UTS UNIVERSITY
OF TECHNOLOGY
SYDNEY

RÉNYI DIVERGENCE INEQUALITIES VIA INTERPOLATION, WITH APPLICATIONS TO GENERALISED ENTROPIC UNCERTAINTY RELATIONS

by

Alexander McKinlay

THESIS SUBMITTED IN FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science (Research) in Computing Sciences

under the supervision of

Dr. Marco Tomamichel
Dr. Christopher Ferrie

UNIVERSITY OF TECHNOLOGY SYDNEY

Faculty of Engineering and Information Technology

January, 2021

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Alexander McKinlay, declare that this thesis is submitted in fulfilment of the requirements for the award of Master of Science (Research) in Computing Sciences, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Production Note:

Signature: Signature removed prior to publication.

Date: 13/05/2021

Acknowledgments

This thesis could not have happened without the consistent and invaluable support of Marco.

The inspiration for the interpolation technique was supplied by Dr. Salman Beigi, whose contributions in discussions about the nature of Pisier norms were more than illuminating.

Lucie, for her patience.

This is a CONVENTIONAL THESIS. Published work arising from research during the candidature of the degree has been included in an appendix.

DECLARATION OF PUBLICATIONS INCLUDED IN THE THESIS

Title	:	Decomposition Rules for Quantum Rényi Mutual Information with an Application to Information Exclusion Relations
Parts included	:	Subsections IV.A and IV.B
Authors	:	Alexander McKinlay, Marco Tomamichel
Contributions	:	AM derived results and developed the method of the proofs in consultation with MT.
Status	:	Published
Journal	:	Journal of Mathematical Physics, vol. 61, no. 7, p. 072202, 2020
Address	:	https://doi.org/10.1063/1.5143862

Contents

List of tables

ABSTRACT

**RÉNYI DIVERGENCE INEQUALITIES VIA INTERPOLATION, WITH
APPLICATIONS TO GENERALISED ENTROPIC UNCERTAINTY RELATIONS**

by

Alexander McKinlay

We investigate quantum Rényi entropic quantities, specifically those derived from 'sandwiched' divergence. This divergence is one of several proposed Rényi generalisations of the quantum relative entropy. We may define Rényi generalisations of the quantum conditional entropy and mutual information in terms of this divergence, from which they inherit many desirable properties. However, these quantities lack some of the convenient structure of their Shannon and von Neumann counterparts. We attempt to bridge this gap by establishing divergence inequalities for valid combinations of Rényi order which replicate the chain and decomposition rules of Shannon and von Neumann entropies. Although weaker in general, these inequalities recover equivalence when the Rényi parameters tend to one.

To this end we present Rényi mutual information decomposition rules, a new approach to the Rényi conditional entropy tripartite chain rules and a more general bipartite comparison. The derivation of these results relies on a novel complex interpolation approach for general spaces of linear operators.

These new comparisons allow us to employ techniques that until now were only available for Shannon and von Neumann entropies. We can therefore directly apply them to the derivation of Rényi entropic uncertainty relations. Accordingly, we establish a family of Rényi information exclusion relations and provide further generalisations and improvements to this and other known relations, including the Rényi bipartite uncertainty relations.

Introduction

This thesis contributes some concrete results in the form of inequalities for quantum Rényi entropies and a direct application to determining Rényi entropy uncertainty relations. However, these results are a consequence of a more general desire to explore the mathematical structure and place of Rényi entropies in the broader theory of quantum information.

Entropy is a concept that was developed in the study of thermodynamics to quantify the disorder of a system. The statistical understanding of entropy was formalised by Boltzmann [1], where it was viewed as a property of thermodynamic systems. The quantities we investigate stem from the information theoretic concept of entropy, characterised by the Shannon entropy $H(X)$ of a random variable [2]. Although having other uses within information theory, entropy is prominently known as a measure of uncertainty (or spread) of a given random variable, quantifying the average information which would be gained (or uncertainty removed) from the observation of that variable.

Associated with the Shannon entropy of a classical random variable X are several entropic quantities which each have their own uses and interpretations. The joint entropy $H(XY)$ describes the entropy of a multivariate random variable and, as for joint probability, is affected by whether or not the variables are dependent. Conditional entropy $H(X|Y)$ quantifies the average uncertainty of X given that Y has been observed. This quantity is well-known and understood, being a fundamental tool in cryptography, statistical analysis and as a general measure of uncertainty. Mutual information $I(X:Y)$ quantifies the level of correlation between X and Y – the difference between the total of the individual entropies of X and Y and the joint entropy of XY. Most prominently, Shannon [2] established that the capacity of any discrete memoryless communication channel is given by the maximal mutual information between the channel's input and output. Beyond its original use in information theory, it has found many other applications in information processing from such a wide range as machine learning (see, e.g., [3, 4, 5]) and computational linguistics (see, e.g., [6]).

The classical Rényi entropies $H_\alpha(X)$ [7] generalise the Shannon entropy. They form a family of entropies parametrised by their order, $\alpha > 0$, that produce measurements of uncertainty that give more or less weight to events with high or low information content. More specifically, they weigh the *surprisal* of outcomes differently depending on the order. We recover the Shannon entropy when $\alpha = 1$ which, naturally, is one of the more useful orders of Rényi entropy, along with *min-entropy* – when $\alpha = \infty$, and the *collision entropy* – when $\alpha = 2$. Formal definitions are given in Section 1.4.

We consider also the quantum generalisation of Rényi entropies. It is prudent therefore to initially cover the quantum analogue of Shannon entropy, the von Neumann entropy $H(\rho)$. This entropy of a probability density matrix ρ is well understood and induces quantum versions of the related entropic quantities that are well behaved and reflect our physical understanding of quantum systems.

The quantum conditional entropy has so far found uses in quantum information such as in the decoupling theorem [8] and more generally as an entanglement witness and as a measure of uncertainty (see [9, 10]). We also have a quantum generalisation of the mutual information which

is compatible with von Neumann entropy and quantum conditional entropy. Quantum mutual information has analogous applications in quantum information, for example characterising the capacity of classical to quantum channels [11, 12, 13] and the quantum channel capacity under entanglement assistance [14, 15, 16, 17]. It has also found applications in other areas of quantum physics, for example as an entanglement and correlation measure (see, e.g., [18]) and to quantify Heisenberg's uncertainty principle (see [19] and [20] for a review on related work).

From this point we will mostly consider quantum entropies and hence omit the 'quantum' and rather specify that an entropy is classical if necessary.

Certain equivalences of entropic quantities are known as chain or decomposition rules. Aside from highlighting the relationships between the quantities, these equivalences can be considered as a method to define them. Indeed, given the relatively simple definitions of the von Neumann entropy and joint entropy we find the conditional entropy *chain rule*, for ρ_{AB} on the system AB with marginal ρ_B on B,

$$H(A|B)_\rho = H(\rho_{AB}) - H(\rho_B). \tag{0.1}$$

This concept can be extended to the tripartite conditional entropy chain rule: for a state ρ_{ABC} on the system ABC,

$$H(AB|C)_\rho = H(A|BC)_\rho + H(A|C)_\rho = H(B|AC)_\rho - H(B|C)_\rho. \tag{0.2}$$

Similarly, mutual information can be described as equivalent to an expression of conceptually simpler entropies, i.e. we define

$$I(A:B)_\rho = H(\rho_A) + H(\rho_B) - H(\rho_{AB}) = H(\rho_A) - H(A|B)_\rho = H(\rho_B) - H(B|A)_\rho. \tag{0.3}$$

These equivalences are interpreted as *decomposition rules*, expressing the mutual information in terms of the von Neumann entropies of the different marginals of the joint state on the systems A and B. By appealing to the intuition that entropy measures uncertainty in a quantum system, they reveal that mutual information measures the uncertainty in A that is due to the lack of knowledge of B, and vice versa.

