_B,\|\cdot\|_A}(\mathcal{N}) := \sup_{\rho\neq\gamma\in\mathcal{D}_d} \frac{\|\mathcal{N}(\rho)-\mathcal{N}(\gamma)\|_B}{\|\rho-\gamma\|_A}$$

$$\check{\eta}^{\|\cdot\|_B,\|\cdot\|_{B'}}(\mathcal{N},\mathcal{M}) := \inf_{\rho\neq\gamma\in\mathcal{D}_d} \frac{\|\mathcal{N}(\rho)-\mathcal{N}(\gamma)\|_B}{\|\mathcal{M}(\rho)-\mathcal{M}(\gamma)\|_{B'}}$$

The special cases $\hat{\eta}_{\mathrm{Tr}}(\mathcal{N}) := \hat{\eta}^{\|\cdot\|_1,\|\cdot\|_1}(\mathcal{N})$, $\check{\eta}_{\mathrm{Tr}}(\mathcal{N},\mathcal{M}) := \check{\eta}^{\|\cdot\|_1,\|\cdot\|_1}(\mathcal{N},\mathcal{M})$ and $\hat{\eta}_2(\mathcal{N}) := \hat{\eta}^{\|\cdot\|_2,\|\cdot\|_2}(\mathcal{N})$, $\check{\eta}_2(\mathcal{N},\mathcal{M}) := \check{\eta}^{\|\cdot\|_2,\|\cdot\|_2}(\mathcal{N},\mathcal{M})$ are the most relevant to us, because the relationships between $\hat{\eta}_{\mathrm{Tr}}, \hat{\eta}_2$, $\hat{\eta}^{\mathrm{Riem}}$ have been studied in the past[37, 59, 28, 6]. Namely, for any $\kappa \in \mathcal{K}$: $\hat{\eta}_{\mathrm{Tr}}(\mathcal{N}) \leqslant \sqrt{\hat{\eta}^{\mathrm{Riem}}_{\kappa}(\mathcal{N})}$ for any channel $\mathcal{N}$[59, 28], $\hat{\eta}_2(\mathcal{N}) \leqslant \sqrt{\hat{\eta}^{\mathrm{Riem}}(\mathcal{N})}$, for any unital channel $\mathcal{N}$ [37]. Recently, $\hat{\eta}_{\mathrm{Tr}}(\mathcal{N})^2 \leqslant \hat{\eta}^{\mathrm{std}}_{\frac{\log x}{x-1}}(\mathcal{N}) \leqslant \hat{\eta}_{\mathrm{Tr}}(\mathcal{N})$ has also been established [6, 31].

It can be helpful to note that $\hat{\eta}^{\|\cdot\|_B,\|\cdot\|_A}(\mathcal{N})$ defines an operator norm on $\mathcal{N}$ as an operator that maps between Banach spaces of traceless matrices, so this relationship informs us that

$$\hat{\eta}^{\mathrm{std}}_{x\log x}(\mathcal{N}) \equiv \hat{\eta}^{\mathrm{Riem}}_{\frac{\log x}{x-1}}(\mathcal{N}) > 0$$

for all non-replacer quantum channels $\mathcal{N}$. A similar relationship, where $\check{\eta}^{\|\cdot\|_B,\|\cdot\|_A}$ is used to lower bound $\check{\eta}^{\mathrm{std}}_f(\mathcal{N})$ or $\check{\eta}^{\mathrm{Riem}}_f(\mathcal{N})$, $f \in \mathcal{F}, \kappa \in \mathcal{K}$, is not possible in general by Theorem 4.3 since

$$\check{\eta}^{\|\cdot\|_B,\|\cdot\|_{B'}}(\mathcal{N},\mathcal{M}) > 0$$

if $\mathcal{N}$ is injective, even when $\mathcal{M} = \mathrm{id}_{\mathcal{B}(\mathcal{H})}$:

Proposition 4.1. *(Norm-Based Relative Expansion Coefficients are Attained)*
$\exists$ traceless Hermitian $X \in \mathcal{B}(\mathcal{H}_A)\backslash\{0\}$ such that for two quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{N}$ is injective and $\mathcal{M}$ is not the replacer channel (otherwise pick any X):

$$\check{\eta}^{\|\cdot\|_B,\|\cdot\|_{B'}}(\mathcal{N},\mathcal{M}) = \frac{\|\mathcal{N}(X)\|_B}{\|\mathcal{M}(X)\|_{B'}} > 0$$

Proof. Observe that for any $\rho \neq \gamma \in \mathcal{D}_d$, $X = \rho - \gamma \neq 0$ is a traceless, Hermitian operator in $\mathcal{B}(\mathcal{H}_A)$. Conversely, for any traceless Hermitian operator $X \in \mathcal{B}(\mathcal{H}_A)$, there exists $\rho \neq \gamma \in \mathcal{D}_d$ such that $X = \varepsilon(\rho - \gamma)$ for $\varepsilon > 0$ sufficiently small.

As a result,

$$\check{\eta}^{\|\cdot\|_B,\|\cdot\|_{B'}}(\mathcal{N},\mathcal{M}) = \inf_{\substack{X\in\mathcal{B}(\mathcal{H}_A),\\ \mathrm{Tr}X=0,\,\|X\|_1=1}} \frac{\|\mathcal{N}(X)\|_B}{\|\mathcal{M}(X)\|_{B'}},$$

since rescaling X has no effect on $\frac{\|\mathcal{N}(X)\|_B}{\|\mathcal{M}(X)\|_{B'}}$.

There exists some $X' \in \mathcal{B}(\mathcal{H}_A)$ with $\mathrm{Tr}\,X' = 0$ and $\|X'\|_1 = 1$, such that

$$\frac{\|\mathcal{N}(X')\|_B}{\|\mathcal{M}(X')\|_{B'}} < \infty.$$

So we define the set

$$S := \left\{ X \in \mathcal{B}(\mathcal{H}_A)\backslash\{0\} : \mathrm{Tr}\,X = 0,\ \|X\|_1 = 1,\ \frac{\|\mathcal{N}(X)\|_B}{\|\mathcal{M}(X)\|_{B'}} \leqslant \frac{\|\mathcal{N}(X')\|_B}{\|\mathcal{M}(X')\|_{B'}} \right\}.$$

Then

$$\check{\eta}^{\|\cdot\|_B,\|\cdot\|_{B'}}(\mathcal{N},\mathcal{M}) = \inf_{X\in S} \frac{\|\mathcal{N}(X)\|_B}{\|\mathcal{M}(X)\|_{B'}},$$

which is an infimum of a real-valued continuous function over a compact set.

□

Corollary 4.2. *(Norm-Based Relative Expansion Coefficients are Positive) Suppose we are given two quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ and $\mathcal{N}$ is injective, then*

$$\check{\eta}^{\|\cdot\|_B,\|\cdot\|_{B'}}(\mathcal{N},\mathcal{M}) > 0.$$

However, we can establish an upper bound using the equivalence between norms on $\mathrm{Im}\,\mathcal{M}$, $\mathrm{Im}\,\mathcal{N}$ when $\mathcal{N},\mathcal{M} \in \mathcal{B}(\mathcal{B}(\mathcal{H}_A))$, $\mathcal{D}_d := \mathcal{D}(\mathcal{H}_A)$, $d = \dim\mathcal{H}_A$, so $\exists\alpha,\beta > 0$ such that $\forall$ traceless Hermitian $X \in \mathcal{B}(\mathcal{H}_A)$,

$$\|\mathcal{N}(X)\|^2_{\kappa,\mathcal{N}(\mathrm{I}/d)} \leqslant \alpha\|\mathcal{N}(X)\|^2_A, \quad \|\mathcal{M}(X)\|^2_{\kappa,\mathcal{M}(\mathrm{I}/d)} \geqslant \beta\|\mathcal{M}(X)\|^2_A$$

As a result,

$$\begin{aligned}\check{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N},\mathcal{M}) &\leqslant \inf_{X\in T_{I/d}D_d} \frac{\alpha\|\mathcal{N}(X)\|^2_A}{\beta\|\mathcal{M}(X)\|^2_A} \\ &\leqslant \frac{\alpha}{\beta}\hat{\eta}^{\|\cdot\|_A,\|\cdot\|_A}(\mathcal{N},\mathcal{M})^2 \\ &= c\hat{\eta}^{\|\cdot\|_A,\|\cdot\|_A}(\mathcal{N},\mathcal{M})^2\end{aligned}$$

for some $c = \frac{\alpha}{\beta} > 0$. In particular, since $\|\cdot\|^2_{\kappa,\mathrm{I}/d} = d\|\cdot\|_2^2$, we have

$$\check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_2(\mathcal{N}, \mathcal{M})^2$$

when $\mathcal{N}, \mathcal{M}$ are unital channels. Moreover, when $\mathcal{M} = \mathrm{id}_{\mathcal{B}(\mathcal{H})}$ and $\mathcal{N}$ is injective, we have

$$\check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}) = 0 < \check{\eta}_2(\mathcal{N}, \mathcal{M}),$$

so equality does not generally hold. We will see later that a converse holds when $\mathcal{N}(\mathcal{D}_d)$ and $\mathcal{M}(\mathcal{D}_d)$ consist of only full rank states.

4.5 The Impact of Classical Output

We will now see that when we consider $\hat{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$, $\hat{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M})$ or $\check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$, $\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M})$ where at least one of $\mathcal{N}$ and $\mathcal{M}$ is a CQ or QC channel, the properties of these contraction and relative expansion coefficients simplify. As the heading suggests, for relative expansion coefficients, QC channels are more interesting, and our focus will be on them. The main purpose of this section will be to review the work by [28] on CQ and QC channels, and modify their results in the context of relative expansion coefficients, and better understand the differences with comparing fully quantum channels.

A quantum channel $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ is called *quantum-classical (QC)* if $\mathrm{Im}\,\mathcal{N}$ is a commutative subalgebra of $\mathcal{B}(\mathcal{H}_B)$ and *classical-quantum (CQ)* if $\mathrm{Im}\,\hat{\mathcal{N}}$ (where $\hat{\mathcal{N}}$ is the adjoint) is in a commutative subalgebra of $\mathcal{B}(\mathcal{H}_A)$. As a result, $\mathcal{N}$ is QC if there is an orthonormal basis $\{|\psi\rangle\}_{1=1}^{d_B}$ of $\mathcal{H}_B$ and a POVM $\{F_i\}_{i=1}^{d_B} \subseteq \mathcal{B}(\mathcal{H}_A)$ such that

$$\mathcal{N}(\rho) = \sum_i (\mathrm{Tr}\, F_i \rho) \,|\psi_i\rangle\langle\psi_i|$$

and $\mathcal{N}$ is CQ iff there is an orthonormal basis $|\phi_i\rangle_{i=1}^{d_A}$ of $\mathcal{H}_A$ and density matrices $\{\gamma_i\}_{i=1}^{d_A} \subseteq \mathcal{D}(\mathcal{H}_B)$ such that

$$\mathcal{N}(\rho) = \sum_i (\mathrm{Tr}\,|\phi_i\rangle\langle\phi_i|\ \rho)\gamma_i$$

Observe that if $\mathcal{N}$ is QC, $\|\mathcal{N}(X)\|^2_{\kappa,\mathcal{N}(\rho)} = \mathrm{Tr}\,\mathcal{N}(X)^2\mathcal{N}(\rho)^{-1} = \|\mathcal{N}(X)\|^2_{\kappa_{\max},\mathcal{N}(\rho)}$ for all $\kappa \in \mathcal{K}$. This causes the Riemannian relative expansion coefficients to completely lose their variation in κ when $\mathcal{N}, \mathcal{M}$ are QC channels:

$$\check{\eta}_\kappa^{\text{Riem}}(\mathcal{N}, \mathcal{M}) = \inf_{\substack{\rho \in \mathcal{D}_d \\ X \in T_\rho \mathcal{D}_d}} \frac{\operatorname{Tr} \mathcal{N}(X)^2 \mathcal{N}(\rho)^{-1}}{\operatorname{Tr} \mathcal{M}(X)^2 \mathcal{M}(\rho)^{-1}} = \check{\eta}_{\kappa_{\max}}(\mathcal{N}, \mathcal{M})$$

for all $\kappa \in \mathcal{K}$.

In fact, we'll soon see that the breakdown in variation is much more severe than this, as $\check{\eta}_\kappa^{\text{Riem}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_f^{\text{std}}(\mathcal{N}, \mathcal{M})$ for all $\kappa \in \mathcal{K}$, $f \in \mathcal{F}$. There turns out to be a breakdown even when only one of $\mathcal{N}, \mathcal{M}$ are QC, adapting from [28, Proposition 5.5]:

Proposition 4.2 ($\hat{\eta}_\kappa^{\text{Riem}}(\mathcal{N})$, $\check{\eta}_\kappa^{\text{Riem}}(\mathcal{N}, \mathcal{M})$, $\check{\eta}_\kappa^{\text{Riem}}(\mathcal{M}, \mathcal{N})$ are monotonic in κ for QC $\mathcal{N}$)**.** *Let $\kappa_1, \kappa_2 \in \mathcal{K}$ be such that $\kappa_1(x) \leqslant \kappa_2(x)$ for all $x > 0$, and let $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be quantum channels. If $\mathcal{N}$ is a QC channel, then*

$$\begin{aligned} \hat{\eta}_{\kappa_1}^{\text{Riem}}(\mathcal{N}) &\geqslant \hat{\eta}_{\kappa_2}^{\text{Riem}}(\mathcal{N}), \\ \check{\eta}_{\kappa_1}^{\text{Riem}}(\mathcal{N}, \mathcal{M}) &\geqslant \check{\eta}_{\kappa_2}^{\text{Riem}}(\mathcal{N}, \mathcal{M}), \\ \check{\eta}_{\kappa_1}^{\text{Riem}}(\mathcal{M}, \mathcal{N}) &\leqslant \check{\eta}_{\kappa_2}^{\text{Riem}}(\mathcal{M}, \mathcal{N}). \end{aligned}$$

Proof. Observe that

$$\|\mathcal{N}(X)\|_{\kappa_1, \mathcal{N}(\rho)}^2 = \|\mathcal{N}(X)\|_{\kappa_2, \mathcal{N}(\rho)}^2 \quad \text{for all } \rho \in \mathcal{D}_d,\ X \in T_\rho \mathcal{D}_d,$$

while

$$\|\mathcal{M}(X)\|_{\kappa_1, \mathcal{M}(\rho)}^2 \leqslant \|\mathcal{M}(X)\|_{\kappa_2, \mathcal{M}(\rho)}^2 \quad \text{for all } \rho \in \mathcal{D}_d,\ X \in T_\rho \mathcal{D}_d,$$

since $\Omega_{\mathcal{M}(\rho)}^{\kappa_1} \leqslant \Omega_{\mathcal{M}(\rho)}^{\kappa_2}$, including the case when $\mathcal{M} = \operatorname{id}_{\mathcal{B}(\mathcal{H}_A)}$.

The result then follows immediately from the definitions. □

If instead $\mathcal{N}$ is a CQ channel, we can choose a subalgebra $\mathcal{A}$ of $\mathcal{B}(\mathcal{H}_A)$ that includes $\operatorname{Im} \hat{\mathcal{N}}$ and a trace-preserving conditional expectation (i.e., an $\mathcal{A} - \mathcal{A}$ bilinear, unital projection) $\mathcal{E} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{A}$ [28]. $\hat{\mathcal{E}}$ is then the inclusion $\mathcal{A} \hookrightarrow \mathcal{B}(\mathcal{H}_A)$ and so $\hat{\mathcal{N}} = \hat{\mathcal{E}} \hat{\mathcal{N}} \implies \mathcal{N} = \mathcal{N} \mathcal{E}$ [28]. The contraction and relative expansion coefficients thus reduce when $\mathcal{N}$ is a CQ channel and $\mathcal{M}$ might not be, as we take the extremum over states in $\mathcal{D}_d \cap \mathcal{A}$ and traceless Hermitian operators in $T_\rho \mathcal{D}_d \cap \mathcal{A}$. For example:

$$\hat{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}) = \sup_{\substack{\rho \in D_d \cap \mathcal{A} \\ X \in T_\rho D_d \cap \mathcal{A}}} \frac{\|\mathcal{N}(X)\|^2_{\kappa, \mathcal{N}(\rho)}}{\operatorname{Tr} X^2 \rho^{-1}},$$

$$\check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{M}, \mathcal{N}) = \inf_{\substack{\rho \in D_d \cap \mathcal{A} \\ X \in T_\rho D_d \cap \mathcal{A}}} \frac{\|\mathcal{M}(X)\|^2_{\kappa, \mathcal{N}(\rho)}}{\|\mathcal{N}(X)\|^2_{\kappa, \mathcal{N}(\rho)}}.$$

CQ channels also cause a breakdown:

Proposition 4.3 ($\hat{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N})$ is monotone in κ for CQ channels $\mathcal{N}$)**.** *[28, Proposition 5.5]. Let $\kappa_1, \kappa_2 \in \mathcal{K}$ be such that $\kappa_1(x) \leqslant \kappa_2(x)$ for all $x > 0$, and let $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be a CQ channel. Then*

$$\hat{\eta}_{\kappa_1}^{\mathrm{Riem}}(\mathcal{N}) \leqslant \hat{\eta}_{\kappa_2}^{\mathrm{Riem}}(\mathcal{N}).$$

Proof. Essentially the same as Proposition 4.2.

We will note that the proof of [28, Theorem 5.7] uses a slightly different definition of the contraction coefficients, namely they optimise over $\rho \neq \gamma \in \mathcal{D}(\mathcal{H}_A)$, ρ, γ full rank. As a result, their proof requires that $\hat{\eta}_f^{\mathrm{std}}$ optimises only over $\rho \neq \gamma \in \mathcal{D}_d$, $\operatorname{supp} \rho = \operatorname{supp} \gamma$. Because of our definitions of the relative expansion coefficients, the proof applies to our setting only when $f(0^+) < \infty$ by Theorem 4.1. The modified theorem and essentially identical proof are given below [28]:

Theorem 4.4. *Let $\kappa \in \mathcal{K}$ and $f \in \mathcal{F}_{\mathrm{sym}}$ be related by $f(x) = (x-1)^2 \kappa(x)$, and suppose that $f(0^+) < \infty$. Let $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\quad \mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$ be quantum channels. If $\mathcal{M}$ is a QC channel, then*

$$\hat{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{M}) \leqslant \hat{\eta}_f^{\mathrm{std}}(\mathcal{M}) \leqslant \hat{\eta}_{\min}^{\mathrm{Riem}}(\mathcal{M}),$$

and

$$\check{\eta}_{\min}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}).$$

Proof. Let $f \in \mathcal{F}_{\mathrm{sym}}$, $\kappa \in \mathcal{K}$ be such that $f(x) = (x-1)^2 \kappa(x)$ (i.e., $\kappa = \kappa_f$).

Define also, $\kappa_s(x) := \frac{1+s}{2}(\frac{1}{x+s} + \frac{1}{1+sx})$, $s \in [0,1]$, the extreme points of the Bauer simplex $\mathcal{K}$ and $f_s \in \mathcal{F}_{\mathrm{sym}}$, $f_s(x) := (x-1)^2 \kappa_s(x)$.

By the expression for symmetrised f-divergences, for $\rho, \gamma \in \mathcal{D}_d = \mathcal{D}(\mathcal{H}_A)$, $\operatorname{supp}\rho = \operatorname{supp}\gamma$, $\mathcal{M}(\rho)$ and $\mathcal{M}(\gamma)$ commute and so,

$$\begin{aligned}
D_{f_s}(\mathcal{M}(\rho)\|\mathcal{M}(\gamma)) &= \Big\langle \mathcal{M}(\rho-\gamma),\; R^{-1}_{\mathcal{M}(\gamma)}\kappa_s\big(L_{\mathcal{M}(\rho)}R^{-1}_{\mathcal{M}(\gamma)}\big)(\mathcal{M}(\rho-\gamma))\Big\rangle_{HS} \\
&= \frac{1+s}{2}\Big\langle \mathcal{M}(\rho-\gamma), \left(\frac{1}{L_{\mathcal{M}(\rho)}+sR_{\mathcal{M}(\gamma)}} + \frac{1}{sL_{\mathcal{M}(\rho)}+R_{\mathcal{M}(\gamma)}}\right)(\mathcal{M}(\rho-\gamma))\Big\rangle_{HS} \\
&= \frac{1+s}{2}\Big\langle \mathcal{M}(\rho-\gamma),\; \big((\mathcal{M}(\rho+s\gamma))^{-1} + \mathcal{M}(s\rho+\gamma)^{-1}\big)(\mathcal{M}(\rho-\gamma))\Big\rangle_{HS} \\
&= \frac{1}{2}\Big\langle \mathcal{M}(\rho-\gamma), \left(\frac{2}{L_{\mathcal{M}(\rho(s))}+R_{\mathcal{M}(\rho(s))}} + \frac{2}{L_{\mathcal{M}(\gamma(s))}+R_{\mathcal{M}(\gamma(s))}}\right)(\mathcal{M}(\rho-\gamma))\Big\rangle_{HS} \\
&= \frac{1}{2}\,\|\mathcal{M}(\rho-\gamma)\|^2_{\min,\,\mathcal{M}(\rho(s))} + \frac{1}{2}\,\|\mathcal{M}(\rho-\gamma)\|^2_{\min,\,\mathcal{M}(\gamma(s))}
\end{aligned}$$

where we define:

$$\rho(s) := \frac{\rho+s\gamma}{1+s}, \gamma(s) := \frac{s\rho+\gamma}{1+s} \in \mathcal{D}(\mathcal{H}_A).$$

Now by the operator convexity of x^{-1} on $(0,\infty)$ for all $\rho,\gamma\in\mathcal{D}_d,$ supp $\rho=$ supp γ:

$$\begin{aligned}
&\frac{1}{2}\left(R^{-1}_{\mathcal{N}(\gamma)}\kappa_s\left(L_{\mathcal{N}(\rho)}R^{-1}_{\mathcal{N}(\gamma)}\right)+R^{-1}_{\mathcal{N}(\rho)}\kappa_s\left(L_{\mathcal{N}(\gamma)}R^{-1}_{\mathcal{N}(\rho)}\right)\right)\\
&\quad=\frac{1+s}{2}\left\{\frac{\left(L_{\mathcal{N}(\rho)}+sR_{\mathcal{N}(\gamma)}\right)^{-1}+\left(sL_{\mathcal{N}(\rho)}+R_{\mathcal{N}(\gamma)}\right)^{-1}}{2}\right.\\
&\qquad\left.+\frac{\left(L_{\mathcal{N}(\gamma)}+sR_{\mathcal{N}(\rho)}\right)^{-1}+\left(sL_{\mathcal{N}(\gamma)}+R_{\mathcal{N}(\rho)}\right)^{-1}}{2}\right\}\\
&\quad=\frac{1+s}{2}\left\{\frac{\left(L_{\mathcal{N}(\rho)}+sR_{\mathcal{N}(\gamma)}\right)^{-1}+\left(sL_{\mathcal{N}(\gamma)}+R_{\mathcal{N}(\rho)}\right)^{-1}}{2}\right.\\
&\qquad\left.+\frac{\left(sL_{\mathcal{N}(\rho)}+R_{\mathcal{N}(\gamma)}\right)^{-1}+\left(L_{\mathcal{N}(\gamma)}+sR_{\mathcal{N}(\rho)}\right)^{-1}}{2}\right\}\\
&\quad\geqslant\frac{1}{2}\left\{\frac{2}{L_{\mathcal{N}(\rho(s))}+R_{\mathcal{N}(\rho(s))}}+\frac{2}{L_{\mathcal{N}(\gamma(s))}+R_{\mathcal{N}(\gamma(s))}}\right\}\\
&\quad=\frac{1}{2}\left\{\Omega^{\min}_{\rho(s)}+\Omega^{\min}_{\gamma(s)}\right\}
\end{aligned}$$

Therefore,

$$\begin{aligned}
D_{fs}(\mathcal{N}(\rho)\|\mathcal{N}(\gamma))&=\Big\langle\mathcal{N}(\rho-\gamma),\ \frac{1}{2}\Big(R^{-1}_{\mathcal{N}(\gamma)}\kappa_s\big(L_{\mathcal{N}(\rho)}R^{-1}_{\mathcal{N}(\gamma)}\big)\\
&\qquad+\ R^{-1}_{\mathcal{N}(\rho)}\kappa_s\big(L_{\mathcal{N}(\gamma)}R^{-1}_{\mathcal{N}(\rho)}\big)\Big)(\mathcal{N}(\rho-\gamma))\Big\rangle_{HS}\\
&\geqslant\frac{1}{2}\Big\langle\mathcal{N}(\rho-\gamma),\ \Big(\Omega^{\min}_{\mathcal{N}(\rho(s))}+\Omega^{\min}_{\mathcal{N}(\gamma(s))}\Big)(\mathcal{N}(\rho-\gamma))\Big\rangle_{HS}\\
&=\frac{1}{2}\big\|\mathcal{N}(\rho-\gamma)\big\|^2_{\min,\mathcal{N}(\rho(s))}+\frac{1}{2}\big\|\mathcal{N}(\rho-\gamma)\big\|^2_{\min,\mathcal{N}(\gamma(s))}
\end{aligned}$$

As a result, for any $\rho \neq \gamma \in \mathcal{D}(\mathcal{H}_A)$, $\operatorname{supp}\rho = \operatorname{supp}\gamma$:

$$\begin{aligned}\frac{D_{f_s}\left(\mathcal{M}(\rho)\|\mathcal{M}(\gamma)\right)}{D_{f_s}\left(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\right)} &\leqslant \frac{\|\mathcal{M}(\rho-\gamma)\|^2_{\min,\mathcal{M}(\rho(s))} + \|\mathcal{M}(\rho-\gamma)\|^2_{\min,\mathcal{M}(\gamma(s))}}{\|\mathcal{N}(\rho-\gamma)\|^2_{\min,\mathcal{N}(\rho(s))} + \|\mathcal{N}(\rho-\gamma)\|^2_{\min\mathcal{N}(\gamma(s))}} \\ &\leqslant \max\left\{\frac{\|\mathcal{M}(\rho-\gamma)\|^2_{\min,\mathcal{M}(\rho(s))}}{\|\mathcal{N}(\rho-\gamma)\|^2_{\min,\mathcal{N}(\rho(s))}}, \frac{\|\mathcal{M}(\rho-\gamma)\|^2_{\min,\mathcal{M}(\gamma(s))}}{\|\mathcal{N}(\rho-\gamma)\|^2_{\min,\mathcal{N}(\gamma(s))}}\right\} \\ &\leqslant \check{\eta}^{\mathrm{Riem}}_{\min}\left(\mathcal{N},\mathcal{M}\right)^{-1}\end{aligned}$$

Consider now that $\kappa \in \mathcal{K}$ must be of the following form for some probability measure $m : [0,1] \to [0,1]$:

$$\kappa(x) = \int_{[0,1]} \kappa_s(x)\, dm(s), \quad x \in (0,\infty).$$

And thus, again using the expression for symmetrized f-divergences:

$$\begin{aligned}D^{\mathrm{std}}_f(\mathcal{M}(\rho)\|\mathcal{M}(\gamma)) &= \int_{[0,1]} D^{\mathrm{std}}_{f_s}(\mathcal{M}(\rho)\|\mathcal{M}(\gamma))\, dm(s) \\ &\leqslant \check{\eta}^{\mathrm{Riem}}_{\min}(\mathcal{N},\mathcal{M})^{-1} \int_{[0,1]} D^{\mathrm{std}}_{f_s}(\mathcal{N}(\rho)\|\mathcal{N}(\gamma))\, dm(s) \\ &= \check{\eta}^{\mathrm{Riem}}_{\min}(\mathcal{N},\mathcal{M})^{-1} D^{\mathrm{std}}_f(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)).\end{aligned}$$

Overall,

$$\check{\eta}^{\mathrm{Riem}}_{\min}(\mathcal{N},\mathcal{M})^{-1} \leqslant \sup_{\substack{\rho\neq\gamma\in\mathcal{D}_d,\\ \operatorname{supp}\rho=\operatorname{supp}\gamma}} \frac{D_f(\mathcal{M}(\rho)||\mathcal{N}(\gamma))}{D_f(\mathcal{N}(\rho)||\mathcal{M}(\gamma))} \leqslant \check{\eta}^{\mathrm{Riem}}_{\kappa}(\mathcal{N},\mathcal{M})^{-1},$$

giving the result ($\hat{\eta}^{\mathrm{Riem}}_{\kappa}(\mathcal{M}) = \check{\eta}^{\mathrm{Riem}}_{\kappa}(\mathrm{id}_{\mathcal{B}(\mathcal{H}_A)},\mathcal{M})^{-1}$) □

We now complete our discussion of QC and CQ channels with the following fundamental result that a quantum output is necessary to cause variation in the divergence and Riemannian relative expansion coefficients. Since we will later see how to prove $\check{\eta}^{\mathrm{std}}_f(\mathcal{N},\mathcal{M}) \leqslant \check{\eta}^{\mathrm{std}}_{\kappa_f}(\mathcal{N},\mathcal{M})$ in general, the direction proved by [12] for the classical f-divergence and Riemannian contraction coefficient actually becomes redundant, so we will not recall it here. However, it does indeed apply, and it can be used to show that for QC channels, $\check{\eta}^{\mathrm{std}}_f(\mathcal{N},\mathcal{M}) \leqslant \check{\eta}^{\mathrm{std}}_{\kappa}(\mathcal{N},\mathcal{M})$ for all thrice differentiable f, such that $f(1) = 0$, $f''(1) > 0$.

Theorem 4.5. *(Relative Expansion Coefficients of QC Channels) [12, 11] Let $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ be QC quantum channels. Then for all $f \in \mathcal{F}$, $\kappa \in \mathcal{K}$,*

$$\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$$

Proof.
Step 1. $\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \geqslant \check{\eta}_f^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$ [11].

By definition of QC, there exist orthonormal bases $\{|\psi_i\rangle\}_{i=1}^{d_B}$ of $\mathcal{H}_B$, $\{|\varphi_i\rangle\}_{i=1}^{d'_B}$ of $\mathcal{H}'_B$, and POVMs $\{F_i\}_{i=1}^{d_B} \subseteq \mathcal{B}(\mathcal{H}_B)$, $\{G_i\}_{i=1}^{d'_B} \subseteq \mathcal{B}(\mathcal{H}'_B)$ such that for all $\rho \in \mathcal{D}_d := \mathcal{D}(\mathcal{H}_A)$:

$$\mathcal{N}(\rho) = \sum_{i=1}^{d_B} \operatorname{Tr} F_i \rho \, |\psi_i\rangle\langle\psi_i| \, , \quad \mathcal{M}(\rho) = \sum_{i=1}^{d'_B} \operatorname{Tr} G_i \rho \, |\varphi_i\rangle\langle\varphi_i| \, .$$

Let $\rho \neq \gamma \in \mathcal{D}_d$, $\operatorname{supp}\rho \leqslant \operatorname{supp}\gamma, X := \rho - \gamma$. Define $\mathbf{x}, \mathbf{y} \in P_{d'_B}$ by $x_i := \operatorname{Tr} F_i\rho,\ y_i = \operatorname{Tr} F_i\gamma,\ x'_i := \operatorname{Tr} G_i\rho,\ y'_i := \operatorname{Tr} G_i\gamma$.

Define also $\mathbf{v} := \mathbf{x} - \mathbf{y}$, $\mathbf{v}' := \mathbf{x}' - \mathbf{y}'$.

By operator convexity, f has an integral representation ((2.6)) that can be written as:

$$f(w) = a(w-1) + b(w-1)^2 + \int_{(1,\infty)} \frac{(w-1)(t(w-1)-1)}{w-1+t} \, dm(t)$$

But observe that $\frac{u(tu-1)}{u+t} - \frac{t^2+1}{t} \cdot \frac{u^2}{u+t} = -\frac{u}{t}$, and define $g_t(w) := \frac{t^2+1}{t} \cdot \frac{(w-1)^2}{w-1+t}$.
We note that $D_{w-1}^{\mathrm{std}}(\rho\|\gamma) = D_{w-1}^{\mathrm{cl}}(\mathbf{x}\|\mathbf{y}) \equiv 0$.

$$\begin{aligned} D_{g_t}^{\mathrm{std}}\left(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\right) &= D_{g_t}^{\mathrm{cl}}(\mathbf{x}\|\mathbf{y}) \\ &= \frac{t^2+1}{t} \sum_{i=1}^{d_B} y_i \cdot \frac{\left(\frac{x_i}{y_i} - 1\right)^2}{\frac{x_i}{y_i} - 1 + t} = \frac{t^2+1}{t} \sum_{i=1}^{d_B} \frac{(\mathbf{x}-\mathbf{y})_i^2}{(\mathbf{x}-\mathbf{y}+t\mathbf{y})_i} \\ &= \frac{t^2+1}{t^2} \sum_{i=1}^{d_B} \frac{v_i^2}{\left(\mathbf{y} + \frac{\mathbf{v}}{t}\right)_i} = \frac{t^2+1}{t} \Phi\left(\mathbf{y} + \frac{\mathbf{v}}{t}; \mathbf{v}\right) \\ &= \frac{t^2+1}{t} \|\mathcal{N}(X)\|^2_{\kappa, \mathcal{N}\left(\gamma + \frac{X}{t}\right)} \end{aligned}$$

$$\begin{aligned} D^{\text{std}}_{(w-1)^2}\left(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\right) &= D^{\text{cl}}_{(w-1)^2}(\mathbf{x}\|\mathbf{y}) \\ &= \sum_i y_i \cdot \left(\frac{x_i}{y_i} - 1\right)^2 = \sum_i \frac{v_i^2}{y_i} \\ &= \Phi(\mathbf{y};\mathbf{v}) = \|\mathcal{N}(X)\|^2_{\mathcal{N}(\gamma)} \end{aligned}$$

Similarly for $\mathcal{M}$:

$$D^{\text{std}}_{g_t}\left(\mathcal{M}(\rho)\|\mathcal{M}(\gamma)\right) = \frac{t^2+1}{t}\,\|\mathcal{M}(X)\|^2_{\kappa,\,\mathcal{M}\left(\gamma+\frac{X}{t}\right)},\; D^{\text{std}}_{(w-1)^2}\left(\mathcal{M}(\rho)\|\mathcal{M}(\gamma)\right) = \|\mathcal{M}(X)\|^2_{\kappa,\,\mathcal{M}(\gamma)}$$

As a result, bearing in mind that $\operatorname{supp} X \leqslant \operatorname{supp}\gamma$:

$$\begin{aligned} D^{\text{std}}_{f}\left(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\right) &= b\, D^{\text{std}}_{(w-1)^2}\left(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\right) + \int_{(1,\infty)} D^{\text{std}}_{g_t}\left(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\right)\, dm(t) \\ &= b\,\|\mathcal{N}(X)\|^2_{\kappa,\mathcal{N}(\gamma)} + \int_{(1,\infty)} \frac{t^2+1}{t}\|\mathcal{N}(X)\|^2_{\kappa,\mathcal{N}\left(\gamma+\frac{X}{t}\right)} dm(t) \\ &\geqslant \check{\eta}^{\text{Riem}}_{\kappa}(\mathcal{N},\mathcal{M})\left(b\,\|\mathcal{M}(X)\|^2_{\mathcal{M}(X)} + \int_{(1,\infty)} \frac{t^2+1}{t}\|\mathcal{M}(X)\|^2_{\mathcal{M}\left(\gamma+\frac{X}{t}\right)} dm(t)\right) \\ &= \check{\eta}^{\text{Riem}}_{\kappa}(\mathcal{N},\mathcal{M})\, D^{\text{std}}_{f}\left(\mathcal{M}(\rho)\|\mathcal{M}(\gamma)\right) \end{aligned}$$

Therefore,

$$\check{\eta}^{\text{std}}_{f}(\mathcal{N},\mathcal{M}) = \inf_{\substack{\rho\neq\gamma\in\mathcal{D}_d \\ \operatorname{supp}\rho\leqslant\operatorname{supp}\gamma}} \frac{D^{\text{std}}_{f}\left(\mathcal{N}(\rho)\|\mathcal{N}(\gamma)\right)}{D^{\text{std}}_{f}\left(\mathcal{M}(\rho)\|\mathcal{M}(\gamma)\right)} \geqslant \check{\eta}^{\text{Riem}}_{\kappa}(\mathcal{N},\mathcal{M})$$

Step 2. $\check{\eta}^{\text{std}}_{f}(\mathcal{N},\mathcal{M}) \leqslant \check{\eta}^{\text{Riem}}_{\kappa}(\mathcal{N},\mathcal{M})$ (see 4.6.2)

We know that $\check{\eta}^{\text{std}}_{f}(\mathcal{N},\mathcal{M}) \leqslant \check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M})$ (see 4.6.2).

All the Riemannian metrics on $\operatorname{Im}\mathcal{N}$ are equal, therefore: $\check{\eta}^{\text{std}}_{f}(\mathcal{N},\mathcal{M}) \leqslant \check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M}) = \check{\eta}^{\text{Riem}}_{\kappa}(\mathcal{N},\mathcal{M})$ □

4.6 The Equivalence Between Relative Expansion Coefficients, and the Maximal Metric

4.6.1 Cases of Equality between the Divergence and Riemannian Contraction and Expansion Coefficients

There are two known cases where the divergence and Riemannian contraction/relative expansion coefficients coincide for all classes of channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B), \mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$, i.e.,

$$\hat{\eta}_f^{\mathrm{std}}(\mathcal{N}) = \hat{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}), \quad \check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$$

Namely: $f(x) = x \log x,\ -\log^* x$ or $(x-1)\log x$ and $\kappa_f(x) = \frac{\log x}{x-1}$; $f(x) = (x-1)^2, \frac{(x-1)^2}{x}$ or $\frac{(x-1)^2(x+1)}{2x}$ and $\kappa_f(x) = \frac{x+1}{2x}$ [28, 6].

Indeed, we can call these only two cases because, e.g.,

$$\begin{aligned} &\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}) \\ &\qquad \Longrightarrow \check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_{\tilde{f}}^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_{f_{\mathrm{sym}}}^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) = \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}). \end{aligned}$$

This reduction from working with a divergence contraction/relative expansion coefficient to the corresponding Riemannian contraction/relative expansion coefficient is useful because the latter can be simpler to upper/lower bound. Generally, we are interested in a single reduction of the form:

$$\alpha \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \leqslant \beta \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M}),$$

where $\alpha, \beta \in (0, \infty)$ are independent of $\mathcal{N}, \mathcal{M}$. Our first result in this section will allow us to always pick $\beta = 1$ for any $f \in \mathcal{F}$. Whether there exists a choice α for any $f \in \mathcal{F}$ is not known, but by drawing upon an apparent connection between the equivalence of $\check{\eta}_f^{\mathrm{std}}$ and $\check{\eta}_{\kappa_f}^{\mathrm{Riem}}$, and the divergence of $\kappa_f(x)$ as $x \to 0^+$ (indeed our aforementioned cases of equality also involve this divergence), we will see how appropriate choices for f can be constructed. This naturally draws our attention to $\kappa_{\max} = \kappa_0$, which is the only κ_s that diverges as $x \to 0^+$, and studying the maximal metric will be a key component of this section.

4.6.2 A Generic Relationship between the Divergence and Riemannian Contraction and Expansion Coefficients

We have already met a result that relates the f-divergence to its corresponding Riemannian semi-norm (Lemma 2.1); we can use this to establish the following relationship, where the inequality for the contraction coefficients was first recognised by [37]:

Theorem 4.6. *Suppose we are provided some $f(x) \in \mathcal{F}$ and its corresponding $\kappa_f(x) = \frac{f(x)+\tilde{f}(x)}{f''(1)(x-1)^2} \in \mathcal{K}$. Given any quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$,*

$$\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) \leqslant \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})$$

Proof.
Define $\mathcal{D}_d := \mathcal{D}(\mathcal{H}_A)$. Observe that for $\rho \in \mathcal{D}_d$, $X \in T_\rho \mathcal{D}_d, \varepsilon \in (0,1], \rho_\varepsilon := \rho + \varepsilon X = (1-\varepsilon)\rho + \varepsilon\gamma$ satisfies $\rho_\varepsilon \in \mathcal{D}_d$ and $\operatorname{supp} \rho_\varepsilon = \operatorname{supp} \rho$ for sufficiently small ε. As a result,

$$\begin{aligned}
\check{\eta}_f^{\mathrm{std}}(\mathcal{N}, \mathcal{M}) &= \inf_{\substack{\rho \neq \gamma \in \mathcal{D}_d, \\ \operatorname{supp} \rho = \operatorname{supp} \gamma}} \frac{D_f^{\mathrm{std}}(\mathcal{N}(\rho) \| \mathcal{N}(\gamma))}{D_f^{\mathrm{std}}(\mathcal{M}(\rho) \| \mathcal{M}(\gamma))} \\
&\leqslant \inf_{\rho \in \mathcal{D}_d, X \in T_\rho \mathcal{D}_d} \lim_{\varepsilon \to 0^+} \frac{D_f^{\mathrm{std}}\left(\mathcal{N}(\rho) \| \mathcal{N}(\rho) + \varepsilon \mathcal{N}(X)\right)}{D_f^{\mathrm{std}}\left(\mathcal{M}(\rho) \| \mathcal{M}(\rho) + \varepsilon \mathcal{M}(X)\right)} \\
&= \inf_{\rho \in \mathcal{D}_d, X \in T_\rho \mathcal{D}_d} \frac{\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon^2} D_f^{\mathrm{std}}\left(\mathcal{N}(\rho) \| \mathcal{N}(\rho) + \varepsilon \mathcal{N}(X)\right)}{\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon^2} D_f^{\mathrm{std}}\left(\mathcal{M}(\rho) \| \mathcal{M}(\rho) + \varepsilon \mathcal{M}(X)\right)} \\
&= \inf_{\rho \in \mathcal{D}_d, X \in T_\rho \mathcal{D}_d} \frac{\|\mathcal{N}(X)\|^2_{\kappa_f, \mathcal{N}(\rho)}}{\|\mathcal{M}(X)\|^2_{\kappa_f, \mathcal{M}(\rho)}} \\
&= \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N}, \mathcal{M})
\end{aligned}$$

In particular,

$$\hat{\eta}_f^{\mathrm{std}}(\mathcal{N}) = \check{\eta}_f^{\mathrm{std}}(\mathrm{id}_{\mathcal{B}_{\mathcal{H}_A}}, \mathcal{N})^{-1} \geqslant \check{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathrm{id}_{\mathcal{B}_{\mathcal{H}_A}}, \mathcal{N})^{-1} = \hat{\eta}_{\kappa_f}^{\mathrm{Riem}}(\mathcal{N})$$

□

Notice that an alternative expression for the Riemannian contraction/relative expansion coefficient was used in the proof:

$$\check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M}) = \inf_{\rho\in\mathcal{D}_d, X\in T_\rho\mathcal{D}_d} \lim_{\varepsilon\to 0^+} \frac{D^{\text{std}}_f(\mathcal{N}(\rho)\|\mathcal{N}(\rho+\varepsilon X))}{D^{\text{std}}_f(\mathcal{M}(\rho)\|\mathcal{M}(\rho+\varepsilon X))}$$

When we can establish $\check{\eta}^{\text{std}}_f(\mathcal{N},\mathcal{M}) = \check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M})$, as in the case of the relative entropy, we can deduce that the relative expansion coefficient is attained by $\frac{D^{\text{std}}_f(\mathcal{N}(\rho)\|\mathcal{N}(\gamma))}{D^{\text{std}}_f(\mathcal{M}(\rho)\|\mathcal{M}(\gamma))}$ in the limit of a pair of states that approach each other [37].