Another important and equivalent pair of definitions for conditional entropy and mutual information is given in terms of quantum relative entropy [21], namely as optimisations of the relative entropy between the joint state ρ_{AB} and particular product states of the marginals:

$$H(A|B)_\rho = -\min_{\sigma_B} D(\rho_{AB}\| \mathbb{1}_A \otimes \sigma_B) = -D(\rho_{AB}\| \mathbb{1}_A \otimes \rho_B) \quad \text{and} \tag{0.4}$$

$$I(A:B)_\rho = \min_{\sigma_A,\sigma_B} D(\rho_{AB}\|\sigma_A \otimes \sigma_B) = \min_{\sigma_B} D(\rho_{AB}\|\rho_A \otimes \sigma_B) = D(\rho_{AB}\|\rho_A \otimes \rho_B), \tag{0.5}$$

where in the expressions on the right we observed that the minima are achieved for the marginals ρ_A and ρ_B of the joint state ρ_{AB}. These expressions reveal a fundamental property of these quantities that is not evident from the chain and decomposition rules, namely the *data-processing inequality*. Specifically, this property entails that the conditional entropy is monotonically non-decreasing and the mutual information non-increasing under any local processing of information on A and B. Its satisfaction directly follows from the monotonicity under quantum channels of the underlying relative entropy and the above equivalences. This property is crucial for many applications relating to the physical interpretations of these quantities, since it corresponds to our intuition that correlations cannot be created by acting on only one of the constituent parts (or by acting on them independently).

On the other hand, Rényi entropic quantities which maintain useful properties are not as

straight-forward to derive and involve a significant increase in complexity. That said, a natural method of deriving suitable Rényi generalisations of the above quantities is to replace the relative entropy in Eqs. (0.4) and (0.5) with a (quantum) Rényi relative entropy which naturally satisfies the data-processing inequality. It is important to note here that, in general, the equivalences in Eqs. (0.2)-(0.5) no longer hold in the case of Rényi entropic quantities defined via this method.

Various generalisations of the concepts of conditional entropy and mutual information to one-parameter families of operationally significant measures have been proposed. Possible candidates for a Rényi conditional entropy have been put forward, generalising both the quantum conditional entropy and various classical Rényi entropies (see [22, 23, 24] for more detailed treatments). Similarly, for mutual information we have propositions in both the classical (see, e.g., [25, 26], for recent discussions) and the quantum setting (see, e.g., [27]). We call such measures (quantum) Rényi conditional entropy or mutual information respectively, if they satisfy the data-processing inequality.

We are essentially looking for quantities which reproduce the relationships between the classical entropic quantities, have operational significance, reflect the physical situation and are mathematically convenient. Of these propositions for such definitions, quantities derived from the so-called 'sandwiched' divergence [28, 29] have in recent years surfaced as some of the most suitable. This parametrised quantum relative entropy, which we from this point refer to as simply 'Rényi divergence', meets the above criteria and has found important applications within quantum information. This thesis essentially provides new results for Rényi divergence based quantum entropic quantities.

Rényi divergence generalises the quantum relative entropy from which we derive the quantum entropic quantities detailed above. It has many desirable properties as a measure of uncertainty and its structure is amenable to formulating inequalities. As an analogue to the quantum relative entropy, it induces definitions of the quantum Rényi conditional entropy [28] and mutual information [27]. These definitions take the form

$$H_\alpha(A|B)_\rho := -\min_{\sigma_B} D_\alpha(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B) \quad \text{and} \quad I_\alpha(A;B)_\rho := \min_{\sigma_B} D_\alpha(\rho_{AB} \| \rho_A \otimes \sigma_B), \quad (0.6)$$

where $D_\alpha(\cdot \| \cdot)$ denotes the Rényi divergence and $\alpha \in \left[\frac{1}{2}, \infty\right)$. Note that, given this definition, we are lacking an obvious way to decompose the Rényi conditional entropy and mutual information into Rényi entropies of the state's marginals.

However, these quantities generalise the conditional entropy and mutual information respectively, which are recovered by setting $\alpha = 1$. These quantities and similar constructions are widely used in the study of strong converses and the analysis of channel coding problems (see, e.g., [30, 31, 32, 29, 33, 34]) and Rényi mutual information has found direct operational interpretation in classical and quantum hypothesis testing [27, 26].

Uncertainty relations are important in any field that involves signals and waves, but have found particular significance in quantum mechanics. These relations are often put in terms of commutators and standard deviations, for example the Robertson relation (see Section 1.6.1), which is a generalisation of the famous *Heisenberg Uncertainty Principle* [35]. Entropic uncertainty relations, on the other hand, take a more information theoretic approach, expressing uncertainty in terms of its entropy.

There are two canonical entropic uncertainty relations, the more well known being the Maassen-Uffink relation [36] which indicates that the total uncertainty of measuring in two different bases is bounded below by a constant depending on the compatibility of the measurements. Even though it is usually expressed in terms of Shannon entropies, the general form of the theorem encompasses a family of relations in terms of classical Rényi entropy. There are some current well-known

extensions and improvement of this relation which we cover in Section 1.6.3.

By extending the Shannon version of the Maassen-Uffink relation to conditional entropies we can then derive the Hall information exclusion principle [19]. This result is an entropic uncertainty relation in terms of mutual information, describing the level of correlation between parts of a quantum system and some memory containing information about the system. This relation, dual to Heisenberg's uncertainty principle, gives upper bounds on the total amount of correlation between a state measured in either one of two incompatible bases and some classical memory with information about how the initial state was prepared, i.e. if there is a high level of correlation between one measurement of the state and the memory, then there is a proportionally low level of correlation for the other measurement, depending on the compatibility of measurements. Hall's original relation has been improved upon, with bounds stronger than that found in the 'Maassen-Uffink' type relations.

The derivation of these relations and their improvements rely on the chain and decomposition rules of quantum entropies, motivating our investigation of feasible Rényi generalisations. We defer the details and formal statements to Section 1.6.

The applications of entropic uncertainty relations and related concepts include quantum randomness, quantum cryptography and entanglement witnessing, see [20] for a more extensive review.

We now outline the core results of this thesis, i.e. establishing Rényi versions of the chain and decomposition rules in Eqs. (0.2) and (0.3). So far the main example of such an extension is Dupuis' chain rules [37], which generalise the tripartite chain rule to conditional entropies based on Rényi divergence.

In particular, we provide decomposition rules for Rényi mutual information which can then be applied to determine Rényi information exclusion relations. These rules take the form

$$I_\alpha(A\,;B)_\rho \geq H_\beta(\rho_A) - H_\gamma(A|B)_\rho \quad \text{and} \quad I_\alpha(A\,;B)_\rho \leq H_{\bar{\beta}}(\rho_A) - H_{\bar{\gamma}}(A|B)_\rho \qquad (0.7)$$

for suitable choices of Rényi orders $\beta, \bar{\beta}, \gamma$ and $\bar{\gamma}$. The formal result is presented in Theorem 3.3. The two inequalities above reduce to the equality in Eq. (0.3) when we take all the parameters to 1.

We employ a novel interpolation approach which additionally yields a general bipartite divergence inequality and an alternative derivation of the tripartite chain rules. For completeness, we include a previously published proof of the decomposition rules in Appendix A which employs a similar method to that used by Dupuis in his demonstration of the chain rules. Interestingly, these different approaches produce slightly varied valid ranges for the Rényi parameters, the explanation of this discrepancy is left as an open question.

The principle mechanism in demonstrating the desired Rényi divergence inequalities is exploiting their connection to Schatten norms. This is done by performing complex interpolation on specific linear operator valued functions which yield our entropic quantities. The main machinery used to this end is an extension of Beigi's three-line theorem [38] to spaces of general linear operators, which we detail in Chapter 4.

The choices we make in these interpolations are not determined by mere luck or simply arrive out of the void, rather they are inspired and motivated by the norms on general L_p-spaces developed by Pisier [39]. These 'Pisier norms' have already found utility in quantum information (see [40, 41]) due to how they relate to Rényi divergence when constrained to finite spaces. In Chapter 2 we discuss some observations of their properties in the context of Rényi entropies and how they influence the derivation of the divergence inequalities.

As a direct application we are able to establish Rényi generalisations of the Maassen-Uffink and Hall relations, and make further improvements on the bounds. Some uncertainty relations of this genre have already been treated, notably the bipartite uncertainty relation found in [42], which provides a quantum Rényi extension to an already improved version of the Maassen-Uffink relation found in [43].

We are able to improve the bipartite uncertainty relation by deriving state-dependent and state-independent bounds analogous to those determined in [44]. These new bounds are determined by the other Rényi orders present in the relation and we may recover known relations with particular choices of parameters. We are then able to establish a new family of information exclusion relations which provide a direct extension of the Hall relation to Rényi entropies. This family of relations can then be further improved, again with an analogous method to that used in [44]. Interestingly, the resulting tighter upper bounds are not order-dependent, unlike those for the improved Rényi bipartite uncertainty relations.

The structure of the thesis is as follows:

In Chapter 1 we cover the mathematical fundamentals, including a brief treatment of probability theory (mainly for the purposes of notation) and an overview of linear operators and Schatten norms. The remainder of this chapter is dedicated to the formal definitions of Rényi divergence and its related quantities, with a summary of some useful properties.

We explore Pisier norms and interpolation in Chapter 2. This involves introducing a family of super-operators and associated norms, which help inform some observations about certain specifications of Pisier norms. We also establish the three-line theorem for spaces of linear operators, integral to the proofs of the subsequent chapter.

Chapter 3 details the Rényi divergence inequalities and their proofs, including the generalised bipartite divergence inequalities, the decomposition rules, the tripartite chain rules and some other comparisons which directly follow. This chapter also includes some specifications of the new interpolation result which facilitate the derivation of the main inequalities.

The improved Rényi uncertainty relations are then established formally in Chapter 4. We first cover the bipartite uncertainty relations, starting with a generalised version of the result from [42], then move to a version with an improved, state-independent bound. We conclude the main results with a Rényi information exclusion relation and its improved version.

We then provide a brief discussion, detailing some possible future work, such as further generalisations and applications. We also examine how these results may fit into the broader theory, notably how it could aid in establishing generalisations of the chain and decomposition rules for smooth Rényi entropies [45, 46] and how it could inform a clearer definition of the Rényi conditional mutual information (see [47]), both currently open questions in quantum information.

Notation and nomenclature

We use standard notation as summarised in the following table:

Table 0.1 : Overview of notation

Symbol	Meaning
$\mathrm{supp}(X)$ $[\mathrm{supp}(\rho)]$	The support of a random variable [probability density matrix], i.e. the values of x such that $P(X = x) \neq 0$ [the vectors such that $\rho \,|\psi\rangle \neq 0$].
$\|X\|$ $[\|\mathrm{supp}(\rho)\|]$	For a random variable X, the cardinality of the support of $P(X = x)$. [For a probability density matrix ρ, the dimension of $\mathrm{supp}(\rho)$.]
$\langle X \rangle$	The expected value of X, $\sum_x x P(X = x)$.
$\log$	The logarithm to base 2, we also take the convention that $0 \log 0 = 0$.
A, B, AB	Quantum system or subsystems
$\mathcal{H}_A, \mathcal{H}_B$ $[\mathcal{H}_{AB}]$	The Hilbert spaces of states corresponding to the quantum systems A and B. $[\mathcal{H}_A \otimes \mathcal{H}_B]$
$\mathcal{L}(A, B) \; [\mathcal{L}(A)]$	Set of linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$ $[\mathcal{H}_A]$.
$\mathcal{P}(A) \; [\mathcal{P}^*(A)]$	The set of positive semi-definite [strictly positive] operators from $\mathcal{H}_A$ to $\mathcal{H}_A$.
$\mathcal{D}(A) \; [\mathcal{D}^*(A)]$	The subset of $\mathcal{P}(A) \; [\mathcal{P}^*(A)]$ of operators with unit trace.
$\mathcal{U}(A) \; [\mathcal{U}(A, B)]$	The set of unitary operators in $\mathcal{L}(A)$ [isometries in $\mathcal{L}(A, B)$].
$\mathrm{CPTP}(A, B)$	The set of completely-positive trace-preserving operator maps from $\mathcal{L}(A)$ to $\mathcal{L}(B)$.
$\mathrm{tr}_A(\cdot)$	The partial trace over A, $\mathrm{tr}_A(K_A \otimes L_B) = \mathrm{tr}(K_A) L_B$.
$\sigma \gg \rho$	σ 'dominates' ρ, i.e the kernel of σ is contained in the kernel of ρ.
$\sigma \perp \rho$ $[\sigma \not\perp \rho]$	σ and ρ are [not] perpendicular, i.e. the span of σ and ρ have [non-]empty intersection.
$\mathbb{1}_A \in \mathcal{L}(A)$	The identity map on $\mathcal{H}_A$.
$\mathrm{Op}_{A \to B}(\cdot)$	The operator mapping from A to B. For basis vectors $|e_i\rangle \in \mathcal{H}_A, |f_j\rangle \in \mathcal{H}_B$, $\mathrm{Op}_{A \to B} : \mathcal{H}_{AB} \to \mathcal{L}(AB), |e_i\rangle \otimes |f_j\rangle \mapsto |f_j\rangle\langle e_i|$.
$\|\cdot\|_p$ $[\|\cdot\|_p^*]$	The p-Schatten norm, $\|X\|_p = \mathrm{tr}\left[(X^\dagger X)^{\frac{p}{2}}\right]^{\frac{1}{p}}$. This is not a norm for $p < 1$. [The dual p-norm - the norm induced on the space of functionals on the space equipped with $\|\cdot\|_p$.]
$\Re(z)$	The real part of the complex number z.
$\lambda_{\max}[M]$	The maximum eigenvalue of M.
$[X, Y]$	The commutator: for $X, Y \in \mathcal{L}(A)$, $[X, Y] = XY - YX$.

Chapter 1

Background and current literature

1.1 Probability theory

Before we discuss entropy, we must establish some definitions and notations for probability.

1.1.1 Events and outcomes

Here we consider only discrete random variables and, accordingly, a set of outcomes is a discrete set which describes the observable results of an experiment. An event, on the other hand, describes a subset of possible outcomes. For example, when rolling a standard die, the numbers 1 through 6 are possible outcomes whereas the outcome being an odd number is an event which includes the outcomes 1, 3 and 5.

Given a set of outcomes Σ, we consider the probability of an event $E \subseteq \Sigma$ as a subadditive function from the power set of Σ to the real unit interval, $P : 2^\Sigma \to [0, 1]$. $P(E) = 1$ means that E is guaranteed to happen and $P(E) = 0$ means E will not happen.