Lemma 4.3. *Suppose we are given an $f(x)\in\mathcal{F}$ and the corresponding $\kappa_f(x) = \frac{f(x)+\tilde{f}(x)}{f''(1)} \in \mathcal{K}$. For any quantum channels $\mathcal{N}:\mathcal{B}(\mathcal{H}_A)\to\mathcal{B}(\mathcal{H}_B)$, $\mathcal{M}:\mathcal{B}(\mathcal{H}_A)\to\mathcal{B}(\mathcal{H}'_B)$,*

$$\check{\eta}^{\text{std}}_f(\mathcal{N},\mathcal{M}) = \check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M}) \implies \check{\eta}^{\text{std}}_f(\mathcal{N},\mathcal{M}) = \lim_{n\to\infty} \frac{D^{\text{std}}_f(\mathcal{N}(\rho_n)\|\mathcal{N}(\gamma_n))}{D^{\text{std}}_f(\mathcal{M}(\rho_n)\|\mathcal{M}(\gamma_n))}$$

for some sequence $\{(\rho_n,\gamma_n)\}_{n\geqslant 1}\subseteq\mathcal{D}_d\times\mathcal{D}_d$, $\rho_n-\gamma_n\to 0, \operatorname{supp}\rho_n = \operatorname{supp}\gamma_n$

Proof.
Simply consider the alternate expression for $\check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M})$.
There exists a sequence $\{\rho_n\}_{n\geqslant 1}\subseteq\mathcal{D}_d$ such that

$$0 \leqslant \inf_{\substack{\gamma\in\mathcal{D}_d\backslash\{\rho_n\},\\ \operatorname{supp}\gamma=\operatorname{supp}\rho_n}} \lim_{\varepsilon\to 0^+} \frac{D^{\text{std}}_f(\mathcal{N}(\rho_n)\|\mathcal{N}(\rho_{n,\varepsilon}(\gamma)))}{D^{\text{std}}_f(\mathcal{M}(\rho_n)\|\mathcal{M}(\rho_n,\varepsilon(\gamma)))} - \check{\eta}^{\text{std}}_f(\mathcal{N},\mathcal{M}) < \frac{1}{n},$$

where $\rho_{n,\varepsilon}(\gamma) := \rho_n + \varepsilon(\gamma-\rho_n)$

$\therefore\ \exists\gamma_n := \rho_n + \varepsilon_n(\gamma-\rho_n)\in\mathcal{D}_d$ s.t. $\frac{D^{\text{std}}_f(\mathcal{N}(\rho_n)\|\mathcal{N}(\gamma_n))}{D^{\text{std}}_f(\mathcal{M}(\rho_n)\|\mathcal{M}(\gamma_n))} - \check{\eta}^{\text{std}}_f(\mathcal{N},\mathcal{M}) < \frac{1}{n}$, where $\varepsilon_n\in(0,1)$ sufficiently small. WLOG, $\varepsilon_n < \frac{1}{n}$.

$$\therefore\ \|\gamma_n-\rho_n\|_\infty = \varepsilon_n\|\gamma-\rho_n\|_\infty \leqslant \varepsilon_n \to 0 \qquad \text{as } n\to\infty$$

□

Unfortunately, this equality is not true for general $f\in\mathcal{F}$: for a particular family of CQ channels $\Phi_{\alpha,\tau}$, it was found that (see Theorem 4.15):

$$\hat{\eta}^{\text{std}}_{(f_s)_{\text{sym}}}(\Phi_{\alpha,\tau}) = \check{\eta}^{\text{std}}_{(f_s)_{\text{sym}}}(\text{id}_{\mathcal{B}(\mathcal{H}_A)},\Phi_{\alpha,\tau})^{-1} > \check{\eta}^{\text{Riem}}_{\kappa_s}(\Phi_{\alpha,\tau}) = \check{\eta}^{\text{Riem}}_{\kappa_s}(\text{id}_{\mathcal{B}(\mathcal{H}_A)},\Phi_{\alpha,\tau})^{-1}$$

for $s\in[0,1]$ sufficiently close to 1, where $(f_s)_{\text{sym}}(x) := (x-1)^2\kappa_s(x)\in\mathcal{F}_{\text{sym}}$ [28].

4.6.3 Consequences of having Bounded Metrics

A relationship of the form $\alpha\check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M}) \leqslant \check{\eta}^{\text{std}}_f(\mathcal{N},\mathcal{M}) \leqslant \beta\check{\eta}^{\text{Riem}}_{\kappa_f}(\mathcal{N},\mathcal{M})$ where $\alpha,\beta \in (0,\infty)$ are independent of $\mathcal{N},\mathcal{M}$ - we henceforth denote this as $\check{\eta}^{\text{std}}_f \cong_{\alpha,\beta} \check{\eta}^{\text{Riem}}_{\kappa_f}$ or $\check{\eta}^{\text{std}}_f \cong \check{\eta}^{\text{Riem}}_{\kappa_f}$ - appears to be closely tied to $\kappa_f(x)$ diverging as $x \to 0^+$. For bounded standard f-divergences, corresponding to bounded $\kappa_f \in \mathcal{K}$ (i.e. $\kappa_f(0^+) < \infty$), we can instead find that the respective relative expansion coefficients are equivalent. i.e. $\eta^{\text{Riem}}_{\kappa_f} \cong \eta^{\text{Riem}}_{\kappa'_f}, \eta^{\text{std}}_f \cong \eta^{\text{std}}_{f'}$ for all $\kappa_f \in \mathcal{K}, f \in \mathcal{F}_{\text{sym}}$ s.t. $\kappa_f(0^+) < \infty$. We still cannot exclude the possibility that $\check{\eta}^{\text{std}}_f(\mathcal{N},\mathcal{M}) > 0$ iff $\check{\eta}^{\text{Riem}}_{\kappa}(\mathcal{N},\mathcal{M}) > 0$, in which case the positivity of the Riemannian relative expansion coefficient would tell us that that a non-zero amount of information about the distinguishability of quantum states w.r.t. D^{std}_f is uniformly preserved by $\mathcal{N}$ relative to $\mathcal{M}'$, even if we cannot use the Riemannian relative expansion coefficient to give a lower bound.

Theorem 4.7 (Equivalence of Bounded Metrics)**.**
Any $\kappa,\kappa' \in \mathcal{K}$ satisfy $\kappa \cong_{\alpha,\beta} \kappa'$ for $\alpha = \frac{2}{\kappa'(0^+)}$, $\beta = \frac{\kappa(0^+)}{2}$ i.e.

$$\frac{2}{\kappa'(0)}\kappa'(x) \leqslant \kappa(x) \leqslant \frac{\kappa(0^+)}{2}\kappa'(x)$$

As a consequence $\|X\|^2_{\kappa,\,\rho} \cong_{\alpha,\beta} \|X\|^2_{\kappa',\,\rho}$ for all traceless Hermitian X, density operators ρ, $\rho, X \in \mathcal{B}(\mathcal{H})$, some finite-dimensional Hilbert space $\mathcal{H}$.

In particular, $\check{\eta}^{\text{Riem}}_{\kappa} \cong_{\frac{\alpha}{\beta},\frac{\beta}{\alpha}} \check{\eta}^{\text{Riem}}_{\kappa'}$.

Proof.
Let $s, s' \in (0,1]$, so that the corresponding $\kappa_s(x), \kappa_{s'}(x)$ are bounded (i.e. $\kappa_s(0^+), \kappa_{s'}(0^+) \in (0,\infty)$).

$$\begin{aligned}
\frac{\kappa_s(x)}{\kappa_{s'}(x)} &= \frac{\frac{(1+s)^2}{2}\cdot\frac{1+x}{(x+s)(1+sx)}}{\frac{(1+s')^2}{2}\cdot\frac{1+x}{(x+s')(1+s'x)}} \\
&= \frac{s'(1+s)^2}{s(1+s')^2}\cdot\frac{(x+s')(x+s'^{-1})}{(x+s)(x+s^{-1})} \\
&= \frac{s'(1+s)^2}{s(1+s')^2}\cdot\left(1+\left[(s'+s'^{-1})-(s+s^{-1})\right]\cdot\frac{x}{x^2+1+(s+s^{-1})x}\right)
\end{aligned}$$

Since
$$\frac{d}{dx}\frac{\kappa_s(x)}{\kappa_{s'}(x)} = \frac{s'(1+s)^2}{s(1+s')^2}\cdot\left[(s'+s'^{-1})-(s+s^{-1})\right]\cdot\frac{1-x^2}{x^2+1+(s+s^{-1})x},$$

$\frac{\kappa_s(x)}{\kappa_{s'}(x)}$ is increasing when $x \leqslant 1$ and $s > s'$ or $x \geqslant 1$ and $s \leqslant s'$; and it is decreasing when $x \geqslant 1$ and $s > s'$, or $x \leqslant 1$ and $s \leqslant s'$. We're interested in taking $s' = 1$ ($\kappa_{\min}(x) := \kappa_1(x)$), so necessarily $s \leqslant s'$.

Note that: $\lim_{x\to\infty}\frac{\kappa_s(x)}{\kappa_{s'}(x)} = \lim_{x\to\infty}\frac{x\kappa_s(x)}{x\kappa_{s'}(x)} = \lim_{x\to\infty}\frac{\kappa_s(x^{-1})}{\kappa_{s'}(x^{-1})} = \frac{\kappa_s(0^+)}{\kappa_{s'}(0^+)}$, and $\kappa_{\min}(0^+) = 2$.

$$\therefore \kappa_{\min}(x) \leqslant \kappa_s(x) \leqslant \frac{\kappa_s(0^+)}{2}\cdot\kappa_{\min}(x) \text{ for all } s \in [0,1]$$

Now consider the integral representation that expresses $\kappa(x), \kappa'(x)$ in terms of the extreme points $\kappa_s(x)$ of the Choquet simplex $\mathcal{K}$.

$$\therefore \quad \kappa_{\min}(x) \leqslant \kappa(x) \leqslant \frac{\kappa(0^+)}{2}\kappa_{\min}(x),\ \kappa_{\min}(x) \leqslant \kappa'(x) \leqslant \frac{\kappa'(0^+)}{2}\kappa_{\min}(x)$$

$$\therefore \quad \frac{2}{\kappa'(0^+)} \leqslant \frac{\kappa(x)}{\kappa'(x)} \leqslant \frac{\kappa(0^+)}{2}$$

This gives the result. ∎

Theorem 4.8 (Equivalence of Bounded Standard Divergences)**.** *Given any $f, g \in \mathcal{F}$, corresponding to $\kappa_f, \kappa_g \in \mathcal{K}$ such that $\kappa_f(0^+), \kappa_g(0^+) < \infty$, then*

$$D_f^{\text{std}} \cong D_g^{\text{std}}$$

In particular, $\check{\eta}_f^{\text{std}} \cong \check{\eta}_g^{\text{std}}$.

Proof.
WLOG let $f'(1) = g'(1) = 0$ (since $D_f^{\text{std}}(\rho\|\gamma) \equiv D_{f-f'(1)(x-1)}^{\text{std}}(\rho\|\gamma)$). By (2.6), this ensures that $f'(\infty), \tilde{f}'(\infty) > 0$.

Since κ_f is bounded (i.e. $\kappa_f(0^+) < \infty$), and defining $f_{\text{sym}}(x) := (x-1)^2\kappa_f(x)$:

$$\begin{aligned} f'_{\text{sym}}(\infty) &:= \lim_{x\to\infty}\frac{f_{\text{sym}}(x)}{x} = \frac{f'(\infty)+\tilde{f}'(\infty)}{2f''(1)} \\ &= \lim_{x\to\infty}\frac{(x-1)^2\kappa(x)}{x} = \lim_{x\to\infty} x\kappa(x) = \lim_{x\to\infty}\kappa(x^{-1}) \\ &= \kappa(0^+) < \infty \end{aligned}$$

$\Rightarrow\ f'(\infty)$ and $\tilde{f}'(\infty) \equiv f(0^+)$ are finite. Similarly, for g.

Note that: $\lim_{x\to 1} \frac{f(x)}{g(x)} = \lim_{x\to 1} \frac{f''(x)}{g''(x)} = \frac{f''(1)}{g''(1)} \in (0,\infty)$ by L'Hôpital's rule.

Therefore by continuity, $\frac{f(x)}{g(x)}$ is bounded, i.e. $\exists 0 < \alpha < \beta$ s.t. $\alpha \leqslant \frac{f(x)}{g(x)} \leqslant \beta$ for all $x \in (0,\infty)$. This implies that

$$\alpha \leqslant \frac{D_f^{\mathrm{std}}(\rho\|\gamma)}{D_g^{\mathrm{std}}(\rho\|\gamma)} \leqslant \beta \quad \text{for all states } \rho \neq \gamma \in \mathcal{B}(\mathcal{H})$$

This gives the result. □

4.6.4 The Impact of the Maximal Metric

The previous section has helped to bolster the possibility that metrics $\kappa \in \mathcal{K}$ that diverge as $x \to 0^+$ are somehow special, since for all bounded metrics $\kappa_f, \kappa_g \in \mathcal{K}$, $\check{\eta}^{\mathrm{Riem}}_{\kappa_f} \cong \check{\eta}^{\mathrm{Riem}}_{\kappa_g}$, $\check{\eta}^{\mathrm{std}}_f \cong \check{\eta}^{\mathrm{std}}_g$ by Theorem 4.7 and Theorem 4.8, suggesting that the corresponding relative expansion coefficients make similar comparisons between all pairs of quantum channels in this bounded case. Recall the following integral representation for $\kappa \in \mathcal{K}$, where m is the unique probability measure on $[0,1]$ corresponding to κ:

$$\kappa(x) = \int_{[0,1]} \kappa_s(x)\, dm(s), \qquad \kappa_s(x) = \frac{1+s}{2}\left(\frac{1}{x+s} + \frac{1}{sx+1}\right) \tag{$\star$}$$

$\kappa_{\max} := \kappa_0(x)$ is the only κ_s that diverges as $x \to 0^+$, so it is only natural to focus on this. However, it is possible for $\kappa(0^+) = \infty$ while $m(0) = 0$, which is the case for the BKM metric (corresponding to the relative entropy). While an obvious connection with $\kappa_{\max}$ may not be necessary, we will now see a simple mechanism for constructing $f \in \mathcal{F}, \kappa_f \in \mathcal{K}$ such that $\check{\eta}^{\mathrm{std}}_f \cong \check{\eta}^{\mathrm{Riem}}_{\kappa_f}$ that uses the convenient properties of the maximal metric.

Lemma 4.4. *(Equivalence based on other Equivalence) For any $f \in \mathcal{F}_{\mathrm{sym}}$ corresponding to $\kappa_f \in \mathcal{K}$, and a $\kappa \in \mathcal{K}$ such that for some $0 < a < b$ (necessarily $a \leqslant 1 \leqslant b$):*

$$a\,\kappa_f(x) \leqslant \kappa(x) \leqslant b\,\kappa_f(x) \quad \textit{for all } x \in (0,\infty)$$

Then, defining $g(x) := (x-1)^2\kappa(x) \in \mathcal{F}_{\mathrm{sym}}$ so that $\kappa = \kappa_g$,

$$\check{\eta}^{\mathrm{std}}_f \cong_{\gamma,\delta} \check{\eta}^{\mathrm{Riem}}_{\kappa_f} \ \textit{for some } 0 < \gamma \leqslant \delta \implies \check{\eta}^{\mathrm{std}}_g \cong_{\alpha,\beta} \check{\eta}^{\mathrm{Riem}}_{\kappa_g}$$

for $\alpha = \frac{a^2\gamma}{b^2}, \beta = 1$.

Proof.
Since $a\,\kappa_f(x) \leqslant \kappa(x) \leqslant b\,\kappa_f(x) \iff a\,f(x) \leqslant g(x) \leqslant b\,f(x)$, we also have $a\|X\|^2_{\kappa_f,\rho} \leqslant \|X\|^2_{\kappa,\rho} \leqslant b\|X\|^2_{\kappa_f,\rho}$ and $aD_f^{\text{std}}(\rho\|\gamma) \leqslant D_g^{\text{std}}(\rho\|\gamma) \leqslant bD_f^{\text{std}}(\rho\|\gamma)$ for any traceless Hermitian X, density operators ρ,γ, supp ρ = supp γ, $X,\rho,\gamma \in \mathcal{B}(\mathcal{H})$ for some Hilbert space $\mathcal{H}$. The latter can also be viewed as a direct consequence of the alternate form of symmetrised f-divergences.

Therefore,

$$\check{\eta}_g^{\text{std}} \geqslant \frac{a}{b}\check{\eta}_f^{\text{std}} \geqslant \frac{a\gamma}{b}\check{\eta}_{\kappa_f}^{\text{Riem}} \geqslant \frac{a^2\gamma}{b^2}\check{\eta}_{\kappa}^{\text{Riem}},$$

so then $\alpha = a^2\gamma/b^2$, and by Theorem 4.6, we get $\beta = 1$ for free, so we are done. ☐

Corollary 4.3. *(Equivalence based on Equality) For any $f \in \mathcal{F}$ corresponding to $\kappa_f \in \mathcal{K}$, and a $\kappa \in \mathcal{K}$ such that for some $0 < a < b$:*

$$a\,\kappa_f(x) \leqslant \kappa(x) \leqslant b\,\kappa_f(x) \quad \forall x \in (0,\infty)$$

Then, defining $g(x) = (x-1)^2\kappa(x) \in \mathcal{F}_{\text{sym}}$ so that $\kappa = \kappa_g$,

$$\check{\eta}_f^{\text{std}} \equiv \check{\eta}_{\kappa_f}^{\text{Riem}} \implies \check{\eta}_g^{\text{std}} \cong_{\alpha,\beta} \check{\eta}_{\kappa_g}^{\text{Riem}}$$

for $\alpha = \frac{a^2}{b^2}$, $\beta = 1$.

Proof. Since $\check{\eta}_f^{\text{std}} \leqslant \check{\eta}_{f_{\text{sym}}}^{\text{std}} \leqslant \check{\eta}_{\kappa_f}^{\text{Riem}}$, for $f_{\text{sym}} := \frac{f+\tilde{f}}{f''(1)} \in \mathcal{F}_{\text{sym}}$, we have $\check{\eta}_{f_{\text{sym}}}^{\text{std}} = \check{\eta}_{\kappa_f}^{\text{Riem}}$, and thus we can take $f \in \mathcal{F}_{\text{sym}}$ WLOG. This reduces to Lemma 4.4 with $\gamma = \delta = 1$. ☐

Theorem 4.9. *(Equivalence via Maximal Metric) For any $f \in \mathcal{F}_{\text{sym}}$ corresponding to $\kappa = \kappa_f \in \mathcal{K}$ of the form $(\star)$ with $m(0) > 0$ (note: $m(0) := \lim_{x\to\infty} 2\kappa(x) = \lim_{x\to\infty} \frac{2f(x)}{x^2}$),*

$$\check{\eta}_f^{\text{std}} \cong_{\alpha,\beta} \check{\eta}_{\kappa_f}^{\text{Riem}}$$

where $\alpha = m(0)^2, \beta = 1$.

Proof. Since $m(0) > 0$ and by the maximality of $\kappa_{\max}$, we have the following equivalence:

$$m(0)\kappa_{\max}(x) \leqslant \kappa(x) \leqslant \kappa_{\max}(x) \quad \text{for all } x \in (0,\infty)$$

We also know that

$$\check{\eta}_{(f_{\max})\text{sym}}^{\text{std}} = \check{\eta}_{\kappa_{\max}}^{\text{Riem}}$$

Therefore, this problem reduces to Corollary 4.3, with $a = m(0), b = 1$. ☐

Remarks:

1. Theorem 4.9 tells us that any convex combination $\kappa_f = (1-\lambda)\kappa' + \lambda\kappa_{\max} \in \mathcal{K}, \kappa' \in \mathcal{K}, \lambda \in (0,1]$ satisfies the equivalence

$$\check{\eta}^{\text{std}}_{f_{\text{sym}}} \cong_{\lambda^2,\,1} \check{\eta}^{\text{Riem}}_{\kappa_f}$$

thus offering a simple way to construct f-divergences with $\kappa_f(0^+) = \infty$ and this equivalence.

2. Indeed, since $\check{\eta}^{\text{std}}_{x\log x} = \check{\eta}^{\text{Riem}}_{\frac{\log x}{x-1}}$, we could similarly apply Corollary 4.3 to construct f-divergences with the equivalence property. It may be a bit less obvious to characterize relevant $\kappa' \in \mathcal{K}$, but certainly we know that $\kappa = (1-\lambda)\kappa_{\min} + \lambda \cdot \frac{\log x}{x-1}, \lambda \in (0,1]$ will satisfy the equivalence since $\kappa_{\min}(x) \leqslant \frac{\log x}{x-1}$ (equality iff $x \to 1$). This suggests we can take $\kappa' = \kappa_s$ for $s \in (0,1]$ sufficiently large, since κ_s is uniformly decreasing in s.

3. Lemma 4.4 and Corollary 4.3 can only be used to construct $g \in \mathcal{F}_{\text{sym}}$ from $f \in \mathcal{F}_{\text{sym}}$ such that $\check{\eta}^{\text{std}}_g \cong \check{\eta}^{\text{std}}_f$, so the corresponding divergence/Riemannian expansion coefficients make similar comparisons between arbitrary pairs of quantum channels.