Mutually exclusive events are events that cannot occur at the same time. Heads or tails being the result of a coin toss are mutually exclusive whereas it raining on the weekend or it reaching $20°$ are not mutually exclusive. We therefore expect that, for mutually exclusive events E and F, if $P(E) = p$ then

$$P(F) \leq 1 - p \implies P(E) + P(F) \leq 1. \tag{1.1}$$

A complete set of mutually exclusive events encompasses all possible outcomes of an experiment. If $\{E_i\}$ is a complete set of mutually exclusive events we have $\bigcap_i E_i = \emptyset$ and $\bigcup_i E_i = \Sigma$. Moreover,

$$P(E_k) = 1 - \sum_{i \neq k} P(E_i) \implies \sum_i P(E_i) = 1. \tag{1.2}$$

Eq. (1.2) is known as the completeness relation.

1.1.2 Random variables

A random variable X associates the outcomes of an experiment with real values, so we may consider the set of outcomes $\Sigma \subseteq \mathbb{R}$. Accordingly, we have the probability of an outcome $x \in \Sigma$, $P(X = x)$ and the probability of an event, $X \in E \subseteq \Sigma$, written $P(X \in E) = \sum_{x \in E} P(X = x)$. Clearly, if $x_1 \neq x_2$, $X = x_1 \implies X \neq x_2$, hence the set $\{x_i\}_i$ is complete and mutually exclusive, i.e. $\sum_i P(X = x_i) = 1$. A discrete random variable X is described by its probability mass function $p(x) = P(X = x)$.

An important property of a random variable is its expected value $\langle X \rangle$. This is a measure of centre which plays a significant role in the analysis of random variables. We define the expected value as the sum of the outcome values, weighted by their probabilities: $\langle X \rangle = \sum_{x \in \Sigma} x p(x)$. Note that in general this does not reproduce the most likely value, but rather characterises the average of the values generated by many independent observations of identical random variables.

1.1.3 Probability and related quantities

The *joint probability* of two events E and F is written $P(E, F)$, read "the probability of E and F". It describes the chance that both E and F will occur simultaneously. In terms of random variables we write $P(X = x, Y = y) = p(xy)$.

The *conditional probability* $P(E|F)$ is similar but instead considers the chance that E occurs given that F occurs. For random variables we write $P(X = x|Y = y) = p(x|y)$.

For independent random variables we have $p(xy) = p(x)p(y)$, otherwise

$$p(xy) = p(x)p(y|x) = p(y)p(x|y) \implies p(x) = \frac{p(xy)}{p(y|x)}, \quad p(x|y) = \frac{p(y|x)p(x)}{p(y)}. \qquad (1.3)$$

The equation on the far right is known as Bayes' Theorem.

This notation will be used for the remainder of this thesis, except for some cases where it may cause ambiguity.

1.2 Banach spaces of operators

The main accepted mathematical model of quantum mechanics and information is based on complex-valued linear algebra and the analysis of vector spaces. In this section we cover some of the theory and tools available to us due to this model. In particular, we detail some results which are useful in the following sections.

1.2.1 Matrices as operators

An operator, in general, is a mapping that takes an element of one space to the element of another space (possibly the same space). In our context of linear operators acting on finite-dimensional Hilbert spaces, we may represent operators with matrices and we use the terms interchangeably when the focus dictates it.

Transformations on operators

For reference we cover some basic manipulations of operators.

- The *inverse* of a square matrix can be considered as the operator which reverses the action of the original. Not all matrices are invertible, these are known as *singular* or *degenerate* matrices. Given invertible $M \in \mathcal{L}(A)$ its inverse, M^{I} is such that if, for $v \in \mathcal{H}_A$,

$$Mv = u \implies v = M^{\mathrm{I}}u = M^{\mathrm{I}}Mv. \qquad (1.4)$$

  As the above indicates, the inverse relationship is symmetric and $MM^{\mathrm{I}} = M^{\mathrm{I}}M = \mathbb{1}_A$.

- The *transpose* is essentially a reflection of the matrix across its diagonal. Given $M = (a_{ij})$, $M^{\top} = (a_{ji})$. Many properties and attributes are preserved under the transpose, the trace for example. If $M \in \mathcal{L}(A, B)$ then $M^{\top} \in \mathcal{L}(B, A)$.

- The *conjugate* $\overline{M}$ of complex-valued matrix M is just the matrix whose entries are the complex conjugates of the entries of M.

- The *adjoint* $M^{\dagger}$ of an operator is the operator such that for the inner product of the Hilbert space it is acting upon $\langle u, Mv \rangle = \langle M^{\dagger}u, v \rangle$. In our context, this simplifies to the conjugate transpose of M, i.e. $M^{\dagger} = \overline{M}^{\top}$.

Types of operators

We have particular categories of operators each with specific properties. We work our way from the most general to the least.

- A *square* matrix is any matrix that has the same number of rows as columns. As an operator it maps between spaces of the same dimension – any operator that maps to the same space must at least be square.

- A *diagonalisable* matrix is a square matrix for which there exists a basis such that it can be written as a diagonal matrix. The entries of this diagonal matrix are the eigenvalues of the original.

- A *normal* matrix is such that $M^\dagger M = MM^\dagger$. As a result all normal matrices map to and from the same space.

- A *self-adjoint* (Hermitian) matrix is such that $M^\dagger = M$. A self-adjoint matrix is thus normal. The eigenvalues of a self-adjoint matrix are real and its eigenvectors form an orthonormal basis.

- A *positive semi-definite* matrix is a self-adjoint matrix whose eigenvalues are greater than or equal to zero. A *strictly positive* or positive definite matrix is similar except that its eigenvalues are all greater than zero. We often shorten positive semi-definite to just *positive*. If $X \in \mathcal{P}(A)$ is positive then there exists an $M \in \mathcal{L}(A, B)$ such that $X = M^\dagger M$ and a self-adjoint $Y \in \mathcal{L}(A)$ such that $X = Y^2$.

- A *density* matrix is a trace-1 positive matrix.

- An *isometry* U is an operator such that $U^\dagger U = \mathbb{1}$. The length of a vector is invariant under isometric operations.

- A *unitary* operator U is an isometry which is also square. Any evolution of a quantum state can be represented as a unitary operator since it takes unit vectors to unit vectors of the same dimension.

Functions on operators

Given a function from $\mathbb{C} \to \mathbb{C}$, we can define an extension of this function on any normal matrix, depending on its domain.

For a subset $\Omega \subseteq \mathbb{C}$ and $f : \Omega \to \mathbb{C}$ we define $f : \Omega^{m \times m} \to \mathbb{C}^{m \times m}$ where, if the eigenvalues of M are $\{\lambda_i\}_i \subseteq \Omega$, the eigenvalues of $f(M)$ are $\{f(\lambda_i)\}_i$.

Most functions are well defined for normal matrices since we are guaranteed a unitarily similar diagonalisation (see Eq. (1.6)). However, sometimes we must restrict our focus for functions with limited natural domains. For example, the logarithm can only be applied to positive matrices since their eigenvalues are all real and non-negative.

We often use the function $f(M) = M^{-1}$ to denote the pseudoinverse [48, Section 5.5.4]. Usually this notation refers to the usual matrix inverse but we require a more general definition. In our context we take advantage of two main properties of the pseudoinverse. Firstly, for matrices with full support, the pseudoinverse coincides with the usual matrix inverse. However, if $M \in \mathcal{L}(A)$ does not have full support (and is therefore singular), $M^{-1}M$ is instead a projector onto $\mathrm{supp}(M) \subset \mathcal{H}_A$. This is problematic as we cannot guarantee a given density operator has full support. This brings us to the second useful property: there are situations where we maintain the usual behaviour of the inverse in the trace, i.e. if $\mathrm{supp}(M) \subset \mathcal{H}_A$ then for $X \in \mathcal{L}(A)$ with $\mathrm{supp}(X) \subseteq \mathrm{supp}(M)$ we have $\mathrm{tr}\, M^{-1}MX = \mathrm{tr}\, X$. In other words, equivalence holds if no information is lost by the action of the projector $M^{-1}M$.