While Theorem 4.9 is convenient, and it secures the importance of the maximal metric, it does not fully explain the connection between $\kappa_f(0^+) = \infty$ and $\check{\eta}^{\text{std}}_f \cong \check{\eta}^{\text{std}}_{\kappa_f}$. We cannot guarantee that $\kappa_f(0^+) = \infty \implies \check{\eta}^{\text{std}}_{f_{\text{sym}}} \cong \check{\eta}^{\text{Riem}}_{\kappa_f}$; nevertheless, the fact this divergence in κ ensures $D^{\text{std}}_{f_{\text{sym}}}(\rho\|\gamma) < \infty \iff \operatorname{supp}\rho = \operatorname{supp}\gamma \implies \operatorname{supp}\rho - \gamma \leqslant \operatorname{supp}\rho, \operatorname{supp}\gamma$ provides some intuition. If it isn't already convincing that for $\kappa(x) = \kappa_{\text{BKM}}(x)$, $m(0) = \lim_{x\to\infty} \frac{2\log x}{x-1} = 0$, we can see in the following integral representation that the BKM metric inherits its divergence as $x \to 0^+$ from its probability measure m having sufficient concentration in the neighbourhood of $s = 0$, despite $m(0) = 0$.

$$\kappa_{\text{BKM}}(x) = \frac{\log x}{x-1} = \int_0^1 \kappa_s(x) \cdot \frac{2}{(1+s)^2} ds$$

The most clear-cut explanation for $\check{\eta}^{\text{std}}_{x\log x} \equiv \check{\eta}^{\text{Riem}}_{\kappa_{\text{BKM}}}$ is by using the integral representation for the relative entropy recognized by [6] for $\operatorname{supp}\rho \leqslant \operatorname{supp}\gamma$:

$$D(\rho\|\gamma) = \int_0^1 \int_0^s \|\rho - \gamma\|^2_{\kappa_{\text{BKM}},\rho_t} dt ds$$

where $\rho_t := (1-t)\gamma + t\rho$. From this, we can deduce for channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B), \mathcal{D}_d = \mathcal{D}(\mathcal{H}_A)$:

$$\begin{aligned}
\check{\eta}^{\text{std}}_{x\log x}(\mathcal{N},\mathcal{M}) &= \inf_{\substack{\rho\neq\gamma\in\mathcal{D}_d,\\ \operatorname{supp}\rho\leqslant\operatorname{supp}\gamma}} \frac{D(\mathcal{N}(\rho)\|\mathcal{N}(\gamma))}{D(\mathcal{M}(\rho)\|\mathcal{M}(\gamma))} \\
&= \inf_{\substack{\rho\neq\gamma\in\mathcal{D}_d,\\ \operatorname{supp}\rho\leqslant\operatorname{supp}\gamma}} \frac{\int_0^1\int_0^s \|\mathcal{N}(\rho-\gamma)\|^2_{\kappa_{\text{BKM}},\mathcal{N}(\rho_t)}\,dt\,ds}{\int_0^1\int_0^s \|\mathcal{M}(\rho-\gamma)\|^2_{\kappa_{\text{BKM}},\mathcal{M}(\rho_t)}\,dt\,ds} \\
&\geqslant \inf_{\substack{\rho\neq\gamma\in\mathcal{D}_d,\\ \operatorname{supp}\rho\leqslant\operatorname{supp}\gamma}} \frac{\int_0^1\int_0^s \check{\eta}^{\text{Riem}}_{\kappa_{\text{BKM}}}(\mathcal{N},\mathcal{M})\|\mathcal{M}(\rho-\gamma)\|^2_{\kappa_{\text{BKM}},\mathcal{M}(\rho_t)}dtds}{\int_0^1\int_0^s \|\mathcal{M}(\rho-\gamma)\|^2_{\kappa_{\text{BKM}},\mathcal{M}(\rho_t)}dtds} \\
&= \check{\eta}^{\text{Riem}}_{\kappa_{\text{BKM}}}(\mathcal{N},\mathcal{M})
\end{aligned}$$

And we combine this with the fact that $\check{\eta}^{\text{std}}_{x\log x}(\mathcal{N},\mathcal{M}) = \check{\eta}^{\text{Riem}}_{\kappa_{\text{BKM}}}(\mathcal{N},\mathcal{M})$ by Theorem 4.6. Originally, in [28], they used another similar integral representation for the symmetrized relative entropy $D_{\text{sym}}(\rho||\gamma) = \int_0^1 \|\rho-\gamma\|^2_{\kappa_{\text{BKM}},\rho_t}dt$ to establish $\check{\eta}^{\text{std}}_{(x-1)\log x}(\mathcal{N},\mathcal{M}) = \check{\eta}^{\text{Riem}}_{\kappa_{\text{BKM}}}(\mathcal{N},\mathcal{M})$, but the above is stronger. Because they sufficed to establish the equality between the divergence and Riemannian relative expansion coefficients, we can investigate whether the relationships between $D^{\text{std}}_{(x-1)^2}$ and D with their corresponding Riemannian semi-norms can be replicated. As it turns out, these specific relationships uniquely define the f-divergence, and proving this will illustrate a convenient method that may be used to check candidate relationships in the future. In effect, we will see from this perspective that, for example, the integral representation for relative entropy by [6] was rather inevitable.

Theorem 4.10. *The only f-divergence satisfying, for all Hilbert spaces $\mathcal{H}$ with $\dim\mathcal{H}\geqslant 2$, $\rho\in\mathcal{D}(\mathcal{H})$, and traceless Hermitian operator $X\in\mathcal{B}(\mathcal{H})$ such that $\operatorname{supp}X\leqslant\operatorname{supp}\rho$ and $\rho_t := \rho + tX \geqslant 0$ for all $t\in[0,1]$:*

$$\frac{d^2}{dt^2}D_f(\gamma\|\rho_t) = w(t)\langle X, \Omega_{\rho_t,\kappa_f},(X)\rangle_{HS} = w(t)\|X\|^2_{\rho_t,\kappa_{f'}} \tag{†}$$

for $f, f'\in\mathcal{F}, w:[0,1]\to\mathbb{R}$, is the reverse-argument relative entropy, $f(x) = -\log^* x$.

Proof. Let $\rho = \lambda |0\rangle\langle 0| + (1-\lambda)|1\rangle\langle 1| = \mu(|0\rangle\langle 0| - |1\rangle\langle 1|)$ where $\lambda \in (0,1), \mu \in (-\lambda, 0) \cup (0, 1-\lambda)$ and $|0\rangle, |1\rangle$ are orthonormal.

Define the density operator $\rho_t := \rho + tX$, $t \in [0,1]$. Then we evaluate the divergence:

$$D_f(\gamma\|\rho_t) = (\lambda_1 + t\mu) f\left(\frac{\lambda}{\lambda + t\mu}\right) + (1-\lambda-t\mu) f\left(\frac{1-\lambda}{1-\lambda-t\mu}\right)$$

Therefore,

$$\begin{aligned}\frac{d^2}{dt^2} D_f(\gamma\|\rho_t) &= \frac{\lambda^2\mu^2}{(\lambda+t\mu)^3} f''\left(\frac{\lambda}{\lambda+t\mu}\right) + \frac{(1-\lambda)^2\mu^2}{(1-\lambda-t\mu)^3} f''\left(\frac{1-\lambda}{1-\lambda-t\mu}\right)\\ &= w(t)\|X\|^2_{\rho_t,\kappa_{f'}} = w(t)\,\mathrm{Tr}\ X^2\rho_t^{-1} = w(t)\left(\frac{\mu^2}{1-\lambda+t\mu} + \frac{\mu^2}{\lambda - t\mu}\right)\end{aligned}$$

Dividing both sides by μ^2 and taking $\mu \to 0$, we compare:

$$\left(\frac{1}{\lambda} + \frac{1}{1-\lambda}\right) f''(1) = \left(\frac{1}{\lambda} + \frac{1}{1-\lambda}\right) w(t) \implies w(t) = f''(1) \quad \forall t \in [0,1] \tag{4.2}$$

Consider re-parametrising to only the variables $\lambda \in [0,1]$, $\chi = \lambda + t\mu \in (0,1)\backslash\{\lambda\}$:

$$\frac{\lambda^2}{\chi^3} f''\left(\frac{\lambda}{\chi}\right) + \frac{(1-\lambda)^2}{(1-\chi)^3} f''\left(\frac{1-\lambda}{1-\chi}\right) = f''(1)\left(\frac{1}{\chi} + \frac{1}{1-\chi}\right) = \frac{f''(1)}{\chi(1-\chi)}$$

$$\therefore (1-\chi)\cdot\left(\frac{\lambda}{\chi}\right)^2 f''\left(\frac{\lambda}{\chi}\right) + \chi \cdot \frac{1-\lambda}{1-\chi} f''\left(\frac{1-\lambda}{1-\chi}\right) = f''(1)$$

Let $\chi \to 0$ while maintaining $\lambda = x\chi$, any fixed $x \in (0,\infty)\backslash\{1\}$ (so that also $\lambda \to 0$):

$$(1-\chi)\cdot\left(\frac{\lambda}{\chi}\right)^2 f''\left(\frac{\lambda}{\chi}\right) + \chi\cdot\frac{1-\lambda}{1-\chi} f''\left(\frac{1-\lambda}{1-\chi}\right) \to x^2 f''(x) + 0\cdot 1^2 \cdot f''(1) = x^2 f''(x)$$

$$\begin{aligned}&\therefore \forall x \in (0,\infty): \quad x^2 f''(x) = f''(1),\ f(1) = 0\\ &\implies f(x) = -f''(1)\log^* x + c(x-1),\ \text{constant } c \in \mathbb{R},\ \forall x \in (0,\infty)\end{aligned}$$

And WLOG taking $f''(1) = 1, c = 0$ (since $D_f = D_{f/f''(1)-c(x-1)}$), we get the result. □

Theorem 4.11. *The only f-divergence satisfying, for all Hilbert spaces $\mathcal{H}$ with $\dim \mathcal{H} \geqslant 2$, $\rho \in \mathcal{D}(\mathcal{H})$, and traceless Hermitian operator $X \in \mathcal{B}(\mathcal{H})$ such that $\operatorname{supp} X \leqslant \operatorname{supp} \rho$ and $\rho_t := \rho + tX \geqslant 0$ for all $t \in [0,1]$:*

$$D_f(\rho \| \rho_t) = w(t) \|X\|^2_{\rho_t, \kappa_{f'}} \tag{††}$$

for $f, f' \in \mathcal{F}$, $w : [0,1] \to \mathbb{R}$, is the maximal standard f-divergence, $f(x) = (x-1)^2$, with $w(t) = t^2$.

Proof. Let $\rho = \lambda |0\rangle\langle 0| + (1-\lambda) |1\rangle\langle 1| = \mu(|0\rangle\langle 0| - |1\rangle\langle 1|)$ where $\lambda \in (0,1), \mu \in (-\lambda, 0) \cup (0, 1-\lambda)$ and $|0\rangle, |1\rangle$ are orthonormal.

Define the density operator $\rho_t := \rho + tX$, $t \in [0,1]$.

$$\begin{aligned} D_f(\rho\|\rho_t) &= (\lambda + t\mu) f\left(\frac{\lambda}{\lambda + t\mu}\right) + (1 - \lambda - t\mu) f\left(\frac{1-\lambda}{1-\lambda+t\mu}\right) \\ &= w(t)\|X\|^2_{\rho_t,\kappa_f} = w(t)\mu^2\left(\frac{1}{\lambda+t\mu} + \frac{1}{1-\lambda-t\mu}\right) \end{aligned} \tag{@}$$

Fix t, and keep λ, μ variable. Define $\chi = \lambda + t\mu$, and let us maintain $\lambda = x\chi$ for fixed $x \in (1, \frac{1}{1-t})$ by choosing μ appropriately.

Thus, also $t\mu = (1-x)\chi$. Dividing both sides of (@) by μt:

$$\frac{1}{1-x} f(x) + \frac{1-x}{(1-x)\chi} f\left(\frac{1-x\chi}{1-\chi}\right) = \frac{w(t)}{t^2}(1 - x + \frac{(1-x)\chi}{1-\chi})$$

Letting $\chi \to 0$ (note: $\chi \in ((1-t)\lambda, \lambda) \cup (\lambda, t + (1-t)\lambda)$, so as we take $\lambda \to 0$, this limit becomes possible):

$$\frac{1}{1-x} f(x) + f'(1) = \frac{w(t)}{t^2}(1-x) \implies \frac{f(x)}{(x-1)^2} + \frac{f'(1)}{x-1} = \frac{w(t)}{t^2} = C$$

for all t, $x \in (1, \frac{1}{1-t})$ (hence all t and all $x > 1$) and some constant $C \geqslant 0$.
And WLOG we take $C = 1$, $f'(1) = 0$ (since $D_f = (D_{(f - f'(1)(x-1))/C})$

Alternatively, take $\chi = 1-\lambda-t\mu, 1-\lambda = x\chi$ for fixed $x \in (1-t,1)$ then if we again divide both sides by $\mu t = (x-1)\chi$:

$$\frac{1-x}{(x-1)\chi} f\left(\frac{1-x\chi}{1-x}\right) + \frac{1}{x-1} f(x) = \frac{w(t)}{t^2}\left(\frac{(x-1)\chi}{1-\chi} + x - 1\right)$$

Letting $\chi \to 0$:

$$f'(1) + \frac{1}{x-1} f(x) = \frac{w(t)}{t^2}(x-1) \implies \frac{f(x)}{(x-1)^2} + \frac{f'(1)}{x-1} = \frac{w(t)}{t^2} = C$$

for all t, $x \in (1-t,1)$ (hence all t and all $x < 1$) and some constant $C \geqslant 0$.
WLOG taking $C = 1$, $f'(1) = 0$, we get the result.

Note: we used

$$\frac{1-\chi}{\chi} f\left(\frac{1-x\chi}{1-\chi}\right) \sim \frac{1}{\chi} \cdot f'(1) \cdot \frac{(1-x)\chi}{1-\chi} \sim (1-x)f'(1) \text{ as } \chi \to 0.$$

$\square$

Remarks:

1. Our strategy in Theorem 4.10 and Theorem 4.11 involves reducing the problem to checking the condition on commuting pairs of qubit states. This turns out to be sufficient in these cases.

2. Notice that in the classical setting, a specific relationship between the classical f-divergence and the Riemannian semi-norm is not necessary for $\check{\eta}_f^{\mathrm{cl}}(A,B) = \check{\eta}^{\mathrm{Riem,cl}}(A,B)$, which simply holds indefinitely.

3. In Theorem 4.10, it is good to bear in mind that the Rényi divergences are not f-divergences, so this needs to be checked separately. In fact, working again in the classical two-dimensional case, it can verified that the Rényi divergences do not satisfy the relation for $\alpha \neq 1$.

4.7 Relative Expansion Coefficients for Quantum Channels with Only Full Rank Output States

We have previously seen that by restricting to QC channels, the divergence and Riemannian expansion coefficients become degenerate, i.e.,

$$\check{\eta}_f^{\text{std}} = \check{\eta}_\kappa^{\text{Riem}}$$

for all $f \in \mathcal{F}, \kappa \in \mathcal{K}$. Here, we will consider another restriction to a family of channels that ensures (the slightly weaker)

$$\check{\eta}_f^{\text{std}} \cong \check{\eta}_\kappa^{\text{Riem}}$$

for all $f \in \mathcal{F}, \kappa \in \mathcal{K}$. Namely, this happens to hold when we are comparing quantum channels whose output density operators have full rank. We will find it useful to define the following notation:

1. $\lambda_{\min}(\rho)$ (resp. $\lambda_{\max}(\rho)$) denotes the minimum (resp. maximum) eigenvalue of a density operator $\rho \in \mathcal{D}(\mathcal{H})$
2. For a CPTP map $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$,

$$\lambda_{\min}(\mathcal{N}) := \min_{\rho \in \mathcal{D}(\mathcal{H}_A)} \lambda_{\min}(\mathcal{N}(\rho)),$$

$$\lambda_{\max}(\mathcal{N}) := \max_{\rho \in \mathcal{D}(\mathcal{H}_A)} \lambda_{\max}(\mathcal{N}(\rho)).$$

3. The *condition number* of a density operator $\rho \in \mathcal{D}(\mathcal{H})$ is $c(\rho) := \frac{\lambda_{\max}(\rho)}{\lambda_{\min}(\rho)}$

Then, we can notice that a quantum channel $\mathcal{N}$ has only full rank output states iff $\lambda_{\min}(\mathcal{N}) > 0$. Also note that $\lambda_{\min}(\rho) \leqslant 1$.

The following results take inspiration from Section 2.4, where it can be noticed from formulae for the Riemannian semi-norms that $\|X\|^2_{\kappa_{\max},\,\rho}$ often shows up as a factor. The proof methods used highlight the importance of the integral representations in (2.6),(2.7).

Theorem 4.12 (Equivalence of Riemannian Semi-Norms on Full Rank States)**.**
Any $\kappa \in \mathcal{K}$ satisfies for all traceless Hermitian X, positive definite density operators ρ, $\rho, X \in \mathcal{B}(\mathcal{H})$, some finite-dimensional Hilbert space $\mathcal{H}$:

$$\|X\|^2_{\kappa,\,\rho} \cong_{\alpha,\beta} \|X\|^2_{\kappa_{\max},\,\rho}$$

For $\alpha = \kappa(c(\rho))$, $\beta = 1 > 0$.

In particular, $\check{\eta}^{\mathrm{Riem}}_{\kappa}(\mathcal{N},\mathcal{M}) \cong \check{\eta}^{\mathrm{Riem}}_{\kappa_{\max}}(\mathcal{N},\mathcal{M})$ for all quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ whose output density operators have full rank.

Proof.
Consider the spectral decomposition $\rho = \sum_{x \,\in\, \mathrm{supp}(\rho)} x P_x$ and the integral representation for $\kappa \in \mathcal{K}$ (2.7):

$$\kappa(x) = \int_{[0,1]} \frac{1+s}{2}\left(\frac{1}{x+s} + \frac{1}{sx+1}\right) dm(s)$$

Then

$$\begin{aligned} 1 \geqslant \frac{\|X\|^2_{\kappa,\rho}}{\|X\|^2_{\kappa_{\max},\rho}} &= \frac{\sum_{x,y \,\in\, \mathrm{supp}(\rho)} \ \mathrm{Tr}\ XP_xXP_y \ \cdot \ \int_{[0,1]} \frac{1+s}{2}\left(\frac{1}{x+sy} + \frac{1}{sx+y}\right) dm(s)}{\sum_{x,y \,\in\, \mathrm{supp}(\rho)} \ \mathrm{Tr}\ XP_xXP_y \ \cdot \ \frac{1}{x}} \\ &\geqslant \min_{x,y} \int_{[0,1]} \frac{1+s}{2}\left(\frac{x}{x+sy} + \frac{x}{sx+y}\right) dm(s) \\ &= \min_{x,y} \kappa\left(\frac{y}{x}\right) = \kappa(c(\rho)) > 0 \end{aligned}$$

Where the first inequality follows from the maximality of $\kappa_{\max}$ and the final equality follows from the fact $\kappa \in \mathcal{K}$ is necessarily (positive, operator) monotone decreasing.
This gives the result. ∎

Theorem 4.13 (Equivalence of Standard Divergences on Full Rank States)**.**
Any $f \in \mathcal{F}$ satisfies for all positive definite density operators $\rho, \gamma \in \mathcal{D}^+(\mathcal{H})$, some finite-dimensional Hilbert space $\mathcal{H}$:

$$D^{\mathrm{std}}_f(\rho\|\gamma) \cong_{\alpha,\beta} D^{\mathrm{std}}_{\max}(\rho\|\gamma)$$

For $\alpha = \nu_f(\lambda_{\max}(\rho)/\lambda_{\min}(\gamma)), \beta = \nu_f(\lambda_{\min}(\rho)/\lambda_{\max}(\gamma)) > 0$, where $\nu_f(x) := \frac{f(x)-f'(1)(x-1)}{(x-1)^2}$.

In particular, $\check{\eta}^{\mathrm{std}}_f(\mathcal{N},\mathcal{M}) \cong \check{\eta}^{\mathrm{std}}_{f_{\max}}(\mathcal{N},\mathcal{M})$ for all quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B), \mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ whose output density operators have full rank.

Proof.
Consider the spectral decompositions $\rho = \sum_{a \in \mathrm{supp}(\rho)} aP_a$, $\gamma = \sum_{b \in \mathrm{supp}(\gamma)} bQ_b$ and the integral representation for $f \in \mathcal{F}$ (where $c \geqslant 0$)(2.6):

$$f(x) = f'(1)(x-1) + c(x-1)^2 + \int_{[0,\infty)} \frac{(x-1)^2}{x+s}\, d\mu(s), \quad x \in (0,\infty).$$

This in particular tells us that:

$$\nu_f(x) = c + \int_{[0,\infty)} \frac{1}{x+s}\, d\mu(s), \quad x \in (0,\infty).$$

Is a decreasing, strictly positive, function and $\nu_f(\infty) = c$. Thus:

$$\begin{aligned}
\frac{D_f^{\mathrm{std}}(\rho\|\gamma)}{D_{\max}^{\mathrm{std}}(\rho\|\gamma)} &= \frac{\sum_{a\in\mathrm{spec}(\rho),b\in\mathrm{spec}(\gamma)} \mathrm{Tr}\ P_aQ_b \cdot [cb(a/b-1)^2 + \int_{[0,\infty)} b \cdot \frac{(a/b-1)^2}{a/b+s} d\mu(s)]}{\sum_{a\in\mathrm{spec}(\rho),b\in\mathrm{spec}(\gamma)} \mathrm{Tr}\ P_aQ_b \cdot b(a/b-1)^2} \\
&\in \left[\min_{a,b} c + \int_{[0,\infty)} \frac{1}{a/b+s}\, d\mu(s),\ \max_{a,b} c + \int_{[0,\infty)} \frac{1}{a/b+s}\, d\mu(s)\right] \\
&= [\nu_f(\lambda_{\max}(\rho)/\lambda_{\min}(\gamma)), \nu_f(\lambda_{\min}(\rho)/\lambda_{\max}(\gamma))]
\end{aligned}$$

□

Corollary 4.4.
For all quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ whose output density operators have full rank,

$$\check{\eta}_f^{\mathrm{std}}(\mathcal{N},\mathcal{M}) \cong \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N},\mathcal{M})$$

Proof. Since $\eta_{f_{\max}}^{\mathrm{std}} \equiv \eta_{\kappa_{\max}}^{\mathrm{Riem}}$, applying Theorems 4.12 and 4.13 gives the result. □

We next proceed to generalising [6, Lemma 4.2] and [6, Proposition 5.4], as this can teach us about the positivity of the relative expansion coefficients:

Lemma 4.5.
Given two density operators $\rho, \gamma \in \mathcal{D}(\mathcal{H})$, some finite-dimensional Hilbert space $\mathcal{H}$, and some $c \in (0,\infty), \kappa \in \mathcal{K}$:

$$\rho \leqslant c\gamma \implies \|X\|_{\kappa,\gamma}^2 \leqslant c\|X\|_{\kappa,\rho}^2 \quad \textit{for all traceless Hermitian operators } X \in \mathcal{B}(\mathcal{H})$$

Proof. By the integral representation of κ, corresponding to the probability measure m on $[0,1]$, in terms of the extreme points $\kappa_s \in \mathcal{K}$:

$$\begin{aligned}\|X\|^2_{\kappa,\rho} &= \int_{[0,1]} \left\langle X, \frac{1}{L_\rho + sR_\rho}(X) \right\rangle_{HS} dm(s) \\ &\geqslant \frac{1}{c}\int_{[0,1]} \left\langle X, \frac{1}{L_\gamma + sR_\gamma}(X) \right\rangle_{HS} dm(s) \\ &= \frac{1}{c}\|X\|^2_{\kappa,\gamma}\end{aligned}$$

where we used the fact that x^{-1} is operator monotone decreasing on $(0,\infty)$ to obtain $\frac{1}{L_\rho+sR_\rho} \geqslant \frac{1}{L_{c\gamma}+sR_{c\gamma}}$.

□

Proposition 4.4.
Suppose we are given quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$, $\mathcal{M} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}'_B)$ and some $\kappa \in \mathcal{K}$, then for all density operators $\rho \in \mathcal{D}(\mathcal{H})$, traceless Hermitian operators $X \in \mathcal{B}(\mathcal{H})$:

$$\lambda^{-1}_{\max}(\mathcal{N})\|\mathcal{N}(X)\|^2_2 \leqslant \|\mathcal{N}(X)\|^2_{\kappa,\,\mathcal{N}(\rho)} \leqslant \lambda^{-1}_{\min}(\mathcal{N})\|\mathcal{N}(X)\|^2_2 \ ,$$

$$\lambda^{-1}_{\max}(\mathcal{M})\|\mathcal{M}(X)\|^2_2 \leqslant \|\mathcal{M}(X)\|^2_{\kappa,\,\mathcal{M}(\rho)} \leqslant \lambda^{-1}_{\min}(\mathcal{M})\|\mathcal{M}(X)\|^2_2$$

In particular, $\check{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N},\mathcal{M}) \geqslant \frac{\lambda_{\min}(\mathcal{M})}{\lambda_{\max}(\mathcal{N})}\check{\eta}_2(\mathcal{N},\mathcal{M})$

Proof. Apply Lemma 4.5 considering

$$\lambda_{\min}(\mathcal{N})I_{d_B} \leqslant \mathcal{N}(\rho) \leqslant \lambda_{\max}(\mathcal{N})I_{d_B}$$

$$\lambda_{\min}(\mathcal{M})I_{d'_B} \leqslant \mathcal{M}(\rho) \leqslant \lambda_{\max}(\mathcal{M})I_{d'_B}$$

and, e.g. $\|\mathcal{N}(X)\|^2_{\kappa,\,I_{d_B}} \equiv \|\mathcal{N}(X)\|^2_2$.

□

Remarks.

1. If $\mathcal{N}$ is injective and $\lambda_{\min}(\mathcal{M}) > 0$, $\check{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N},\mathcal{M}) \geqslant \frac{\lambda_{\min}(\mathcal{M})}{\lambda_{\max}(\mathcal{N})}\check{\eta}_2(\mathcal{N},\mathcal{M}) > 0$
2. [37] conjectured for unital channels that $\hat{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N}) \equiv \hat{\eta}_2(\mathcal{N})$. For unital channels with a unique fixed point (the maximally mixed state), the above result gives something close.

4.7.1 Primitive Channels and Quantum Markov Chains

The ideas in this section can be applied to primitive quantum channels, which are quantum channels $\mathcal{N} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ such that for sufficiently large $m \in \mathbb{N}$, $\mathcal{N}^m$ has only full rank output states, i.e. $\lambda_{\min}(\mathcal{N}^m) > 0$. [55] showed that they are equivalently characterised as quantum channels that have a unique full-rank fixed point, or as quantum channels that have eventually full Kraus rank. The former alternative characterisation makes it quite convenient for Riemannian contraction coefficients to apply to convergence rates of discrete time-inhomogeneous Markov chains based on primitive quantum channels, perhaps because it 'allows easy access to the spectral properties of the map' [59, 19]. In this section, we will see that, conversely, relative expansion coefficients can also be applied to the convergence rates. We obtain this as follows:

Theorem 4.14. *Let $\mathcal{N} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be an injective quantum channel with a unique, full-rank, fixed point π and $\kappa \in \mathcal{K}$. There exists $M \in \mathbb{N}$ s.t. for all $m \geqslant M$:*

$$\check{\eta}_\kappa^{\mathrm{Riem}}(m, \mathcal{N}, \pi) := \inf_{\rho \in \mathcal{D}(\mathcal{H})} \frac{\|\mathcal{N}^m(\rho) - \pi\|_{\kappa,\pi}^2}{\|\mathcal{N}^{m-1}(\rho) - \pi\|_{\kappa,\pi}^2} \geqslant \check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N}^m, \mathcal{N}^{m-1}) > 0 \tag{4.3}$$

Proof.
Since $\mathcal{N}$ is a primitive quantum channel, for m sufficiently large, $\mathcal{N}^{m-1}$ satisfies $\lambda_{\min}(\mathcal{N}^{m-1}) > 0$. By the first remark of Proposition 4.4, we have the result. □

Corollary 4.5. *Let $\mathcal{N} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ be an injective quantum channel with a unique, full-rank, fixed point π and $\kappa \in \mathcal{K}$. There exists $M \in \mathbb{N}$ s.t. for all $m \geqslant M$:*

$$\|\mathcal{N}^m(\rho) - \pi\|_1 \geqslant \lambda_{\min}^{1/2}(\mathcal{N}^m) \eta_\kappa^{\mathrm{Riem}}(M, \mathcal{N}, \pi)^{\frac{m-M+1}{2}} \|\mathcal{N}^{M-1}(\rho) - \pi\|_{\kappa,\,\pi}$$

Proof.

$$\begin{aligned}
\|\mathcal{N}^m(\rho) - \pi\|_1 &\geqslant \|\mathcal{N}^m(\rho) - \pi\|_2 \geqslant \lambda_{\min}^{1/2}(\mathcal{N}^m) \|\mathcal{N}^m(\rho) - \pi\|_{\kappa,\,\pi} \\
&\geqslant \lambda_{\min}^{1/2}(\mathcal{N}^m) (\check{\eta}_\kappa^{\mathrm{Riem}}(m, \mathcal{N}, \pi) \ldots \eta_\kappa^{\mathrm{Riem}}(M, \mathcal{N}, \pi))^{1/2} \|\mathcal{N}^{M-1}(\rho) - \pi\|_{\kappa,\,\pi} \\
&\geqslant \lambda_{\min}^{1/2}(\mathcal{N}^m) \eta_\kappa^{\mathrm{Riem}}(M, \mathcal{N}, \pi)^{\frac{m-M+1}{2}} \|\mathcal{N}^{M-1}(\rho) - \pi\|_{\kappa,\,\pi}
\end{aligned}$$

□

Remark. This result accompanies [19, Proposition 24] and [59, Theorem 9], which demonstrate an exponential upper bound on $\|\mathcal{N}^m(\rho) - \pi\|_1$ in terms of a contraction coefficient. We now have a computable exponential lower bound on the convergence rate.

4.8 Explicit Lower Bounds on the Riemannian Relative Expansion Coefficients

We will now apply some of the theory from previous sections to demonstrate for many important parametrised families of channels on $\mathcal{D}(\mathcal{H})$, $\{\Phi_x\}_{x\in\mathcal{X}}$, $\mathcal{X}$ some valid parameter family, such that for all $x, x'' \in \mathcal{X}$, $\Phi_{x'} \circ \Phi_x = \Phi_{x''}$ for some $x' \in \mathcal{X}$, the channels are perhaps sufficiently similar that we obtain positive relative expansion coefficients $\check{\eta}^{\mathrm{Riem}}(\Phi_x, \Phi_{x''}) > 0$. If $\Phi_{x''} = \Phi_{x'} \circ \Phi_x$ for some $x' \in \mathcal{X}$, this indeed tells us that $\Phi_{x'}$ uniformly preserves a certain proportion of the distinguishability of states in $\Phi_x(\mathcal{D}(\mathcal{H}))$. Except when we are dealing with primitive channels, it is not clear how we can show the positivity of $\check{\eta}_f^{\mathrm{std}}(\Phi_x, \Phi_{x''})$, $f \in \mathcal{F}_{\mathrm{sym}}$, in general. Even for the Riemannian expansion coefficients, we are mostly dealing with qubit channels. In this setting, it is possible to write the Riemannian semi-norms (as we've seen with some examples in Section 2.4) and essentially decompose the problem into a few independent optimisation problems; it helps that all traceless Hermitian operators X acting on qubits satisfy $X^2 \propto I$. This section extends the calculations by [6] for the BKM metric, via an observation that the properties they used for the corresponding Riemannian semi-norm are not unique, but apply to Riemannian semi-norms in general.

4.8.1 Generalised Dephasing Channel

We already know that when a quantum channel $\mathcal{N}$ is primitive, i.e. it has a unique full rank fixed point, there exists another (primitive) quantum channel $\mathcal{M}$ such that

$$\check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{N} \circ \mathcal{M}, \mathcal{M}),\ \check{\eta}_f^{\mathrm{std}}(\mathcal{N} \circ \mathcal{M}, \mathcal{M}) > 0 \quad \text{for all } \kappa \in \mathcal{K},\ f \in \mathcal{F}_{\mathrm{sym}}.$$

The generalised dephasing channels $\Phi_\Gamma : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$, $\dim \mathcal{H} = d$, are parametrised by

$$\Gamma \in \mathcal{X} := \left\{\Gamma' \in \mathcal{B}(\mathcal{H}) : \Gamma' \geqslant 0,\ \Gamma'_{ij} \in [0,1],\ \Gamma'_{ij} = 1,\ 0 \leqslant i,j \leqslant d-1\right\} \setminus \{I_d\},$$

and have the form:

$$\Phi_\Gamma(\rho) = \Gamma \odot \rho = \sum_{i,j=0}^{d-1} \Gamma_{ij}\rho_{ij} \left|ii\right\rangle\left\langle jj\right|$$

As generalisations of the qubit dephasing channels, these channels model decoherence and preserve all density operators that are diagonal w.r.t. the standard basis. Since at least some of the channels' fixed points are not full rank, they are not primitive. To deal with the relative expansion coefficients in this case, we cater to the specific form of these channels, and compare generalised dephasing channels whose parameters are *close*, exactly as [6], did.

To be able to do this, we consider the following generalisation of [6, Lemma 4.3] (note: a slight improvement has been made):

Lemma 4.6. *Given quantum channels $\mathcal{M}, \mathcal{N}, \Phi \in \mathcal{B}(\mathcal{B}(\mathcal{H}_A), \mathcal{B}(\mathcal{H}_B))$ and $\varepsilon \in (0, \frac{1}{2})$ such that:*

$$\mathcal{N} = (1-\varepsilon)\mathcal{M} + \varepsilon\Phi$$

If there exists a quantum channel $\mathcal{D} \in \mathcal{B}(\mathcal{B}(\mathcal{H}_B), \mathcal{B}(\mathcal{H}_B))$ such that $\Phi = \mathcal{D} \circ \mathcal{N}$ and $\mathcal{D}(\omega) \leqslant c\omega$ for some fixed density operator ω and $c > 0$. Then for any operator $X \in \mathcal{B}(\mathcal{H}_A)$ and $\kappa \in \mathcal{K}$:

$$\|\mathcal{N}(X)\|^2_{\kappa,\omega} \geqslant \frac{(1-\varepsilon)(1-2\varepsilon)}{1+c\varepsilon(1-2\varepsilon)} \|\mathcal{M}(X)\|^2_{\kappa,\omega}$$

Proof.
For $X, Y \in \mathcal{B}(\mathcal{H}_B)$, $\operatorname{supp} X$, $\operatorname{supp} Y \leqslant \operatorname{supp} \omega$: $X, Y \mapsto \langle X, R_\omega^{-1}\kappa(L_\omega R_\omega^{-1})(Y)\rangle_{HS}$ defines an inner product, then necessarily $X \mapsto \|X\|^2_{\kappa,\rho}$ defines a genuine norm, and in particular the triangle inequality holds.

$$\therefore \|\mathcal{N}(X)\|_{\kappa,\omega} \geqslant (1-\varepsilon)\|\mathcal{M}(X)\|_{\kappa,\omega} - \varepsilon\|\Phi(X)\|_{k,\omega}$$
$$\therefore \|\mathcal{N}(X)\|^2_{\kappa,\omega} \geqslant (1-\varepsilon)^2\|\mathcal{M}(X)\|^2_{\kappa,\omega} - 2\varepsilon(1-\varepsilon)\|\mathcal{M}(X)\|_{\kappa,\omega}\|\Phi(X)\|_{\kappa,\omega} + \varepsilon^2\|\Phi(X)\|^2_{\kappa,\omega}$$

$$\begin{aligned} &\overset{2ab\leqslant a^2+b^2}{\geqslant} (1-\varepsilon)^2\|\mathcal{M}(X)\|^2_{\kappa,\omega} - \varepsilon(1-\varepsilon)(\|\mathcal{M}(X)\|^2_{\kappa,\omega} + \|\Phi(X)\|^2_{\kappa,\omega}) \\ &\qquad\qquad + \varepsilon^2\|\Phi(X)\|^2_{\kappa,\omega} \\ &= (1-\varepsilon)(1-2\varepsilon)\|\mathcal{M}(X)\|^2_{\kappa,\omega} - \varepsilon(1-2\varepsilon)\|\Phi(X)\|^2_{\kappa,\omega} \end{aligned}$$

$$\begin{aligned} &\overset{\text{Lemma 4.5}}{\geqslant} (1-\varepsilon)(1-2\varepsilon)\|\mathcal{M}(X)\|^2_{\kappa,\omega} - c\varepsilon(1-2\varepsilon)\|\mathcal{D}\circ\mathcal{N}(X)\|^2_{\kappa,\mathcal{D}(\omega)} \\ &\overset{\text{Monotonicity}}{\geqslant} (1-\varepsilon)(1-2\varepsilon)\|\mathcal{M}(X)\|^2_{\kappa,\omega} - c\varepsilon(1-2\varepsilon)\|\mathcal{N}(X)\|^2_{\kappa,\omega} \end{aligned}$$

This gives the result. □

As a consequence, we can achieve the same result as [6] about the possibility of a positive Riemannian expansion coefficient, for general choices $\kappa \in \mathcal{K}$:

Proposition 4.5. *[6]*
Let $\Gamma = (\Gamma_{ij}), \Gamma' = (\Gamma'_{ij}) \in \mathcal{X}$. Suppose there exists $\varepsilon \in (0,1)$ such that

1. $(1-\varepsilon)\Gamma \leqslant \Gamma' \leqslant (1+\varepsilon)\Gamma$
2. $\hat{\Gamma} = (\hat{\Gamma}_{ij})$ *defined via the following is in $\mathcal{X}$ (in particular, we want Γ' PSD):*

$$\hat{\Gamma}_{ij} := \begin{cases} 0, & \text{if } \Gamma'_{ij} = 0, \\ \frac{\Gamma'_{ij} - (1-\varepsilon)\Gamma_{ij}}{\varepsilon \Gamma'_{ij}}, & \text{if } \Gamma'_{ij} > 0. \end{cases}$$

Then for any $\kappa \in \mathcal{K}$:

$$\check{\eta}_\kappa^{\mathrm{Riem}}(\Phi_{\Gamma'}, \Phi_\Gamma) \geqslant \frac{(1-2\varepsilon)(1-\varepsilon)}{(1+2\varepsilon)(1+\varepsilon)}.$$

Proof.
The idea is to apply Lemma 4.6 with $\mathcal{N} = \Phi_{\Gamma'}, \mathcal{M} = \Phi_\Gamma$, because this directly gives a bound $\check{\eta}(\Phi_{\Gamma'}, \Phi_\Gamma) \geqslant \frac{(1-\varepsilon)(1-2\varepsilon)}{(1+\varepsilon)(1+c\varepsilon(1-2\varepsilon))}$ if arbitrary choices of ω in the image $\Phi_{\Gamma'}(\mathcal{D}(\mathcal{H}))$ work.

Let us construct $\widetilde{\Gamma}_{ij} := \frac{\Gamma'_{ij} - (1-\varepsilon)\Gamma_{ij}}{\varepsilon}$, $\widetilde{\Gamma} = (\widetilde{\Gamma}_{ij})_{0 \leqslant i,j \leqslant d-1} \in \mathcal{X}$ ($\widetilde{\Gamma} = \frac{\Gamma' - (1-\varepsilon)\Gamma}{\varepsilon} \geqslant 0$ by assumption), then we have

$$\Phi_{\Gamma'} = (1-\varepsilon)\Phi_\Gamma + \varepsilon \Phi_{\widetilde{\Gamma}},$$

It remains to show that:

1. There exists a quantum channel $\mathcal{D}$ such that $\Phi_{\widetilde{\Gamma}} = \mathcal{D} \circ \Phi_{\Gamma'}$.
2. There exists a universal constant $c > 0$ such that for any density operator γ, $\mathcal{D}(\Phi_{\Gamma'}(\gamma)) \leqslant c\Phi_{\Gamma'}(\gamma)$.

For the first point, we note that $\hat{\Gamma}$ was constructed so that $\hat{\Gamma} \odot \Gamma' = \widetilde{\Gamma}$, which implies:

$$\Phi_{\hat{\Gamma}} \circ \Phi_{\Gamma'} = \Phi_{\widetilde{\Gamma}}.$$

This means that we can take $\mathcal{D} := \Phi_{\widetilde{\Gamma}}$.

For the other condition, we observe that $c = \frac{2}{1-2\varepsilon}$ is a valid choice:

$$\mathcal{D} \circ \Phi_{\Gamma'} = \Phi_{\widetilde{\Gamma}} = \frac{\Phi_{\Gamma'} - (1-\varepsilon)\Phi_\Gamma}{\varepsilon} \leqslant_{\mathrm{cp}} \frac{(1+\varepsilon)\Phi_\Gamma - (1-\varepsilon)\Phi_\Gamma}{\varepsilon} = \frac{2\varepsilon\Phi_\Gamma}{\varepsilon} = 2\Phi_\Gamma \leqslant_{\mathrm{cp}} \frac{2}{1-2\varepsilon}\Phi_{\Gamma'}.$$

Finally, we observe for all $\omega = \Phi_{\Gamma'}(\gamma)$ for some $\gamma \in \mathcal{D}(\mathcal{H})$:

$$\begin{aligned}\|\Phi_{\Gamma'}(X)\|^2_{\kappa,\Phi_{\Gamma'}(\gamma)} &\overset{\text{Lemma 4.6}}{\geqslant} \frac{(1-2\varepsilon)(1-\varepsilon)}{1+2\varepsilon}\|\Phi_{\Gamma}(X)\|^2_{\kappa,\Phi_{\Gamma'}(\gamma)} \\ &\overset{\text{Lemma 4.5}}{\geqslant} \frac{(1-2\varepsilon)(1-\varepsilon)}{(1+2\varepsilon)(1+\varepsilon)}\|\Phi_{\Gamma}(X)\|^2_{\kappa,\Phi_{\Gamma}(\gamma)}.\end{aligned}$$

Where for the final inequality, we used the fact $\Gamma' \leqslant (1+\varepsilon)\Gamma$ (and thus $\Phi'_{\Gamma} \leqslant_{cp} (1+\varepsilon)\Phi_{\Gamma}$). □

4.9 Qubit Calculations

4.9.1 A Case of Strict Inequality between the Riemannian and Divergence Coefficients

The point of this section is to review the proof of an important result in [28] that demonstrates that the Riemannian and divergence coefficients don't have to coincide. To better understand the motivation for this result, some extra elaboration is included in the proof of the first assertion. This result establishes that $\kappa_f(0^+)$ must be sufficiently large for it to be possible that the Riemannian and divergence coefficients are always equal, perhaps again alluding to a connection with $\kappa_f(0^+) = \infty$.

Theorem 4.15. *[28]*
Define $f_s(x) := (x-1)^2\kappa_s(x) = \frac{(1+s)(x-1)^2}{2}(\frac{1}{x+s} + \frac{1}{1+sx})$. *Let* $\Phi_{\alpha,\tau} : \mathcal{B}(\mathbb{C}^2) \to \mathcal{B}(\mathbb{C}^2)$, $\frac{1}{2}(I + \mathbf{w}\cdot\sigma) \mapsto \frac{1}{2}(I + \alpha w_1\sigma_1 + \tau\sigma_3)$. *If* $4\tau^2 > (1-\alpha^2)(4-\alpha^2)$, *then*

$$\hat{\eta}^{\text{std}}_{f_s}(\Phi_{\alpha,\tau}) \geqslant \hat{\eta}^{\text{std}}_{(f_s)_{\text{sym}}}(\Phi_{\alpha,\tau}) = \check{\eta}^{\text{std}}_{(f_s)_{\text{sym}}}(\text{id}_{\mathcal{B}(\mathcal{H}_A)}, \Phi_{\alpha,\tau})^{-1} > \hat{\eta}^{\text{Riem}}_{\kappa_s}(\Phi_{\alpha,\tau}) = \check{\eta}^{\text{Riem}}_{\kappa_s}(\text{id}_{\mathcal{B}(\mathcal{H}_A)}, \Phi_{\alpha,\tau})^{-1}$$

for any $s \leqslant 1$ *sufficiently near 1 (depending on* α, τ*). Moreover, if* $s > \sqrt{\frac{5-\sqrt{21}}{2}}$*, then* $\hat{\eta}^{\text{std}}_{(f_s)_{\text{sym}}}(\Phi_{\alpha,\tau}) \geqslant \check{\eta}^{\text{std}}_{f_{(f_s)_{\text{sym}}}}(\text{id}_{\mathcal{B}(\mathcal{H}_A)}, \Phi_{\alpha,\tau})^{-1} > \hat{\eta}^{\text{Riem}}_{\kappa_s}(\Phi_{\alpha,\tau})$ *when* $\alpha^2 = 1-\tau^2$ *is sufficiently small.*

Proof. For $2\rho =: P = I + \mathbf{w}\cdot\sigma$, $2\gamma =: Q = I + \mathbf{x}\cdot\sigma$ and $0 \leqslant s \leqslant 1$, by Lemma 2.2 we have

$$\begin{aligned}\frac{H_{f_s}(Q,P)}{2(1+s)} &= \frac{1}{2(1+s)}\,\text{Tr}\,(\mathbf{y}\cdot\sigma)\,\frac{1}{L_P + sR_Q}\,(\mathbf{y}\cdot\sigma) \\ &= \left\langle \mathbf{y}, \left[\{(1+s)^2 - |\mathbf{u}|^2\}I + \mathbf{u}\mathbf{u}^T - \mathbf{v}\mathbf{v}^T\right]^{-1}\mathbf{y}\right\rangle,\end{aligned}$$

where $\mathbf{y} = \mathbf{w} - \mathbf{x}$, $\mathbf{u} = \mathbf{w} - s\mathbf{x}$, $\mathbf{v} = \mathbf{w} + s\mathbf{x}$. $\mathbf{y}$ is orthogonal to $\mathbf{u} \times \mathbf{v} = 2s\,\mathbf{w} \times \mathbf{x}$. The formula for $H_{f_s}(P, Q)$ is similar with $\mathbf{w}, \mathbf{x}$ interchanged.

To derive the lower bound $\hat{\eta}^{\text{std}}_{(f_s)_{\text{sym}}}(\Phi_{\alpha,\tau}) > \hat{\eta}^{\text{Riem}}_{\kappa_s}(\Phi_{\alpha,\tau})$, it turns out to be sufficient if we take $\mathbf{w} = (w_1, 0, 0)^t$ and $\mathbf{x} = 0$, for which:

$$\frac{H_{f_s}(Q,P)}{2(1+s)} = \frac{w_1^2}{(1+s)^2 - w_1^2}, \quad \frac{H_{f_s}(P,Q)}{2(1+s)} = \frac{w_1^2}{(1+s)^2 - s^2 w_1^2}.$$

The Bloch vector representations of the output states are $2\Phi_{\alpha,\tau}(\rho) = \Phi_{\alpha,\tau}(P) = I + \tilde{\mathbf{w}} \cdot \sigma$ and $2\Phi_{\alpha,\tau}(\gamma) = \Phi_{\alpha,\tau}(Q) = I + \tilde{\mathbf{x}} \cdot \sigma$ where $\tilde{\mathbf{w}} = (\alpha w_1, 0, \tau)^t$ and $\tilde{\mathbf{x}} = (0, 0, \tau)^t$, so we want to evaluate:

$$\frac{H_{f_s}(\Phi_{\alpha,\tau}(Q), \Phi_{\alpha,\tau}(P))}{2(1+s)} = \left\langle \tilde{\mathbf{y}}, \left[\{(1+s)^2 - |\tilde{\mathbf{u}}|^2\} I + \tilde{\mathbf{u}}\tilde{\mathbf{u}}^T - \tilde{\mathbf{v}}\tilde{\mathbf{v}}^T\right]^{-1} \tilde{\mathbf{y}} \right\rangle$$

where $\tilde{\mathbf{y}} = \tilde{\mathbf{w}} - \tilde{\mathbf{x}} = (\alpha w_1, 0, 0)^t$, $\tilde{\mathbf{u}} = \tilde{\mathbf{w}} - s\tilde{\mathbf{x}} = (\alpha w_1, 0, (1-s)\tau)^t$, and $\tilde{\mathbf{v}} = \tilde{\mathbf{w}} + s\tilde{\mathbf{x}} = (\alpha w_1, 0, (1+s)\tau)^t$.