1.2.2 Operator decompositions

As an aid to computation we often make use of operator decompositions. These observations essentially allow us to decompose certain classes of operators into combinations of simpler operators which are more mathematically convenient.

The spectral decomposition

Also known as the 'eigendecomposition', the spectral decomposition allows us to reformulate a given matrix in terms of its eigenvalues and eigenvectors:

Proposition 1.1. *Any diagonalisable matrix A can be factorised as*

$$A = P\Lambda P^{-1}, \tag{1.5}$$

where Λ is a diagonal matrix whose ith entry is the ith eigenvalue, λ_i, of A and the ith column of P is the eigenvector corresponding to λ_i. For normal matrices this reduces to

$$X = U\Lambda U^{\dagger}, \tag{1.6}$$

where U is unitary.

The polar decomposition

The polar decomposition allows us to express any square matrix in a much more convenient form, especially in the context of Schatten norms (see the next section).

Proposition 1.2. *Any operator $M \in \mathcal{L}(A)$ can be decomposed into the product of a positive operator and a unitary operator i.e. for $P, Q \in \mathcal{P}(A)$ and $U \in \mathcal{U}(A)$*

$$M = UP = QU. \tag{1.7}$$

The above are known respectively as the left and right polar decompositions.

The singular value decomposition

We may also decompose non-square matrices in a similar way to the spectral decomposition.

Proposition 1.3. *For any $M \in \mathcal{L}(A, B)$ we may write*

$$M = UDV \tag{1.8}$$

where $U \in \mathcal{U}(C, B), V \in \mathcal{U}(A, C)$ are isometries and D is a positive semi-definite diagonal matrix in $\mathcal{L}(C)$ with dimension equal to the number of non-zero eigenvalues of $M^{\dagger}M$.

The diagonal entries of D are known as the *singular values* of M which we denote $\{s_i(M)\}_i$.

Most expositions of this proposition instead present the decomposition in terms of unitary U and V in $\mathcal{U}(B)$ and $\mathcal{U}(A)$, respectively, and D as a rectangular diagonal matrix in $\mathcal{L}(A, B)$. However, for our purposes, the above form is more convenient (See, for example, Theorem 2.12). Hence we include a brief proof for clarity.

Proof of Proposition 1.3. Let $M \in \mathcal{L}(A, B)$ be an $n \times m$ matrix. We have that $M^{\dagger}M$ is positive semi-definite so by Eq. (1.6) there exists a $\tilde{V}^{\dagger} \in \mathcal{U}(A)$ such that

$$\tilde{V}^{\dagger}M^{\dagger}M\tilde{V} = \tilde{D} = \begin{bmatrix} D^2 & 0 \\ 0 & 0 \end{bmatrix} \tag{1.9}$$

where $D^2 \in \mathcal{L}(C)$ is diagonal and positive semi-definite with $\dim(\mathcal{H}_C)$ equal to the number of non-zero eigenvalues of $M^{\dagger}M$. Again by Eq. (1.6), we may write

$$\tilde{V}^{\dagger} = \begin{bmatrix} V^{\dagger} & V_0^{\dagger} \end{bmatrix} \tag{1.10}$$

where $V^\dagger$ and $V_0^\dagger$ are matrices whose columns are the eigenvectors of the, respectively, non-zero and zero eigenvalues. In this sense we can write

$$
\begin{bmatrix} D^2 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} VM^\dagger M V^\dagger & VM^\dagger M V_0^\dagger \\ V_0 M^\dagger M V^\dagger & V_0 M^\dagger M V_0^\dagger \end{bmatrix},
\tag{1.11}
$$

hence $D^2 = VM^\dagger M V^\dagger$ with $V^\dagger \in \mathcal{U}(C, A)$.

Define $U = MV^\dagger D^{-1} \in \mathcal{L}(C, B)$, thus

$$
UDV = MV^\dagger D^{-1} DV = MV^\dagger V = M.
\tag{1.12}
$$

Moreover,

$$
U^\dagger U = (MV^\dagger D^{-1})^\dagger MV^\dagger D^{-1} = D^{-1} V M^\dagger M V^\dagger D^{-1} = D^{-1} D^2 D^{-1} = \mathbb{1}_B,
\tag{1.13}
$$

Hence $U \in \mathcal{U}(C, B)$. $\qquad \square$

The Schmidt decomposition

Finally, we have the Schmidt decomposition, considered an extension of the singular value decomposition to vectors, which allows us to represent any pure state as a set of orthonormal vectors localised on distinct subsystems.

Theorem 1.4. *We may write any vector $v \in \mathcal{H}_{AB}$ in the form*

$$
v = \sum_i r_i \left(e_A^i \otimes f_B^i \right)
\tag{1.14}
$$

where $\{e_A^i\}_i$ and $\{f_B^i\}_i$ form orthonormal bases on their respective subsystems and the Schmidt coefficients $\{r_i\}_i$ are real and non-negative.

For proofs and related material see [49, Section 2.5].

1.2.3 The Schatten operator norm

If we introduce a norm on a set of a finite complex-valued linear operators we form a Banach space. The Schatten norms form a family which generalise some of the more common norms associated with spaces of operators. Essentially, they can be considered as a matrix generalisation of the p-norms on Lebesgue spaces.

For $M \in \mathcal{L}(A, B)$, we define

$$
\|M\|_p = \left[\mathrm{tr} \left(M^\dagger M \right)^{\frac{p}{2}} \right]^{\frac{1}{p}}, \quad p \geq 1.
\tag{1.15}
$$

Note these norms are well defined for rectangular matrices since $M^\dagger M$ is by definition self-adjoint. In fact, the eigenvalues of $|M| = \left(M^\dagger M \right)^{\frac{1}{2}}$ coincide with the singular values of M (and $M^\dagger$) (see [9, Section 1.1.3]). As such, we may write the above in the form

$$
\|M\|_p = \left[\sum_{i=1}^k s_i \left(M \right)^p \right]^{\frac{1}{p}},
\tag{1.16}
$$

where k is the rank of M.

Properties

The above definition induces some useful properties:

Proposition 1.5. *The Schatten norm is unitarily invariant, i.e. for $M \in \mathcal{L}(A, B)$, $U \in \mathcal{U}(B)$ and $V \in \mathcal{U}(A)$,*

$$\|UMV\|_p = \|UM\|_p = \|MV\|_p = \|M\|_p. \tag{1.17}$$

Proof. By the singular value decomposition we may write $UMV = UU'DV'V$, where $D \in \mathcal{L}(C)$ is a diagonal matrix, $UU' \in \mathcal{U}(C, B)$ and $V'V \in \mathcal{U}(A, C)$. This demonstrates that the singular values of M and UVM coincide, hence by Eq. (1.16) we have the equivalences in Eq. (1.17). $\square$

Proposition 1.6. *For $M \in \mathcal{L}(A, B)$,*

$$\|M\|_p = \|M^\dagger\|_p = \|\overline{M}\|_p = \|M^\top\|_p = \||M|\|_p. \tag{1.18}$$

Proof. We simply observe that the singular values of M are invariant under these transformations. $\square$

Proposition 1.7. *Given $M \in \mathcal{L}(A, B)$ we may write $\|M^\dagger M\|_p = \|M\|_{2p}^2$.*

Proof.

$$\|M^\dagger M\|_p = \left[\operatorname{tr} \left(M^\dagger M M^\dagger M \right)^{\frac{p}{2}} \right]^{\frac{1}{p}} \tag{1.19}$$

$$= \left[\operatorname{tr} \left(M^\dagger M \right)^{\frac{2p}{2}} \right]^{\frac{2}{2p}} \tag{1.20}$$

$$= \|M\|_{2p}^2. \tag{1.21}$$

$\square$

We conclude with an observation about the form of positive operators in the unit ball of a given norm:

Proposition 1.8. *For all $\sigma \in \mathcal{D}(A)$ $\left\| \sigma^{\frac{1}{p}} \right\|_p = 1$.*

Proof. This is evident by considering the definition of the Schatten norm and the fact that $\operatorname{tr}(\sigma) = 1$. $\square$

Examples

The Schatten norms generalise some well known norms on spaces of operators:

- The *trace* norm is equivalent to $\| \cdot \|_1$ and recovers the trace for self-adjoint operators.

- The *Hilbert-Schmidt (or Frobenius)* norm is equivalent to $\| \cdot \|_2$. This norm is induced by the inner product $\langle M, N \rangle = \operatorname{tr}(M^\dagger N)$, hence $\mathcal{L}(A, B)$ equipped with this norm is a Hilbert space.

- The *operator* norm is equivalent to $\| \cdot \|_\infty$, defined as

$$\|M\|_\infty = \sup_{v \in \mathcal{H}_A} \frac{\|Mv\|_2}{\|v\|_2} \qquad \left[\|v\|_2 = \sqrt{\langle v, v \rangle} \right]. \tag{1.22}$$

1.3 Quantum mechanics and information

The following follows the material presented in [49].

1.3.1 The state of a quantum system

A quantum system can be any physical system, but generally we only consider systems where quantum effects are significant, such as an electron, a molecule or the crystalline structure of a material like doped silicon. A quantum system A is associated with a separable* Hilbert space $\mathcal{H}_A$ over $\mathbb{C}$. Here, we consider only finite Hilbert spaces equipped with an inner product

$$\langle \cdot, \cdot \rangle : \mathcal{H}_A \times \mathcal{H}_A \longrightarrow \mathbb{C} \tag{1.23}$$

$$: \quad \langle u, v \rangle \longmapsto \sum_i \overline{u}_i v_i, \tag{1.24}$$

where u_i and v_i are respectively the ith elements of u and v. The possible states of the system A are described by the elements of $\mathcal{H}_A$.

We use bra-ket notation to represent quantum states: for $\psi \in \mathcal{H}_A$, its 'ket' is written $|\psi\rangle$ and it can be considered as a column vector such that for an operator $M \in \mathcal{L}(A, B)$,

$$M |\psi\rangle = M\psi. \tag{1.25}$$

From this point we use the shorthand $|\psi\rangle = \psi \in \mathcal{H}_A$. Considering $|\psi\rangle$ as a matrix, we define the 'bra' of ψ as $\langle\psi| = |\psi\rangle^{\dagger}$.

We find this definition notationally convenient as, considering Eq. (1.24), we may write $\langle\varphi|\psi\rangle = \langle\varphi, \psi\rangle$ and

$$\langle\varphi, A\psi\rangle = \langle\varphi|A\psi\rangle = \left\langle A^{\dagger}\varphi|\psi\right\rangle = \langle\varphi| \, A \, |\psi\rangle. \tag{1.26}$$

1.3.2 State vectors and probability density matrices

A finite dimensional state vector can be represented by a complex-valued ket $|\psi\rangle$, for which $\langle\psi|\psi\rangle = 1$. This is to say that in an orthonormal basis $\{|e_i\rangle\}_i$ where $|\psi\rangle = \sum_i a_i |e_i\rangle$, the square sum of the moduli of the complex coefficients, $\sum_i |a_i|^2 = 1$.

The probability density matrix (or probability operator) ρ of a quantum system is determined by an ensemble of pure states, $\{p_i, |\psi_i\rangle\}_i$. We define

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|, \tag{1.27}$$

where ρ is a density operator and as such we may express its spectral decomposition (see section 1.2.2) in terms of bra-ket notation:

$$\rho = \sum_j \lambda_j |j\rangle \langle j|, \tag{1.28}$$

where the eigenvectors form an orthonormal eigenbasis and $\sum_j \lambda_j = 1$.

A system is considered to be in a pure state when its probability density matrix can be represented by a single state vector, i.e. $\rho = |\psi\rangle \langle\psi|$. This reflects that we have complete knowledge of the state of the system. Note we must still consider superposition and the underlying uncertainty when measuring a pure state in a particular basis. A system that is not in a pure state is said to be in a mixed state.

---

*Having a countable orthonormal basis.

If ρ has the spectral decomposition as in Eq. (1.28) then

$$\rho^2 = \sum_{i,j} \lambda_i \lambda_j \ket{i}\bra{i}j\rangle\bra{j} \tag{1.29}$$

$$= \sum_j \lambda_j^2 \ket{j}\bra{j} \tag{1.30}$$

$$\implies \mathrm{tr}(\rho^2) = \sum_j \lambda_j^2. \tag{1.31}$$

For all j, $0 \leq \lambda_j \leq 1$, so $\lambda_j^2 \leq \lambda_j$, with equality if and only if λ_j is the only eigenvalue, i.e. equality if and only if ρ is a pure state.

This gives us the criterion:

$$\mathrm{tr}(\rho^2) \leq 1, \tag{1.32}$$

with equality if and only if ρ is a pure state.

1.3.3   Composite systems

A composite quantum system is one that is made up of distinct subsystems. We write that $\rho_{AB} \in \mathcal{D}(AB)$ is a probability density matrix of the composite system of A and B. To recover the state on one of the subsystems we 'trace out' the other subsystem with the *partial trace*: given a basis $\{\ket{e_i}\}_i \subset \mathcal{H}_B$ and $M \in \mathcal{L}(AB)$, we define the marginal of M in A:

$$\mathrm{tr}_B(M) = \sum_i \bra{e_i} M \ket{e_i} = M_A \in \mathcal{L}(A). \tag{1.33}$$

Given $\rho_{AB} \in \mathcal{D}(AB)$, the *reduced state* $\rho_A \in \mathcal{D}(A)$ is the marginal of ρ_{AB} in A, i.e. $\mathrm{tr}_B(\rho_{AB}) = \rho_A$. We can consider the reduced state as a description of the state on a subsystem, at least when it comes to making measurements only on this subsystem [49]. The state of a composite system may or may not be entangled. A pure state on a composite system can be decomposed as $\rho_{AB} = \rho_A \otimes \rho_B$ if and only if no entanglement exists between the subsystems.

Classical-quantum states

It is possible to define joint classical-quantum systems, required in the case of measurement on a single subsystem (see next section). In order to consider classical subsystems in the context of density operators we introduce an auxiliary Hilbert space $\mathcal{H}_X$ for the random variable X, with orthonormal basis $\{\ket{x}\}_x$ that acts as a 'classical register' [20]. Accordingly, we have

$$\rho_{XB} = \sum_x p(x) \ket{x}\bra{x} \otimes \rho_B^x, \tag{1.34}$$

where ρ_B^x is the quantum state of the system B conditioned on $X = x$, i.e. we have

$$\mathrm{tr}_X \rho_{XB} = \sum_x p(x) \rho_B^x = \rho_B. \tag{1.35}$$

Observations about pure states

Given a pure state is represented by a vector in its respective Hilbert space, we can use the Schmidt decomposition to derive some useful equivalences:

Proposition 1.9. *Given the pure state* $\rho_{AB} = \ket{\varphi}\bra{\varphi} \in \mathcal{D}(AB)$ *and the Schmidt decomposition* $\ket{\varphi} = \sum_i r_i \ket{i}_A \otimes \ket{i}_B$, *if we let* $X \in \mathcal{L}(A, B)$ *such that* $X = \sum_i r_i \ket{i}_B \bra{i}_A$ *then*

$$X^\dagger X = \rho_A \quad and \quad XX^\dagger = \rho_B. \tag{1.36}$$

Proof. We prove the first statement explicitly, then the second follows from a symmetric argument.