The matrix form of $\left[\{(1+s)^2 - |\tilde{\mathbf{u}}|^2\} I + \tilde{\mathbf{u}}\tilde{\mathbf{u}}^T - \tilde{\mathbf{v}}\tilde{\mathbf{v}}^T\right]$ is

$$\begin{pmatrix} \xi_s(\tau^2) - \alpha^2 w_1^2 & 0 & -2s\alpha\tau w_1 \\ 0 & \xi_s(\tau^2) - \alpha^2 w_1^2 & 0 \\ -2s\alpha\tau w_1 & 0 & \xi_s(\tau^2) - 4s\tau^2 - \alpha^2 w_1^2 \end{pmatrix},$$

where we recall that $\xi_s(x) := (1+s)^2(1-x) + 4sx$. We only care about the (1,1)-entry of the inverse of this matrix, since $\tilde{\mathbf{y}} \propto \mathbf{e}_3$ (so just cross the last two rows and divide by the determinant to derive). Therefore, the exact form is:

$$\frac{H_{f_s}(\Phi_{\alpha,\tau}(Q), \Phi_{\alpha,\tau}(P))}{2(1+s)} = \frac{\alpha^2 w_1^2 [\xi_s(\tau^2) - \alpha^2 w_1^2 - 4s\tau^2]}{[\xi_s(\tau^2) - \alpha^2 w_1^2][\xi_s(\tau^2) - \alpha^2 w_1^2 - 4s\tau^2] - 4s^2\alpha^2\tau^2 w_1^2}.$$

Similarly,

$$\frac{H_{f_s}(\Phi_{\alpha,\tau}(P), \Phi_{\alpha,\tau}(Q))}{2(1+s)} = \frac{\alpha^2 w_1^2 [\xi_s(\tau^2) - s^2\alpha^2 w_1^2 - 4s\tau^2]}{[\xi_s(\tau^2) - s^2\alpha^2 w_1^2][\xi_s(\tau^2) - s^2\alpha^2 w_1^2 - 4s\tau^2] - 4s^2\alpha^2\tau^2 w_1^2}.$$

Define:

$$H(s) \equiv \lim_{|w_1| \to 1} \frac{H_{f_s}(Q,P) + H_{f_s}(P,Q)}{2(1+s)},$$
$$\widetilde{H}(s) \equiv \lim_{|w_1| \to 1} \frac{H_{f_s}(\Phi_{\alpha,\tau}(Q), \Phi_{\alpha,\tau}(P)) + H_{f_s}(\Phi_{\alpha,\tau}(P), \Phi_{\alpha,\tau}(Q))}{2(1+s)}.$$

We already have $\hat{\eta}^{\mathrm{RelEnt}}_{f_s}(\Phi_{\alpha,\tau}) \geqslant \frac{H_{f_s}(\Phi_{\alpha,\tau}(\gamma),\Phi_{\alpha,\tau}(\rho))+H_{f_s}(\Phi_{\alpha,\tau}(\rho),\Phi_{\alpha,\tau}(\gamma))}{H_{f_s}(\gamma,\rho)+H_{f_s}(\rho,\gamma)} = \widetilde{H}(s)/H(s)$ for every $s \in [0,1]$. Note that

$$\hat{\eta}^{\mathrm{Riem}}_{\kappa_s}(\Phi_{\alpha,\tau}) = \frac{(1+s)^2\alpha^2}{(1+s)^2-(1-s)^2\tau^2} = \frac{\alpha^2}{1-\left(\frac{1-s}{1+s}\right)^2\tau^2} \to \alpha^2 \text{ as } s \to 1$$

And observe that $H(1) = 2/3$ and

$$\widetilde{H}(1) = \frac{2\alpha^2(4-\alpha^2-4\tau^2)}{(4-\alpha^2)(4-\alpha^2-4\tau^2)-4\alpha^2\tau^2} = \frac{2\alpha^2(4-\alpha^2-4\tau^2)}{(4-\alpha^2)^2-16\tau^2}.$$

The denominator $(4-\alpha^2)^2-16\tau^2$ is positive ($\because \alpha^2+\tau^2 \leqslant 1$ and $\tau^2 < 1$). Assume that $4\tau^2 > (1-\alpha^2)(4-\alpha^2)$ (in particular, $\alpha^2 > 0$). To ensure $\widetilde{H}(1)/H(1) > \alpha^2$, we must enforce:

$$\frac{3}{2}\cdot 2(4-\alpha^2-4\tau^2) - \left[(4-\alpha^2)^2-16\tau^2\right] = 4\tau^2-(1-\alpha^2)(4-\alpha^2) > 0,$$

Thus we arrive at the first assertion stated in the theorem, via the continuity of $\hat{\eta}^{\mathrm{Riem}}_{\kappa_s}(\Phi_{\alpha,\tau})$ w.r.t. s.

For the second assertion, when $\alpha^2+\tau^2 = 1$, a tedious computation gives

$$\frac{\widetilde{H}(s)}{H(s)} = \alpha^2 \cdot \frac{s(s+2)(2s+1)\left[12s(s+1)^2+(2s^4+s^3+s+2)\alpha^2\right]}{(s^2+4s+1)\left[4s(s+1)+s^3\alpha^2\right]\left[4s(s+1)+\alpha^2\right]}.$$

The limit of $\left[\widetilde{H}(s)/H(s)\right]/\hat{\eta}^{\mathrm{Riem}}_{\kappa_s}(\Phi_{\alpha,\tau})$ as $\alpha^2 = 1-\tau^2 \to 0$ is

$$\frac{3s(s+2)(2s+1)}{(s^2+4s+1)(s+1)^2}.$$

The numerator minus the denominator of the above ratio is $-(s^4-5s^2+1)$, which is positive when $s^2 > \frac{5-\sqrt{21}}{2}$. This yields the second assertion of the theorem.

□

Remark. This $\Phi_{\alpha,\tau}$ is necessarily non-unital, since $\hat{\eta}^{\mathrm{std}}_f(\mathcal{N}) = \hat{\eta}^{\mathrm{Riem}}_\kappa(\mathcal{N}) = \hat{\eta}_{\mathrm{Tr}}(\mathcal{N})^2$ for all unital qubit channels [28, 37].

Theorem 4.16 (Inequivalence of Bounded vs. Unbounded Cases)**.** *If $\kappa \in \mathcal{K}$ is bounded, then:*

$$\frac{1}{\alpha^2}\hat{\eta}_\kappa^{\mathrm{Riem}}(\Phi_{\alpha,\sqrt{1-\alpha^2}}) < \infty \tag{4.4}$$

If instead $\kappa \in \mathcal{K}$ is unbounded, then:

$$\frac{1}{\alpha^2}\hat{\eta}_\kappa^{\mathrm{Riem}}(\Phi_{\alpha,\sqrt{1-\alpha^2}}) = \infty \tag{4.5}$$

In particular, if $\kappa \in \mathcal{K}$ is bounded and $\kappa' \in \mathcal{K}$ is unbounded,

$$\check{\eta}_\kappa^{\mathrm{Riem}} \not\asymp \check{\eta}_{\kappa'}^{\mathrm{Riem}}$$

Proof.
The bounded case (4.4) is immediately deduced via $\eta_{\kappa_{\min}}^{\mathrm{Riem}}(\Phi_{\alpha,\sqrt{1-\alpha^2}}) = \alpha^2$ [28, Theorem 6.2] and Theorem 4.7. We henceforth consider only the unbounded setting of κ.

For $\rho = \frac{I+\mathbf{w}\cdot\sigma}{2}$, $X = \mathbf{y}\cdot\sigma$, $\xi_s(x) := (1+s)^2 - (1-s)^2 x = (1+s)^2(1-x) + 4sx$. We will use the following results (the latter is Lemma 2.2):

$$\kappa(x) = \int_{[0,1]} \kappa_s(x)\, dm(s),\ \|X\|^2_{\kappa_s,\rho} = \frac{4(1+s)^2}{\xi_s(|\mathbf{w}|^2)}\left[|\mathbf{y}|^2 + \frac{4s(\mathbf{w}\cdot\mathbf{y})^2}{(1+s)^2(1-|\mathbf{w}|^2)}\right]$$

We can apply these results to the image of $\Phi_{\alpha,\sqrt{1-\alpha^2}}$, recalling $\Phi_{\alpha,\sqrt{1-\alpha^2}} : \frac{1}{2}(I+\mathbf{w}\cdot\sigma) \mapsto \frac{1}{2}(I + \alpha w_1\sigma_1 + \sqrt{1-\alpha^2}\,\sigma_3)$, and that $\xi_s(x) := (1+s)^2 - (1-s)^2 x$ is a decreasing function with $\xi_s(1) = 4s$:

$$\begin{aligned}\|\Phi_{\alpha,\sqrt{1-\alpha^2}}(X)\|^2_{\kappa,\Phi(\rho)} &= \int \frac{4(1+s)^2}{\xi_s(1-\alpha^2(1-w_1^2))}\cdot\alpha^2\left(y_1^2 + \frac{4s(w_1y_1)^2}{(1+s)^2(1-w_1^2)}\right) dm(s)\\ &\geqslant \int \frac{(1+s)^2}{s}\cdot\alpha^2 y_1^2\left(1 + \frac{4sw_1^2}{(1+s)^2(1-w_1^2)}\right) dm(s)\end{aligned}$$

To compute a lower bound on the contraction coefficient $\eta_\kappa^{\mathrm{Riem}}(\Phi_{\alpha,\sqrt{1-\alpha^2}})$, WLOG we take $y_1 = 1$, which makes $\|\Phi_{\alpha,\sqrt{1-\alpha^2}}(X)\|^2_{\kappa,\Phi(\rho)}$ independent of $\mathbf{y}$. We therefore proceed to minimise and upper bound $\|X\|^2_{\kappa_s,\rho}$ w.r.t. $\mathbf{y}$ for fixed $\mathbf{w}$:

$$\min_{\mathbf{y}:y_1=1} \frac{4(1+s)^2}{\xi_s(|\mathbf{w}|^2)}\left[|\mathbf{y}|^2 + \frac{4s(\mathbf{w}\cdot\mathbf{y})^2}{(1+s)^2(1-|\mathbf{w}|^2)}\right] = \frac{\mu(\mu+\nu|\mathbf{w}|^2)}{\mu+\nu(|\mathbf{w}|^2 - w_1^2)}$$

where $\mu := \int \frac{4(1+s)^2}{\xi(|\mathbf{w}|^2)}\, dm(s), \nu := \int \frac{16}{\xi_s(|\mathbf{w}^2|)} \cdot \frac{s}{1-|\mathbf{w}|^2}\, dm(s)$. Note that $\mu + \nu|\mathbf{w}|^2 = \frac{4}{1-|\mathbf{w}|^2}$, and thus:

$$\frac{\mu \cdot \frac{4}{1-|\mathbf{w}|^2}}{\frac{4}{1-|\mathbf{w}|^2} - \nu \cdot w_1^2} = \frac{4\mu}{4 - \nu w_1^2(1-|\mathbf{w}|^2)} = \frac{\int \frac{16(1+s)^2}{\xi_s(|\mathbf{w}|^2)}\, dm(s)}{4 - w_1^2 \int \frac{16s}{\xi_s(|\mathbf{w}|^2)}\, dm(s)}$$
$$\leqslant \sup_s \frac{16(1+s)^2}{4\xi_s(|\mathbf{w}^2|) - w_1^2 \cdot 16s} = \sup_s \frac{4}{1-w_1^2} = \frac{4}{1-w_1^2}$$

Now, we can establish a lower bound on $\eta_\kappa^{\text{Riem}}(\Phi_{\alpha,\sqrt{1-\alpha^2}})$:

$$\sup_{\mathbf{w},\mathbf{y}:|\mathbf{w}|<1, y_1=1} \frac{\|\Phi_{\alpha,\sqrt{1-\alpha^2}}(X)\|^2_{\kappa,\Phi_{\alpha,\sqrt{1-\alpha^2}}(\rho)}}{\|X\|^2_{\kappa,\rho}} \geqslant \sup_{\mathbf{w}:|\mathbf{w}|<1} \frac{\int \frac{(1+s)^2}{s} \cdot \alpha^2 \left(1 + \frac{4sw_1^2}{(1+s)^2(1-w_1^2)}\right)\, dm(s)}{\frac{\mu(\mu+\nu|\mathbf{w}|^2)}{\mu+\nu(|\mathbf{w}|^2-w_1^2)}}$$
$$\geqslant \frac{\alpha^2}{4} \sup_{|w_1|<1} \int \frac{(1+s)^2}{s}(1 - w_1^2 + \frac{4sw_1^2}{(1+s)^2})\, dm(s)$$
$$= \frac{\alpha^2}{4} \sup_{|w_1|<1} \int \frac{(1+s)^2}{s}(1 - \frac{(1-s)^2}{(1+s)^2} w_1^2)\, dm(s)$$
$$= \frac{\alpha^2}{4} \int \frac{(1+s)^2}{s}\, dm(s) = \frac{\alpha^2}{2}\kappa(0^+)$$

This gives the result in the unbounded case:

$$\frac{1}{\alpha^2}\hat{\eta}_\kappa^{\text{Riem}}(\Phi_{\alpha,\sqrt{1-\alpha^2}}) \geqslant \frac{\kappa(0^+)}{2} = \infty$$

□

4.9.2 Dephasing Channel and Amplitude Damping Channel

For the dephasing and amplitude damping channels, some care needs to be taken around their pure fixed points, and this makes it difficult to compute a lower bound on the relative expansion coefficients. However, the strategy for showing that the relative expansion coefficients are positive is similar for both channels. Recall that for $s \in [0, 1]$, density operator

$\rho = \frac{1}{2}(\mathbb{I}_2 + \mathbf{w}\cdot\sigma) \in \mathcal{D}(\mathbb{C}^2)$, traceless Hermitian operator $X = \mathbf{y}\cdot\sigma \in \mathcal{B}(\mathbb{C}^2)$ (2.2),

$$\begin{aligned}\|X\|^2_{\kappa_s,\rho} &= \frac{2|\mathbf{y}|^2}{1-|\mathbf{w}|^2}\cdot\frac{(1+s^2)(1-|\mathbf{w}|^2)+4s|\mathbf{w}|^2\cos^2\theta}{(1+s^2)(1-|\mathbf{w}|^2)+4s|\mathbf{w}|^2}\\ &= \frac{2|\mathbf{y}|^2}{1-|\mathbf{w}|^2}\left(h_s(|\mathbf{w}|^2)+(1-h_s(|\mathbf{w}|^2))\cos^2\theta\right)\\ &= \|X\|^2_{\kappa_{\max},\rho}\left(h_s(|\mathbf{w}|^2)+(1-h_s(|\mathbf{w}|^2))\cos^2\theta\right)\end{aligned}$$

where $h_s(x) := \frac{(1+s)^2(1-x)}{(1+s)^2(1-x)+4sx}$. Since every $\kappa \in \mathcal{K}$ has the form $\kappa(x) = \int_{[0,1]}\kappa_s(x)\,dm(s)$ for all $x \in (0,\infty)$, we generally have:

$$\|X\|^2_{\kappa,\rho} = \|X\|^2_{\kappa_{\max},\rho}(h(|\mathbf{w}|^2)+(1-h(|\mathbf{w}|^2))\cos^2\theta), \quad h(x) := \int_{[0,1]} h_s(x)\,dm(s)$$

and $h(|\mathbf{w}|^2) = 0 \iff |\mathbf{w}| = 1,$ otherwise $h(|\mathbf{w}|^2) \in (0,1]$ (as $|\mathbf{w}| < 1$).

If we are considering two qubit quantum channels

$$\mathcal{N} : \frac{1}{2}(\mathbb{I}_2 + \mathbf{w}\cdot\sigma) \mapsto \frac{1}{2}(\mathbb{I}_2 + (T\mathbf{w}+\mathbf{t})\cdot\sigma)\ , \quad \mathcal{M} : \frac{1}{2}(\mathbb{I}_2 + \mathbf{w}\cdot\sigma) \mapsto \frac{1}{2}(\mathbb{I}_2 + (T'\mathbf{w}+\mathbf{t}')\cdot\sigma)$$

the problem of lower bounding the relative Riemannian expansion coefficient reduces slightly:

$$\check{\eta}^{\mathrm{Riem}}_{\kappa}(\mathcal{N},\mathcal{M}) \geqslant \check{\eta}^{\mathrm{Riem}}_{\kappa_{\max}}(\mathcal{N},\mathcal{M})\cdot\inf_{\mathbf{y},\mathbf{w}:|\mathbf{w}|\leqslant 1}\frac{h_{\mathcal{N}}(\mathbf{w},\mathbf{y})}{h_{\mathcal{M}}(\mathbf{w},\mathbf{y})}$$

where we denote $h_{\mathcal{N}}(\mathbf{w},\mathbf{y}) = |(T\mathbf{w}+\mathbf{t})\cdot T\mathbf{y}/|T\mathbf{y}||^2(1-h(|T\mathbf{w}+\mathbf{t}|^2))+|T\mathbf{w}+\mathbf{t}|^2h(|T\mathbf{w}+\mathbf{t}|^2)$, and similarly for $\mathcal{M}$.

Proposition 4.6. *Let $\Phi_p(\rho) = \frac{1}{2}(\mathbb{I}_2 + T_p\mathbf{w}\cdot\sigma)$, $T_p := \mathrm{diag}(1-p,1-p,1)$, denote the dephasing channel. For $0 < p_2 < p_1 < 1$ and any $\kappa \in \mathcal{K}$, we have*

$$\check{\eta}^{\mathrm{Riem}}_{\kappa}(\Phi_{p_1},\Phi_{p_2}) > 0$$

Proof.
Denote $\mathbf{w}_p = T_p\mathbf{w}$, $\mathbf{y}_p = T_p\mathbf{y}, \theta_p := \cos^{-1}\frac{|\mathbf{w}_p\cdot\mathbf{y}_p|}{|\mathbf{w}_p\|\mathbf{y}_p|}$ and $S_\varepsilon := B(\mathbf{e}_3,\varepsilon)^c \cap B(-\mathbf{e}_3,\varepsilon)^c$. For any

$\varepsilon > 0$, $c(p, \varepsilon) := \inf_{\mathbf{w} \in S_\varepsilon} h(|\mathbf{w}_p|^2) > 0$ since S_ε is compact (so the infimum is attained). Thus for any $\mathbf{w} \in S_\varepsilon$:

$$\frac{h_{\Phi_{p_1}}(\mathbf{w}, \mathbf{y})}{h_{\Phi_{p_2}}(\mathbf{w}, \mathbf{y})} = \frac{\cos^2 \theta_{p_1} + \sin^2 \theta_{p_1} h(|\mathbf{w}_{p_1}|^2)}{\cos^2 \theta_{p_2} + \sin^2 \theta_{p_2} h(|\mathbf{w}_{p_2}|^2)} \geqslant c(p_1, \varepsilon) > 0$$

Now consider $\mathbf{w} \mapsto \pm \mathbf{e}_3 \cdot$ WLOG we consider $|\mathbf{y}| = 1$, and define (by minimisation of a continuous function over a compact set):

$$\mathbf{y}(\mathbf{w}) := \arg\min_{|\mathbf{y}|=1} \frac{h_{\Phi_{p_1}}(\mathbf{w}, \mathbf{y})}{h_{\Phi_{p_2}}(\mathbf{w}, \mathbf{y})}$$

and $\mathbf{y}_{p_i}(\mathbf{w}) := T_{p_i} \mathbf{y}(\mathbf{w})$.