First note that since $\{|i\rangle_A\}_i$ is an orthonormal basis for $\mathcal{H}_A$ we may write the partial trace over A as

$$\rho_B = \mathrm{tr}_A(\rho_{AB}) \tag{1.37}$$

$$= \sum_k \langle k|_A \sum_{i,j} r_i r_j \left(|i\rangle_A \otimes |i\rangle_B\right) \left(\langle j|_A \otimes \langle j|_B\right) |k\rangle_A \tag{1.38}$$

$$= \sum_{i,j,k} r_i r_j \left(\langle k|i\rangle_A \otimes |i\rangle_B\right) \left(\langle j|k\rangle_A \otimes \langle j|_B\right) \tag{1.39}$$

$$= \sum_i r_i^2 |i\rangle_B \langle i|_B . \tag{1.40}$$

Moreover,

$$XX^\dagger = \sum_{i,j} r_i r_j |i\rangle_B \langle i|j\rangle_A \langle j|_B \tag{1.41}$$

$$= \sum_i r_i^2 |i\rangle_B \langle i|_B . \tag{1.42}$$

$\square$

Proposition 1.10. *Given a pure state $|\varphi\rangle \in \mathcal{H}_{ABC}$ with Schmidt decompositions*

$$|\varphi\rangle = \sum_i r_i |i\rangle_A \otimes |i\rangle_{BC} = \sum_i s_i |i\rangle_{AB} \otimes |i\rangle_C , \tag{1.43}$$

we have, for $K_A \in \mathcal{L}(A)$ and $L_C \in \mathcal{L}(C)$,

$$\left\| L_C \sum_i r_i |i\rangle_{BC} \langle i|_A K_A \right\|_2 = \left\| L_C \sum_i s_i |i\rangle_C \langle i|_{AB} K_A \right\|_2 . \tag{1.44}$$

This implies that we are free to choose any valid Schmidt decomposition in the product of an argument of the Schatten 2-norm – effectively 'shifting' a subsystem to the other side – given that subsystem is not otherwise represented in the product.

Proof. We first rewrite the left-hand side as a trace:

$$\left\| L_C \sum_i r_i |i\rangle_{BC} \langle i|_A K_A \right\|_2^2 = \mathrm{tr}\left(K_A^\dagger \sum_i r_i |i\rangle_A \langle i|_{BC} L_C^\dagger L_C \sum_j r_j |j\rangle_{BC} \langle j|_A K_A \right) \tag{1.45}$$

$$= \mathrm{tr}\left(\sum_{i,j} r_i r_j K_A^\dagger |i\rangle_A \langle i|_{BC} L_C^\dagger L_C |j\rangle_{BC} \langle j|_A K_A \right) \tag{1.46}$$

$$= \sum_{i,j,k} \langle k|_A r_i r_j K_A^\dagger |i\rangle_A \langle i|_{BC} L_C^\dagger L_C |j\rangle_{BC} \langle j|_A K_A |k\rangle_A \tag{1.47}$$

$$= \sum_{i,j,k} r_i r_j \langle i|_A K_A |k\rangle_A \langle k|_A K_A^\dagger |j\rangle_A \langle i|_{BC} L_C^\dagger L_C |j\rangle_{BC} \tag{1.48}$$

$$= \sum_{i,j} r_i r_j \langle i|_A \otimes \langle i|_{BC} \left(K_A K_A^\dagger \otimes L_C^\dagger L_C \right) |j\rangle_A \otimes |j\rangle_{BC} \tag{1.49}$$

$$= \langle \varphi| \left(K_A K_A^\dagger \otimes L_C^\dagger L_C \right) |\varphi\rangle \tag{1.50}$$

$$= \sum_{i,j} s_i s_j \langle i|_{AB} \otimes \langle i|_C \left(K_A K_A^\dagger \otimes L_C^\dagger L_C \right) |j\rangle_{AB} \otimes |j\rangle_C . \tag{1.51}$$

From this point, reversing the process obtains the right-hand side. $\qquad\qquad$ $\square$

1.3.4 Measurement

Uncertainty relations are a characterisation of how much information we can extract from a system by measurement, hence we require a formal understanding of what measurement means in the context of quantum information theory.

Although an interesting avenue for further research, this thesis does not treat general projector operator-valued measurements – or POVMs, instead we cover the simpler situation of measurements in orthonormal bases, or ONBs for short.

An ONB $\mathbb{X}$ can be represented by a set of rank 1 projectors: $\mathbb{X} = \{|x\rangle \langle x|\}$, whose spans are mutually orthogonal.

In a real physical situation, if a state is measured in this basis the system will collapse to one of the pure states described by the basis vectors. The probability of observing the system in that basis state can be determined by acting on the state with its associated measurement map:

$$\mathcal{M}_X(\cdot) : \mathcal{L}(A) \longrightarrow \mathcal{L}(X) \tag{1.52}$$

$$: \quad \rho_A \longmapsto \sum_x |x\rangle\langle x| \, \rho_A \, |x\rangle\langle x| = \sum_x \langle x| \, \rho_A \, |x\rangle \, |x\rangle\langle x| \,, \tag{1.53}$$

$\mathcal{M}_X \in \text{CPTP}(A, X)$ and we often denote $\mathcal{M}_X(\rho_A) = \rho_X$.

By the Born Rule, when a pure state $|\psi\rangle$ is measured in an ONB with basis vectors $\{|x_i\rangle\}_i$, the square modulus of the coefficients λ_x of each eigenvector can be interpreted as the probability $p(k)$ that the system will be observed in the state $|x_k\rangle$, i.e.

$$\langle\psi|x_k\rangle \, \langle x_k|\psi\rangle = \sum_i |\lambda_{x_i}|^2 \, \langle x_i|x_k\rangle \, \langle x_k|x_i\rangle \tag{1.54}$$

$$= \sum_i |\lambda_{x_i}|^2 \delta_{ik} \tag{1.55}$$

$$= |\lambda_k|^2 = p(k). \tag{1.56}$$

For a mixed state the probability is instead the weighted sum of the coefficients for each pure state in the ensemble.

We may also formalise measurement maps on only part of a multipartite system. In this case we consider $\mathcal{M}_X^* \in \text{CPTP}(AB, XB)$ noting that even though it is defined on the larger space it only acts on the relevant subsystem. With this in mind, the expansion to any additional subsystem is assumed in a given context and we denote $\mathcal{M}_X(\rho_{AB}) = \rho_{XB}$, where it does not cause ambiguity.

The probability that measuring ρ in $\mathbb{X}$ will have the outcome x, i.e. the sum of the joint probabilities that ρ is in the state $|\psi_i\rangle$ and that it collapses to $|x\rangle$, induces the random variable X with probability mass function: $\langle x| \, \rho \, |x\rangle = p(x)$, given by the eigenvalues of $\mathcal{M}_X(\rho)$.