To simplify the expression for $\mathbf{y}(\mathbf{w})$:

$$\begin{aligned}
|\mathbf{w}_{p_2} \cdot \mathbf{y}_{p_2}|^2 &= |(1 - p_2)^2 (y_1 w_1 + y_2 w_2) + y_3 w_3|^2 \\
&= |\mathbf{w}_{p_1} \cdot \mathbf{y}_{p_1} + ((1 - p_2)^2 - (1 - p_1)^2)(y_1 w_1 + y_2 w_2)|^2 \\
&\leqslant 2|\mathbf{w}_{p_1} \cdot \mathbf{y}_{p_1}|^2 + 2((1 - p_2)^2 - (1 - p_1)^2)^2 |y_1 w_1 + y_2 w_2|^2 \\
&\overset{C-S}{\leqslant} 2|\mathbf{w}_{p_1} \cdot \mathbf{y}_{p_1}|^2 + 2(2 - p_1 - p_2)^2 (p_1 - p_2)^2 (y_1^2 + y_2^2)(w_1^2 + w_2^2) \\
&\leqslant 2|\mathbf{w}_{p_1} \cdot \mathbf{y}_{p_1}|^2 + \frac{2(2 - p_1 - p_2)^2 (p_1 - p_2)^2}{(1 - p_2)^2} \cdot |\mathbf{y}_{p_2}|^2 \cdot \frac{|\mathbf{w}|^2 - |\mathbf{w}_{p_2}|^2}{p_2(2 - p_2)} \\
&\leqslant 2|\mathbf{w}_{p_1} \cdot \mathbf{y}_{p_1}|^2 + \frac{2(2 - p_1 - p_2)^2 (p_1 - p_2)^2}{(1 - p_2)^2 p_2 (2 - p_2)} \cdot |\mathbf{y}_{p_2}|^2 \cdot (1 - |\mathbf{w}_{p_2}|^2) \\
&\leqslant 2|\mathbf{w}_{p_1} \cdot \mathbf{y}_{p_1}|^2 + \frac{2(2 - p_1 - p_2)^2 (p_1 - p_2)^2}{(1 - p_2)^2 p_2 (2 - p_2)} \cdot |\mathbf{y}_{p_2}|^2 \cdot h(|\mathbf{w}_{p_2}|^2)
\end{aligned}$$

$$\begin{aligned}
\because \qquad |\mathbf{y}_{p_2}|^2 &\geqslant (\mathbf{y}_{p_2})_1^2 + (\mathbf{y}_{p_2})_2^2 = (1-p_2)^2(y_1^2+y_2^2)\\
1-|\mathbf{w}_{p_2}|^2 &= 1-(1-p_2)^2(w_1^2+w_2^2)-w_3^2\\
&= 1-|\mathbf{w}|^2+(1-(1-p_2)^2(w_1^2+w_2^2))\\
&= 1-|\mathbf{w}|^2+p_2(2-p_2)(w_1^2+w_2^2)\\
h(x) \geqslant h_1(x) &= \frac{4(1-x)}{4(1-x)+4x} = 1-x \quad \forall x\in[0,1]
\end{aligned}$$

Define $c'(p_1,p_2) := \frac{(2-p_1-p_2)^2(p_1-p_2)^2}{2(1-p_2)^2p_2(2-p_2)}$, then:

$$\begin{aligned}
&\liminf_{\substack{\mathbf{w}\to\pm\mathbf{e}_3,\\|\mathbf{w}|\leqslant 1}} \inf_{|\mathbf{y}|=1} \frac{h_{\Phi_{p_1}}(\mathbf{w},\mathbf{y})}{h_{\Phi_{p_2}}(\mathbf{w},\mathbf{y})}\\
&\quad\geqslant \liminf_{\substack{\mathbf{w}\to\pm\mathbf{e}_3,\\|\mathbf{w}|\leqslant 1}} \inf_{|\mathbf{y}|=1} \frac{|\mathbf{w}_{p_1}\cdot\mathbf{y}_{p_1}/|\mathbf{y}_{p_1}||^2(1-h(|\mathbf{w}_{p_1}|^2))+|\mathbf{w}_{p_1}|^2h(|\mathbf{w}_{p_1}|^2)}{2|\mathbf{w}_{p_1}\cdot\mathbf{y}_{p_1}/|\mathbf{y}_{p_2}||^2(1-h(|\mathbf{w}_{p_2}|^2))+(|\mathbf{w}_{p_2}|^2+c'(p_1,p_2))h(|\mathbf{w}_{p_2}|^2)}\\
&\geqslant \min\left\{\liminf_{\substack{\mathbf{w}\to\pm\mathbf{e}_3,\\|\mathbf{w}|\leqslant 1}} \inf_{|\mathbf{y}|=1} \frac{|\mathbf{y}_{p_2}|^2}{2|\mathbf{y}_{p_1}|^2}\cdot\frac{1-h(|\mathbf{w}_{p_1}|^2)}{1-h(|\mathbf{w}_{p_2}|^2)}, \frac{1}{1+c'(p_1,p_2)} \liminf_{\substack{\mathbf{w}\to\pm\mathbf{e}_3,\\|\mathbf{w}|\leqslant 1}} \frac{h(|\mathbf{w}_{p_1}|^2)}{h(|\mathbf{w}_{p_2}|^2)}\right\}\\
&= \min\left\{\frac{1}{2},\ \frac{1}{1+c'(p_1,p_2)}\right\} > 0
\end{aligned}$$

$\therefore$ for $\varepsilon>0$ sufficiently small, $\inf_{\mathbf{w}\in B(\mathbf{e}_3,\varepsilon)\cup B(-\mathbf{e}_3,\varepsilon)} \frac{h_{\Phi_{p_1}}(\mathbf{w},\mathbf{y})}{h_{\Phi_{p_2}}(\mathbf{w},\mathbf{y})} > 0$.

Now, we only have to show $\check{\eta}^{\mathrm{Riem}}_{\kappa_{\max}}(\Phi_{p_1},\Phi_{p_2}) > 0$:

$$\begin{aligned}
\check{\eta}^{\mathrm{Riem}}_{\kappa_{\max}}(\Phi_{p_1},\Phi_{p_2}) &\geqslant \inf_{\mathbf{y}:|\mathbf{y}_{p_1}|=1}\frac{|\mathbf{y}_{p_1}|^2}{|\mathbf{y}_{p_2}|^2}\cdot\inf_{\mathbf{w}:|\mathbf{w}|\leqslant 1}\frac{1-|\mathbf{w}_{p_2}|^2}{1-|\mathbf{w}_{p_1}|^2}\\
&= \left(\frac{1-p_1}{1-p_2}\right)^2\cdot\inf_{\mathbf{w}:|\mathbf{w}|\leqslant 1}\frac{1-|\mathbf{w}|^2+p_2(2-p_2)(w_1^2+w_2^2)}{1-|\mathbf{w}|^2+p_1(2-p_1)(w_1^2+w_2^2)}\\
&\geqslant \left(\frac{1-p_1}{1-p_2}\right)^2\min\left\{1,\frac{p_2(2-p_2)}{p_1(2-p_1)}\right\} > 0
\end{aligned}$$

□

Proposition 4.7. *Let* $\mathcal{A}_\gamma(\rho) := \frac{1}{2}(\mathbb{I}_2 + (T_\gamma \mathbf{w} + \mathbf{t}_\gamma)\cdot\sigma)$, $T_\gamma = \mathrm{diag}(\sqrt{1-\gamma}, \sqrt{1-\gamma}, 1-\gamma)$, $t_\gamma = \gamma\mathbf{e}_3$, *denote the amplitude damping channel. For* $0 < \gamma_2 < \gamma_1 < 1$, *and any* $\kappa \in \mathcal{K}$ *we have*

$$\check{\eta}_\kappa^{\mathrm{Riem}}(\mathcal{A}_{\gamma_1}, \mathcal{A}_{\gamma_2}) > 0.$$

Proof.
This time, denote

$$\mathbf{w}_\gamma := T_\gamma \mathbf{w} + \mathbf{t}_\gamma = (\sqrt{1-\gamma}w_1, \sqrt{1-\gamma}w_2, (1-\gamma)w_3 + \gamma)$$

$$\mathbf{y}_\gamma := T_\gamma \mathbf{y}_1, (\sqrt{1-\gamma}y_1, \sqrt{1-\gamma}y_2, (1-\gamma)y_3),$$

$$\theta_\gamma := \cos^{-1}\frac{|\mathbf{w}_\gamma \cdot \mathbf{y}_\gamma|}{|\mathbf{w}_\gamma||\mathbf{y}_\gamma|},$$

and $S_\varepsilon := B(\mathbf{e}_3, \varepsilon)^c$.

For any $\varepsilon > 0, c(\gamma, \varepsilon) := \inf_{\mathbf{w}\in S_\varepsilon} h(|\mathbf{w}_\gamma|^2) > 0$ since S_ε is compact (so the infimum is attained). Thus for any $\mathbf{w} \in S_\varepsilon$:

$$\frac{h_{\mathcal{A}_{\gamma_1}}(\mathbf{w}, \mathbf{y})}{h_{\mathcal{A}_{\gamma_2}}(\mathbf{w}, \mathbf{y})} = \frac{\cos^2\theta_{\gamma_1} + \sin^2\theta_{\gamma_1} h(|\mathbf{w}_{\gamma_1}|^2)}{\cos^2\theta_{\gamma_2} + \sin^2\theta_{\gamma_2} h(|\mathbf{w}_{\gamma_2}|^2)} \geqslant c(\gamma_1, \varepsilon) > 0$$

Now consider $\mathbf{w} \to \pm\mathbf{e}_3$, WLOG we consider $|\mathbf{y}| = 1$, and define (by the minimisation of a continuous function over a compact set)

$$\mathbf{y}(\mathbf{w}) := \underset{|\mathbf{y}|=1}{\mathrm{argmin}} \frac{h_{\mathcal{A}_{\gamma_1}}(\mathbf{w}, \mathbf{y})}{h_{\mathcal{A}_{\gamma_2}}(\mathbf{w}, \mathbf{y})}$$

and $\mathbf{y}_{\gamma_i}(\mathbf{w}) := T_{\gamma_i}\mathbf{y}(\mathbf{w})$

To simplify the expression for $\mathbf{y}(\mathbf{w})$:

$$
\begin{aligned}
|\mathbf{w}_{\gamma_2} \cdot \mathbf{y}_{\gamma_2}|^2 &= |(1-\gamma_2)(\mathbf{w} \cdot \mathbf{y} + \gamma_2 y_3 (1-w_3))|^2 \\
&= \left(\frac{1-\gamma_2}{1-\gamma_1}\right)^2 |(1-\gamma_1)(\mathbf{w} \cdot \mathbf{y}) + \gamma_1 y_3(1-w_3) + (\gamma_2 - \gamma_1) y_3 (1-w_3))|^2 \\
&\overset{|a+b|^2 \leqslant 2(|a|^2+|b|^2)}{\leqslant} (\frac{1-\gamma_2}{1-\gamma_1})^2 (|\mathbf{w}_{\gamma_1} \cdot \mathbf{y}_{\gamma_1}|^2 + (1-\gamma_2)^2 (\gamma_1 - \gamma_2)^2 y_3^2 (1-w_3)^2) \\
&\leqslant 2 \left(\frac{1-\gamma_2}{1-\gamma_1}\right)^2 (|\mathbf{w}_{\gamma_1} \cdot \mathbf{y}_{\gamma_1}|^2 + \left(\frac{\gamma_1 - \gamma_2}{\gamma_2}\right)^2 |\mathbf{y}_{\gamma_2}|^2 (|\mathbf{w}|^2 - |\mathbf{w}_{\gamma_2}|^2)) \\
&\leqslant 2 \left(\frac{1-\gamma_2}{1-\gamma_1}\right)^2 (|\mathbf{w}_{\gamma_1} \cdot \mathbf{y}_{\gamma_1}|^2 + \left(\frac{\gamma_1 - \gamma_2}{\gamma_2}\right)^2 |\mathbf{y}_{\gamma_2}|^2 (1 - |\mathbf{w}_{\gamma_2}|^2)) \\
&\leqslant 2 \left(\frac{1-\gamma_2}{1-\gamma_1}\right)^2 (|\mathbf{w}_{\gamma_1} \cdot \mathbf{y}_{\gamma_1}|^2 + \left(\frac{\gamma_1 - \gamma_2}{\gamma_2}\right)^2 |\mathbf{y}_{\gamma_2}|^2 h(|\mathbf{w}_{\gamma_2}|^2))
\end{aligned}
$$

$$
\begin{aligned}
\therefore |\mathbf{y}_{\gamma_2}|^2 &\geqslant (1-\gamma_2)^2 y_3^2 \\
1 - |\mathbf{w}_{\gamma_2}|^2 &= 1 - (1-\gamma_2)(w_1^2 + w_2^2) + ((1-\gamma_2) w_3 + \gamma_2)^2 \\
&= 1 - (1-\gamma_2)(w_1^2 + w_2^2) + (w_3 + \gamma_2 (1-w_3))^2 \\
&= 1 - |\mathbf{w}|^2 + \gamma_2 (w_1^2 + w_2^2) + 2\gamma_2 w_3 (1-w_3) + \gamma_2^2 (1-w_3)^2 \\
&\geqslant 1 - |\mathbf{w}|^2 + \gamma_2^2 (1-w_3)^2 \\
h(x) \geqslant h_1(x) &= \frac{4(1-x)}{4(1-x) + 4x} = 1 - x \quad \forall x \in [0,1]
\end{aligned}
$$

Define $c_1'(\gamma_1, \gamma_2) = 2\left(\frac{1-\gamma_2}{1-\gamma_1}\right)^2$ and $c_2'(\gamma_1, \gamma_2) = 2\left(\frac{\gamma_1 - \gamma_2}{\gamma_2}\right)^2$, then:

$$\liminf_{\substack{\mathbf{w}\to\mathbf{e}_3,\\ |\mathbf{w}|\leqslant 1}} \inf_{|\mathbf{y}|=1} \frac{h_{\mathcal{A}_{\gamma_1}}(\mathbf{w},\mathbf{y})}{h_{\mathcal{A}_{\gamma_2}}(\mathbf{w},\mathbf{y})}$$

$$= \liminf_{\substack{\mathbf{w}\to\mathbf{e}_3,\\ |\mathbf{w}|\leqslant 1}} \inf_{|\mathbf{y}|=1} \frac{|\mathbf{w}_{\gamma_1}\cdot\mathbf{y}_{\gamma_1}/|\mathbf{y}_{\gamma_1}||^2(1-h(|\mathbf{w}_{\gamma_1}|^2)) + |\mathbf{w}_{\gamma_1}|^2 h(|\mathbf{w}_{\gamma_1}|^2)}{c_1'(\gamma_1,\gamma_2)|\mathbf{w}_{\gamma_1}\cdot\mathbf{y}_{\gamma_1}/|\mathbf{y}_{\gamma_2}||^2(1-h(|\mathbf{w}_{\gamma_2}|^2)) + (|\mathbf{w}_{\gamma_2}|^2 + c_2'(\gamma_1,\gamma_2))h(|\mathbf{w}_{\gamma_2}|^2)}$$

$$\geqslant \min\left\{\liminf_{\substack{\mathbf{w}\to\mathbf{e}_3,\\ |\mathbf{w}|\leqslant 1}} \inf_{|\mathbf{y}|=1} \frac{1}{c_1'(\gamma_1,\gamma_2)}\cdot\frac{|\mathbf{y}_{\gamma_2}|^2}{|\mathbf{y}_{\gamma_1}|^2}\cdot\frac{1-h(|\mathbf{w}_{\gamma_1}|^2)}{1-h(|\mathbf{w}_{\gamma_2}|^2)}, \frac{1}{1+c_2'(\gamma_1,\gamma_2)} \liminf_{\substack{\mathbf{w}\to\mathbf{e}_3,\\ |\mathbf{w}|\leqslant 1}} \inf_{|\mathbf{y}|=1} \frac{h(|\mathbf{w}_{\gamma_1}|^2)}{h(|\mathbf{w}_{\gamma_2}|^2)}\right\}$$

$$\geqslant \min\left\{\frac{1}{c_1'(\gamma_1,\gamma_2)},\ \frac{1}{1+c_2'(\gamma_1,\gamma_2)}\right\} > 0$$

$\therefore$ for $\varepsilon > 0$ sufficiently small,

$$\inf_{\mathbf{w}\in B(\mathbf{e}_3,\varepsilon)} \frac{h_{\mathcal{A}_{\gamma_1}}(\mathbf{w},\mathbf{y})}{h_{\mathcal{A}_{\gamma_2}}(\mathbf{w},\mathbf{y})} > 0$$

Now, we only have to show $\check{\eta}^{\mathrm{Riem}}_{\kappa_{\max}}(\mathcal{A}_{\gamma_1},\mathcal{A}_{\gamma_2}) > 0$:

$$\begin{aligned}
\check{\eta}^{\mathrm{Riem}}_{\kappa_{\max}}(\mathcal{A}_{\gamma_1},\mathcal{A}_{\gamma_2}) &\geqslant \inf_{\mathbf{y}:|\mathbf{y}|=1} \frac{|\mathbf{y}_{\gamma_1}|^2}{|\mathbf{y}_{\gamma_2}|^2} \inf_{\mathbf{w}:|\mathbf{w}|\leqslant 1} \frac{1-|\mathbf{w}_{\gamma_2}|^2}{1-|\mathbf{w}_{\gamma_1}|^2} \\
&= \left(\frac{1-\gamma_1}{1-\gamma_2}\right)^2 \cdot \inf_{\mathbf{w}:|\mathbf{w}|\leqslant 1} \frac{1-|\mathbf{w}_{\gamma_2}|^2}{1-|\mathbf{w}_{\gamma_1}|^2} \\
&= \left(\frac{1-\gamma_1}{1-\gamma_2}\right)^2 \cdot \inf_{\mathbf{w}:|\mathbf{w}|\leqslant 1} \frac{(1-\gamma_2)(1-|\mathbf{w}|^2+\gamma_2(w_3-1)^2)}{(1-\gamma_1)(1-|\mathbf{w}|^2+\gamma_1(w_3-1)^2)} \\
&\geqslant \left(\frac{1-\gamma_1}{1-\gamma_2}\right) \min\left\{1, \frac{\gamma_2}{\gamma_1}\right\} > 0
\end{aligned}$$

□

Chapter 5

Conclusion

This study of the relative expansion coefficient has hopefully illustrated to the reader that it is a rather subtle object. From the beginning, the optimisation regions that define the coefficients were selected carefully to capture the lack of reverse data processing on the set of all input states (Section 4.3) w.r.t. an f-divergence or Riemannian metric. However, we saw that the injective primitive channels provide cases where we can have a restricted reverse data processing inequality (for sufficiently large $m \in \mathbb{N}$): $\|\mathcal{N}(X)\|^2_{\kappa, \mathcal{N}(\rho)} \geqslant \eta \|X\|^2_{\kappa, \rho}, \eta > 0$, for all $\rho \neq \gamma \in \mathcal{N}^{m-1}(\mathcal{D}_d)$ (this generalises to arbitrary D_f via the equivalence result Corollary 4.4). This re-established a connection with quantum Markov chains that has been previously studied only in the context of Riemannian contraction coefficients [59, 19, 20].

Alternatively, two cases were explored where restricting the set of quantum channels recovers (at least in part) the equivalence between the divergence and Riemannian expansion coefficients that holds indefinitely in the classical case, namely: QC channels (Section 4.5) and quantum channels that have only full rank output states (Section 4.7). Note that it is the generalisation from the contraction coefficient to the relative expansion coefficient that makes these results possible (as we have full control over $\mathcal{N}$ and $\mathcal{M}$).

If we are interested in the general equivalence between $\check{\eta}_f^{\text{std}}$ and $\check{\eta}_{\kappa_f}^{\text{Riem}}$ over all channels, I have suggested a possible connection with $\kappa_f(0^+) = \infty$. This makes sense from the perspective that, uniquely (Theorem 4.11), the maximal metric coincides with the standard maximal f-divergence, so it is reasonable to suppose that a similarity between κ and $\kappa_{\max}$ is important. This was solidified via explicit constructions of this equivalence using $\kappa_{\max}$ (Theorem 4.9) and from the observation that for bounded $\kappa_f, \kappa_{f'} \in \mathcal{K}$: $\check{\eta}_f^{\text{std}} \cong \check{\eta}_{f'}^{\text{std}}$,

$\check{\eta}^{\mathrm{Riem}}_{\kappa_f} \cong \check{\eta}^{\mathrm{Riem}}_{\kappa_{f'}}$ (Theorems 4.7 and 4.8). The latter point does not definitively exclude the possibility of equivalence, but it does enforce a clear distinction between the bounded and unbounded cases.

Finally, this thesis has also made significant progress on an open problem posed in [6] about whether their results could be generalised to other information measures. Via the convenient integral representation (2.7) for $\mathcal{K}$, I recognised that the techniques they used were in fact valid for any Riemannian metric (i.e. not only the BKM metric). In Section 4.8, their results on the positivity of the Riemannian relative expansion coefficients for various classes of channels were found in the general case.

References

[1] G. Androulakis and T. C. John. *"Quantum f-divergences via Nussbaum–Szkoła distributions and applications to f-divergence inequalities"*. Reviews in Mathematical Physics **36**(09): 2360002, (2024).

[2] R. Araiza, M. Junge, and P. Wu. *"Transportation cost and contraction coefficient for channels on von Neumann algebras"*, (2025). Available online: https://arxiv.org/abs/2506.04197.

[3] K. M. Audenaert, J. Calsamiglia, R. Munoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete. *"Discriminating states: The quantum Chernoff bound"*. Physical review letters **98**(16): 160501, (2007).

[4] J. Aujla, S. S. Dragomir, M. Khosravi, and M. Moslehian. *"Refinements of Choi–Davis–Jensen's inequality"*. Bull. Math. Anal. Appl **3**(2): 127–133, (2011).

[5] S. Beigi. *"Sandwiched Rényi divergence satisfies data processing inequality"*. Journal of Mathematical Physics **54**(12) (2013).

[6] P. Belzig, L. Gao, G. Smith, and P. Wu. *"Reverse-type Data Processing Inequality"*, (2024). Available online: https://arxiv.org/abs/2411.19890.

[7] M. Berta, V. B. Scholz, and M. Tomamichel. *"Rényi divergences as weighted non-commutative vector-valued L p-spaces"*. In *Annales Henri Poincaré*, volume 19, pages 1843–1867, (2018).

[8] M. Beth Ruskai, S. Szarek, and E. Werner. *"An analysis of completely-positive trace-preserving maps on M2"*. Linear Algebra and its Applications **347**(1): 159–187 (2002).

[9] F. Buscemi. *"Degradable channels, less noisy channels, and quantum statistical morphisms: An equivalence relation"*. Problems of Information Transmission **52**(3): 201–213 (2016).

[10] M.-D. Choi. *"A schwarz inequality for positive linear maps on C^*-algebras"*. Illinois Journal of Mathematics **18**(4): 565–574, (1974).

[11] M.-D. Choi, M. B. Ruskai, and E. Seneta. *"Equivalence of certain entropy contraction coefficients"*. Linear algebra and its applications **208**: 29–36, (1994).

[12] J. E. Cohen, Y. Iwasa, G. Rautu, M. B. Ruskai, E. Seneta, and G. Zbaganu. *"Relative entropy under mappings by stochastic matrices"*. Linear algebra and its applications **179**: 211–235, (1993).

[13] E. B. Davies. *Quantum Theory of Open Systems*. Academic Press, London (1976).

[14] C. Davis. *"A Schwarz inequality for convex operator functions"*. Proceedings of the American Mathematical Society **8**(1): 42–44, (1957).

[15] R. L. Frank and E. H. Lieb. *"Monotonicity of a relative Rényi entropy"*. Journal of Mathematical Physics **54**(12), (2013).

[16] J. Fujii. *"Jensen's inequalities on any interval for operators"*. Proc. of the 3rd Int. Cof. on Nonlinear Analysis and Convex Analysis, 2004 , (2004).

[17] L. Gao, M. Junge, N. LaRacuente, and H. Li. *"Complete positivity order and relative entropy decay"*. In *Forum of Mathematics, Sigma*, volume 13, page e31, (2025).

[18] L. Gao and C. Rouzé. *"Complete Entropic Inequalities for Quantum Markov Chains"*. Archive for Rational Mechanics and Analysis **245**(1): 183–238 (2022).

[19] I. George and M. Tomamichel. *"A Unified Approach to Quantum Contraction and Correlation Coefficients"*. arXiv preprint arXiv:2505.15281 , (2025).

[20] I. George, A. Zheng, and A. Bansal. *"Divergence Inequalities with Applications in Ergodic Theory"*, (2024). Available online: `https://arxiv.org/abs/2411.17241`.

[21] M. Hayashi. *"Quantum hypothesis testing and discrimination of quantum states"*. In *Quantum Information Theory: Mathematical Foundation*, pages 95–153. Springer (2016).

[22] C. W. Helstrom. *"Quantum detection and estimation theory"*. Journal of Statistical Physics **1**(2): 231–252, (1969).

[23] F. Hiai. *"Quantum f-divergences in von Neumann Algebras"*. Math. Phys. Stud., Springer, Singapore (2021).

[24] F. Hiai, H. Kosaki, D. Petz, and M. B. Ruskai. *"Families of completely positive maps associated with monotone metrics"*. Linear Algebra and its Applications **439**(7): 1749–1791, (2013).

[25] F. Hiai and M. Mosonyi. *"Different quantum f-divergences and the reversibility of quantum operations"*. Reviews in Mathematical Physics **29**(07): 1750023, (2017).

[26] F. Hiai, M. Mosonyi, D. Petz, and C. Bény. *"Quantum f-divergences and error correction"*. Reviews in Mathematical Physics **23**(07): 691–747 (2011).

[27] F. Hiai and D. Petz. *"The proper formula for relative entropy and its asymptotics in quantum probability"*. Communications in Mathematical Physics **143**: 99–114 (1991).

[28] F. Hiai and M. B. Ruskai. *"Contraction coefficients for noisy quantum channels"*. Journal of Mathematical Physics **57**(1): 015211 (2015).

[29] F. Hiai and M. B. Ruskai. *"Contraction coefficients for noisy quantum channels"*. Journal of Mathematical Physics **57**(1) (2015).

[30] C. Hirche, C. Rouzé, and D. Stilck França. *"On contraction coefficients, partial orders and approximation of capacities for quantum channels"*. Quantum **6**: 862 (2022).

[31] C. Hirche and M. Tomamichel. *"Quantum Rényi and f-Divergences from Integral Representations"*. Communications in Mathematical Physics **405**(9) (2024).

[32] A. S. Holevo. *"Investigations in the general theory of statistical decisions"*. Trudy Matematicheskogo Instituta imeni VA Steklova **124**: 3–140, (1976).

[33] I. D. Ivanovic. *"How to differentiate between non-orthogonal states"*. Physics Letters A **123**(6): 257–259, (1987).

[34] A. Jenčová. *"Rényi relative entropies and noncommutative L p-spaces"*. In *Annales Henri Poincaré*, volume 19, pages 2513–2542, (2018).

[35] A. Jenčová. *"Rényi relative entropies and noncommutative L p-spaces II"*. In *Annales Henri Poincaré*, volume 22, pages 3235–3254, (2021).

[36] M. Junge, R. Renner, D. Sutter, M. M. Wilde, and A. Winter. *"Universal recoverability in quantum information"*. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2494–2498, (2016).

[37] A. Lesniewski and M. B. Ruskai. *"Monotone Riemannian metrics and relative entropy on noncommutative probability spaces"*. Journal of Mathematical Physics **40**(11): 5702–5724 (1999).

[38] K. Li. *"Discriminating quantum states: The multiple Chernoff distance"*. The Annals of Statistics **44**(4) (2016).

[39] K. Li and Y. Yao. *"Operational interpretation of the sandwiched Rényi divergence of order 1/2 to 1 as strong converse exponents"*. Communications in Mathematical Physics **405**(2): 22, (2024).

[40] G. Lindblad. *"Completely positive maps and entropy inequalities"*. Communications in Mathematical Physics **40**: 147–151 (1975).

[41] K. Löwner. *"Über monotone matrixfunktionen"*. Mathematische Zeitschrift **38**(1): 177–216, (1934).

[42] K. Matsumoto. *"A new quantum version of f-divergence"*. In *Nagoya Winter Workshop: Reality and Measurement in Algebraic Quantum Theory*, pages 229–273, (2015).

[43] E. A. Morozova and N. N. Chentsov. *"Markov invariant geometry on state manifolds"*. Itogi Nauki i Tekhniki. Seriya" Sovremennye Problemy Matematiki. Noveishie Dostizheniya" **36**: 69–102, (1989).

[44] A. Müller-Hermes and D. Reeb. *"Monotonicity of the Quantum Relative Entropy Under Positive Maps"*. Annales Henri Poincaré **18**(5): 1777–1788 (2017).

[45] M. Nussbaum and A. Szkoła. *"The Chernoff lower bound for symmetric quantum hypothesis testing"*. The Annals of Statistics **37**(2) (2009).

[46] A. Parusiński and A. Rainer. *"Perturbation theory of polynomials and linear operators"*. In *Handbook of Geometry and Topology of Singularities VII*, pages 121–202. Springer (2025).

[47] D. Petz. *"Quasi-entropies for finite quantum systems"*. Reports on mathematical physics **23**(1): 57–65, (1986).

[48] D. Petz. *"Sufficient subalgebras and the relative entropy of states of a von Neumann algebra"*. Communications in mathematical physics **105**(1): 123–131, (1986).

[49] D. Petz. *"Monotone metrics on matrix spaces"*. Linear algebra and its applications **244**: 81–96, (1996).

[50] D. Petz. *“Monotonicity of quantum relative entropy revisited”*. Reviews in Mathematical Physics **15**(01): 79–91 (2003).

[51] D. Petz and M. B. Ruskai. *“Contraction of generalized relative entropy under stochastic mappings on matrices”*. Infinite Dimensional Analysis, Quantum Probability and Related Topics **1**(01): 83–89, (1998).

[52] D. Petz and C. Sudár. *“Extending the Fisher metric to density matrices”*. Geometry in Present Day Science **21**, (1999).

[53] Y. Polyanskiy. *“Lecture notes on f-divergences”*, (2020).

[54] F. Rellich. *Perturbation theory of eigenvalue problems.* CRC Press (1969).

[55] M. Sanz, D. Perez-Garcia, M. M. Wolf, and J. I. Cirac. *“A quantum version of Wielandt's inequality”*. IEEE Transactions on Information Theory **56**(9): 4668–4673, (2010).

[56] B. Schumacher and M. D. Westmoreland. *“Relative entropy in quantum information theory”*. arXiv preprint quant-ph/0004045 , (2000).

[57] G. Smith and P. Wu. *“Additivity of quantum capacities in simple non-degradable quantum channels”*. IEEE Transactions on Information Theory (2025).

[58] M. Tehranchi. *“Lecture notes on Advanced Stochastic Models”*, (2020).

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

[60] A. Uhlmann. *“Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory”*. Communications in Mathematical Physics **54** (1977).

[61] 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”*. Communications in Mathematical Physics **331**(2): 593–622, (2014).

[62] H. K. Wimmer. *“Rellich's perturbation theorem on hermitian matrices of holomorphic functions”*. Journal of mathematical analysis and applications **114**(1): 52–54, (1986).

[63] E. Zeidler. *Applied functional analysis: applications to mathematical physics.* volume 108, Springer Science & Business Media (2012).

APPENDICES

Appendix A

Local Symmetry of quantum f-divergences

It turns out that all standard f-divergences are locally symmetric, even if $f \notin \mathcal{F}$, and we only require that f is continuously twice differentiable. This proof is based on [53], which dealt with only the classical f-divergence.

Theorem A.1 (Standard Quantum f-divergences are Locally Symmetric)**.** *Let f be a continuously twice differentiable function such that $\lim_{x\to\infty} \frac{f(x)}{x^2} < \infty$, then for any density operator ρ and traceless Hermitian operator X, with $\operatorname{supp} X \leqslant \operatorname{supp} \rho$, acting on the finite d-dimensional Hilbert space $\mathcal{H}$, we define $\rho_\varepsilon := \rho + \varepsilon X \geqslant 0$ for $\varepsilon \in \mathbb{R}$ sufficiently small. Then*

$$D_f^{\mathrm{std}}(\rho_\varepsilon \| \rho) = D_f^{\mathrm{std}}(\rho \| \rho_\varepsilon) + o(\varepsilon^2) \tag{A.1}$$

Proof.
We can find an open neighbourhood $I \subseteq \mathbb{R}$ of 0 such that $\rho_\varepsilon \geqslant 0 \ \forall \varepsilon \in I$. From Rellich's Theorem Version I (Theorem C.1), we know that the real eigenvalues $\lambda_1(\varepsilon), ..., \lambda_d(\varepsilon)$ and their corresponding orthogonal eigenvectors $|\phi_1(\varepsilon)\rangle, ..., |\phi_d(\varepsilon)\rangle$ of ρ_ε are analytic in some neighbourhood $I' \subseteq I$ of 0. i.e.

$$\begin{aligned}
|\phi_i(\varepsilon)\rangle &= |\phi_i^{(0)}\rangle + \varepsilon|\phi_i^{(1)}\rangle + \varepsilon^2|\phi_i^{(2)}\rangle + ... \\
\lambda_i(\varepsilon) &= \lambda_i^{(0)} + \varepsilon\lambda_i^{(1)} + \varepsilon^2\lambda_i^{(2)} + ... \\
\rho_\varepsilon|\phi_i(\varepsilon)\rangle &= \lambda_i(\varepsilon)|\phi_i(\varepsilon)\rangle
\end{aligned}$$

We only want the eigenvectors of ρ_ε to be orthogonal, because then it is possible to conveniently take WLOG $\langle\phi_i^{(0)}|\phi_i^{(k)}\rangle = 0\ \forall i,k$, but the eigenvectors of ρ are chosen to be orthonormal. Also note that ρ and ρ_ε are allowed to have degenerate eigenvalues here.

Denote $|\phi_i\rangle := |\phi_i^{(0)}\rangle, \lambda_i := \lambda_i^{(0)}$. If we define $P_i(\varepsilon) := \frac{|\phi_i(\varepsilon)\rangle\langle\phi_i(\varepsilon)|}{\langle\phi_i(\varepsilon)|\phi_i(\varepsilon)\rangle} = \frac{P_i'(\varepsilon)}{\langle\phi_i(\varepsilon)|\phi_i(\varepsilon)\rangle}$ and $P_i := |\phi_i\rangle\langle\phi_i|$, then these rank-1 projections are also analytic in ε, so $\lim_{\varepsilon\to 0} P_i(\varepsilon) = P_i$.

By the definition of the standard f-divergence:

$$D_f^{\mathrm{std}}(\rho_\varepsilon\|\rho) = \sum_{ij}\lambda_j f\left(\frac{\lambda_i(\varepsilon)}{\lambda_j}\right)\mathrm{Tr}\ P_i(\varepsilon)P_j$$

WLOG, we have $f(1) = f'(1) = 0$. This is because we can otherwise choose $\widetilde{f}, \widetilde{f}(x) = f(x) - f'(1)(x-1)$, which satisfies $D_f^{\mathrm{std}} \equiv D_{\widetilde{f}}, \|X\|^2_{\kappa_f,\rho} = \|X\|^2_{\kappa_{\widetilde{f}},\rho}$.

The proof now proceeds by (degenerate) perturbation theory. Consider the following:

- $P_i'(\varepsilon) = |\phi_i\rangle\langle\phi_i| + \varepsilon(|\phi_i^{(1)}\rangle\langle\phi_i| + |\phi_i\rangle\langle\phi_i^{(1)}|) + \varepsilon^2(|\phi_i^{(2)}\rangle\langle\phi_i|) + |\phi_i^{(1)}\rangle\langle\phi_i^{(1)}| + |\phi_i\rangle\langle\phi_i^{(2)}|) + o(\varepsilon^2)$
- $\langle\phi_i(\varepsilon)|\phi_j(\varepsilon)\rangle = \langle\phi_i(\varepsilon)|\phi_i(\varepsilon)\rangle\delta_{ij} = \delta_{ij} + \varepsilon(\langle\phi_i^{(1)}|\phi_j\rangle + \langle\phi_i|\phi_j^{(1)}\rangle) + o(\varepsilon)$
  $\implies \langle\phi_i^{(1)}|\phi_j\rangle = -\langle\phi_i|\phi_j^{(1)}\rangle$ for $i \neq j$
- By the continuous twice differentiability of f, $f(x) = (x-1)^2\int_0^1(1-t)f''(1+t(x-1))dt$
- $\mathrm{Tr}\ P_iP_j = \delta_{ij}$ and

$$\begin{aligned}\mathrm{Tr}\ P_i'(\varepsilon)P_j =&\mathrm{Tr}\ |\phi_i\rangle\langle\phi_i||\phi_j\rangle\langle\phi_j| + \varepsilon\mathrm{Tr}\ (|\phi_i^{(1)}\rangle\langle\phi_i| + |\phi_i\rangle\langle\phi_i^{(1)}|)|\phi_j\rangle\langle\phi_j|)\\ &+ \varepsilon^2\mathrm{Tr}\ ((|\phi_i^{(2)}\rangle\langle\phi_i| + |\phi_i^{(1)}\rangle\langle\phi_i^{(1)}| + |\phi_i\rangle\langle\phi_i^{(2)}|)|\phi_j\rangle\langle\phi_j|) + o(\varepsilon^2)\\ =&\delta_{ij} + \varepsilon^2\mathrm{Tr}\ |\langle\phi_i^{(1)}|\phi_j\rangle|^2 + o(\varepsilon^2)\end{aligned}$$

$$\begin{aligned}D_f^{\mathrm{std}}(\rho_\varepsilon\|\rho) &= \sum_{ij}\lambda_j f\left(\frac{\lambda_i(\varepsilon)}{\lambda_j}\right)\mathrm{Tr}\ P_i(\varepsilon)P_j\\ &= \sum_i \frac{\lambda_i}{\langle\phi_i(\varepsilon)|\phi_i(\varepsilon)\rangle}f\left(\frac{\lambda_i(\varepsilon)}{\lambda_i}\right) + \varepsilon^2\sum_{i\neq j}\frac{\lambda_j}{\langle\phi_i(\varepsilon)|\phi_i(\varepsilon)\rangle}f\left(\frac{\lambda_i(\varepsilon)}{\lambda_j}\right)|\langle\phi_i^{(1)}|\phi_j\rangle|^2 + o(\varepsilon^2)\end{aligned}$$

$f\left(\frac{\lambda_i(\varepsilon)}{\lambda_i}\right) = (\frac{\lambda_i(\varepsilon)-\lambda_i}{\lambda_i})^2 \int_0^1 (1-t)f''(1+t\left(\frac{\lambda_i(\varepsilon)-\lambda_i}{\lambda_i}\right))dt$

Since $\lambda_i(\varepsilon) \to \lambda_i$ as $\varepsilon \to 0\ \forall i$, $\exists \delta > 0$ s.t. $|\frac{\lambda_i(\varepsilon)-\lambda_i}{\lambda_i}| < \frac{1}{2}\ \forall \varepsilon \in (-\delta, \delta) \forall i$ and $(-\delta, \delta) \subseteq I'$

Note that by the fact $\lim_{x\to\infty} \frac{f(x)}{x^2} < \infty$, $\sup_{x\in[\frac{1}{2},\infty)} f''(x) < C$ for some constant $C > 0$. By the dominated convergence theorem:

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} \left(\frac{\lambda_i(\varepsilon) - \lambda_i}{\lambda_i}\right)^2 \int_0^1 (1-t)f''(1+t\left(\frac{\lambda_i(\varepsilon)-\lambda_i}{\lambda_i}\right))dt$$
$$= (\lambda_i^{(1)})^2 f''(1) \int_0^1 1 - t\ dt = \frac{1}{2}(\lambda_i^{(1)})^2 f''(1)$$

And thus we finally obtain:

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} D_f^{\text{std}}(\rho_\varepsilon \| \rho) = \frac{f''(1)}{2} \sum_i \lambda_i (\lambda_i^{(1)})^2 + \sum_{i\neq j} \lambda_j f\left(\frac{\lambda_i}{\lambda_j}\right) |\langle \phi_i^{(1)} | \phi_j \rangle|^2$$

But note that the transpose $\tilde{f}(x) := xf(x^{-1})$ satisfies $\tilde{f}''(1) = f''(1)$ and $\lambda_j f\left(\frac{\lambda_i}{\lambda_j}\right) = \lambda_i \tilde{f}\left(\frac{\lambda_j}{\lambda_i}\right)$, which means that:

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} D_f^{\text{std}}(\rho_\varepsilon \| \rho) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} D_{\tilde{f}}^{\text{std}}(\rho_\varepsilon \| \rho) \equiv \lim_{\varepsilon\to 0} \frac{1}{\varepsilon^2} D_f^{\text{std}}(\rho \| \rho_\varepsilon)$$

□

Appendix B

Lemmas for the DPI of Sandwiched Rényi Divergences

Theorem (Araki-Lieb-Thirring):
For $A, B \geqslant 0$, $q \geqslant 0$, and $0 \leqslant r \leqslant 1$, the following inequality holds:

$$\mathrm{Tr}\left[(A^r B^r A^r)^q\right] \leqslant \mathrm{Tr}\left[(ABA)^{rq}\right]$$

Theorem (Riesz-Thorin theorem for $L_{\rho,\sigma}$ spaces):
Let $\Phi : \mathcal{L}(\mathcal{H}) \to \mathcal{L}(\mathcal{H}')$ be a linear superoperator. Assume that $1 \leqslant p_0 \leqslant p_1 \leqslant \infty$ and $1 \leqslant q_0 \leqslant q_1 \leqslant \infty$. Let $0 \leqslant \theta \leqslant 1$ and define p_θ and q_θ as $p_\theta := (\theta p_0^{-1} + (1-\theta)p_1^{-1})^{-1}$, $q_\theta := ((1-\theta)q_0^{-1} + \theta q_1^{-1})^{-1}$. Finally, assume $\sigma \in \mathcal{L}(\mathcal{H})$, $\sigma' \in \mathcal{L}(\mathcal{H}')$ are positive definite. Then we have:

$$\|\Pi\|_{(p_\theta,\sigma)\to(q_\theta,\sigma')} \leqslant \|\Phi\|^{1-\theta}_{(p_0,\sigma)\to(q_0,\sigma')} \|\Phi\|^{\theta}_{(p_1,\sigma)\to(q_1,\sigma')}$$

Theorem (Russo-Dye):
Let $\mathcal{A}$ and $\mathcal{B}$ be C^*-algebras with units, and let $\Phi : \mathcal{A} \to \mathcal{B}$ be a positive map. Then $\|\Phi\| = \|\Phi(1)\|$

Appendix C

Some Perturbation Theory

C.1 Rellich's Theorem

Rellich's theorem formalises our intuition about time-dependent perturbation theory.

Theorem C.1 (Rellich Version I). *[54, 46]*
Let $H(\varepsilon) \in \mathcal{B}(\mathcal{H})$, $dim\mathcal{H} = d < \infty$, be a family of Hermitian operators whose entries (in any chosen basis) are (real) analytic in an open neighbourhood $I \subseteq \mathbb{R}$ of $\varepsilon = 0$. Then the eigenvalues and eigenvectors of $H(\varepsilon)$ can be chosen to be real analytic in ε.

More precisely, there exist real-valued functions $\lambda_i(\varepsilon)$ and complex-valued d-vectors $v_i(\varepsilon)$,
$i = 1, ..., d$, depending analytically on ε, such that for ε in an open neighbourhood $I' \subseteq \mathbb{R}$ of $\varepsilon = 0$:

1. *The eigenvectors $|\psi_1(\varepsilon)\rangle, ..., |\psi_d(\varepsilon)\rangle$ form an orthonormal basis*
2. *The eigenvalues $\lambda_i(t)$ are real $\quad \forall i \in \{1, ..., d\}$*
3. *$H(\varepsilon)|\psi_i(\varepsilon)\rangle = \lambda_i(\varepsilon)|\psi_i(\varepsilon)\rangle \quad \forall i \in \{1, ..., d\}$*

Theorem C.2 (Rellich Version II). *[54, 62]*
Let $H(z) \in \mathcal{B}(\mathcal{H})$, $dim\mathcal{H} = d < \infty$, be a family of Hermitian operators whose entries (in any chosen basis) are (complex) analytic in an open neighbourhood $\Omega \subseteq \mathbb{C}$ of $z = 0$. Then the unitary $U(z)$ and diagonal matrix $D(z)$ of $H(z)$'s eigendecomposition can be chosen to be real analytic in z.

More precisely, there exists a unitary $U(z)$ *and diagonal matrix* $D(z)$ *depending analytically on* z*, such that for* z *in an open neighbourhood* $J \subseteq \Omega$ *of* $z = 0$*:*

1. $H(z) = U^*(z)D(z)U(z)$
2. $U^*(z)U(z) = \mathbb{I}$

Appendix D

Proof of Lemma 4.2

The proof proceeds by a pigeonhole principle argument. By definition, there exists a pure state $|\varphi_1\rangle$ such that $\operatorname{rank} \mathcal{N}(|\varphi_1\rangle\langle\varphi_1|) \geqslant 2$. Then we can extend $|\varphi_1\rangle$ to an orthonormal basis $\{|\varphi_i\rangle\}_{1 \leqslant i \leqslant d_A}$ of $\mathcal{H}_A$, and define a chain of orthogonal projections:

$$P_k = \sum_{i=1}^{k} |\varphi_i\rangle\langle\varphi_i| \,,\; 1 \leqslant k \leqslant d_A \,.$$

As $\mathcal{N}$ is a positive map, $\operatorname{supp} \mathcal{N}(P_k) \leqslant \operatorname{supp} \mathcal{N}(P_{k+1})$, for all $1 \leqslant k \leqslant d_A - 1$, and so

$$2 \leqslant \operatorname{rank} \mathcal{N}(P_1) \leqslant \operatorname{rank} \mathcal{N}(P_2) \leqslant \;...\; \leqslant \operatorname{rank} \mathcal{N}(P_{d_A}) \leqslant d_B \leqslant d_A$$

By the pigeonhole principle, i.e. we have d_A ranks between 2 and d_A, there exists $1 \leqslant k_0 \leqslant d_A - 1$ such that $\operatorname{rank} \mathcal{N}(P_{k_0}) = \operatorname{rank} \mathcal{N}(P_{k_0+1})$, which means that $\operatorname{supp} \mathcal{N}(P_{k_0}) = \operatorname{supp} \mathcal{N}(P_{k_0+1}) = \operatorname{supp} \mathcal{N}(P_{k_0} + |\varphi_{k_0+1}\rangle\langle\varphi_{k_0+1}|)$. Thus, we choose $P_A = P_{k_0}$, $|\psi\rangle = |\varphi_{k_0+1}\rangle$ and this gives the result.