1.4 Entropy

Entropy itself is, in broad terms, a measurement of chaos. Systems that are well structured and predictable have low entropy – consider a closed container of liquid water and air – whereas high entropy systems have little structure and are hard to predict, such as the same container holding a homogenous mixture of water vapour and air. Basically, a system that is more "mixed up" has more entropy.

From an information theoretic perspective, entropy is instead viewed in terms of random variables. A random variable with a uniform distribution – where each outcome is equally likely – has

high entropy and one with a degenerate distribution – where there is only one possible outcome – has low entropy. In terms of the above example, we can consider the distribution of whether or not a water molecule will be observed at a particular position. In the low entropy, liquid state we would be guaranteed one way or the other depending on which side of the surface we measure, but in the high entropy, gaseous state we would be much less sure of the outcome. This statistical interpretation is well formalised by the Shannon entropy.

1.4.1 Classical Entropy

The Shannon entropy $H(X)$ of a random variable X is the most common, and arguably the most useful, definition of entropy. It satisfies the following postulates, desirable for a measurement of uncertainty [50]: If $n = |X|$, then

(a) $H(X) = H(p(x_1), p(x_2), ..., p(x_n))$ *is symmetric,*

(b) $H(p(x), 1 - p(x))$ *is continuous for* $0 < p(x) \leq 1$,

(c) $H(\underbrace{1/n, 1/n, ..., 1/n}_{n \text{ times}}) = \log_2 n$,

(d) $H(tp(x_1), (1 - t)p(x_1), p(x_2), ..., p(x_n)) = H(p(x_1), p(x_2), ..., p(x_n)) + p(x_1)H(t, 1 - t)$, *for* $0 \leq t \leq 1$.

Shannon entropy and surprisal

To determine a rigorous definition of entropy we first consider surprisal [20, Section III.A.1].

Let's restrict our focus to a single outcome. The amount of information gained from observing this outcome is related to its probability – we gain more information when less likely outcomes occur, i.e. we can consider surprisal to quantify how surprised one would be if the outcome occurred. Additionally, we want the information gained from two independent outcomes to be the sum of the information gained from each outcome, i.e. we want a function $f(P(X = x)) = f(x)$ such that

$$p(x_1) < p(x_2) \implies f(x_1) > f(x_2) \text{ and } f(x, y) = f(p(x)p(y)) = f(x) + f(y). \tag{1.57}$$

Any function of the form $-\lambda \log_2 p(x)$, $\lambda \in \mathbb{R}^+$ satisfies these conditions, and we take $\lambda = 1$ for simplicity [2], giving us the definition $f(x) := -\log_2 p(x)$. [cite]

Since we want to know the expected amount of information gained from observing the random variable we take the average of the surprisal to obtain the Shannon entropy:

$$H(X) = -\sum_x p(x) \log p(x). \tag{1.58}$$

This satisfies the postulates (a)-(d) and is an adequate if not preferable measurement of uncertainty in many situations.

For any random variable X its Shannon entropy is bounded: $0 \leq H(X) \leq \log |X|$, with equality on the left if and only if the distribution of X is degenerate and equality on the right if and only if the distribution of X is uniform on its support.

Joint entropy

Joint entropy is the entropy of the random variable described by the joint probability of two or more random variables. Given two random variables X and Y, their joint entropy is defined

$$H(XY) = -\sum_{x,y} p(xy) \log p(xy). \tag{1.59}$$

By Bayes' Theorem (see Eq. (1.3)) and Jensen's inequality, we have

$$H(XY) \leq H(X) + H(Y), \tag{1.60}$$

with equality if and only if X and Y are independent.

Conditional entropy

The conditional entropy of X given Y is average the amount of information left to gain about X once Y has been observed. This is the remaining uncertainty of XY when given access to the side information provided by Y about X. As such

$$H(X|Y) = H(XY) - H(Y). \tag{1.61}$$

From Eq. (1.60) we can see that

$$H(X|Y) \leq H(X). \tag{1.62}$$

The conditional entropy is equivalently defined [51]

$$H(X|Y) = \sum_y p(y) H(X|Y = y) \tag{1.63}$$

$$= -\sum_{x,y} p(y) p(x|y) \log p(x|y) \tag{1.64}$$

$$= -\sum_{x,y} p(xy) \log \frac{p(xy)}{p(y)}. \tag{1.65}$$

Note that Eq. (1.61) is the bivariate form of the more general multivariate chain rule:

$$H(X|YZ) = H(XY|Z) + H(Y|Z). \tag{1.66}$$

Mutual information

Mutual information, on the other hand, is the amount of information shared between random variables, quantifying how well they are correlated. This notion is conveyed through its definition in terms of the decomposition rules:

$$I(X:Y) = H(X) + H(Y) - H(XY) \tag{1.67}$$

$$= H(XY) - H(X|Y) - H(Y|X) \tag{1.68}$$

$$= H(X) - H(X|Y) = H(Y) - H(Y|X). \tag{1.69}$$

From Eqs. (1.62) and (1.69) we can see $I(X:Y) \geq 0$.

Relative entropy

Relative entropy, also know as Kullback-Leibler divergence, is a quantity that can be used to measure the closeness of two distributions over the same index set in terms of their entropies.

For X_1 and X_2, $P(X_1 = x) = p_1(x)$, $P(X_2 = x) = p_2(x)$ we define

$$D(X_1 \| X_2) = \sum_x p_1(x) \log \frac{p_1(x)}{p_2(x)}. \tag{1.70}$$

In the case that $p_2(x)$ does not dominate $p_1(x)$, i.e. there exists a x in the index set such that $p_1(x) > 0$ and $p_2(x) = 0$, then $D(X_1 \| X_2) = \infty$.

We can express classical entropies in terms of this quantity:

$$H(X) = \log|X| - D\left(p(x) \middle\| \frac{\{1\}_x}{|X|}\right) \tag{1.71}$$

$$= -\sum_x p(x) \log p(x), \tag{1.72}$$

$$H(X|Y) = -D(p(xy)\|p(y)) \tag{1.73}$$

$$= -\sum_{x,y} p(xy) \log \frac{p(xy)}{p(y)} \tag{1.74}$$

$$= \sum_y p(y) \log p(y) - \sum_{x,y} p(xy) \log p(xy) \tag{1.75}$$

$$= H(XY) - H(Y), \tag{1.76}$$

$$I(X:Y) = D(p(xy)\|p(x)p(y)) \tag{1.77}$$

$$= \sum_{x,y} p(xy) \log \frac{p(xy)}{p(x)p(y)} \tag{1.78}$$

$$= \sum_{x,y} p(xy) \log p(xy) - \sum_x p(x) \log p(x) - \sum_y p(y) \log p(y) \tag{1.79}$$

$$= H(X) + H(Y) - H(XY) \tag{1.80}$$

$$= H(X) - H(X|Y). \tag{1.81}$$

1.4.2 Quantum entropies

We want to find applications of entropy in quantum information, so here we discuss the quantum analogues of the classical entropies covered in the previous sections. But first, a run-down of the concepts and notation.

Von Neumann entropy

Related to the Shannon entropy is its quantum analogue, von Neumann entropy. Instead of taking a random variable as input, the von Neumann entropy takes the probability density matrix ρ describing the state of a quantum system. It is defined similarly to Shannon entropy:

$$H(\rho) = -\operatorname{tr}(\rho \log \rho). \tag{1.82}$$

Fortunately, when ρ is expressed in terms of its spectral decomposition (see Eq. (1.28)), the above simplifies to the Shannon entropy of the random variable with probability mass function given by the eigenvalues of ρ. i.e. $H(\rho) = -\sum_j \lambda_j \log \lambda_j$. We can use the von Neumann entropy as a stepping off point to derive the quantum generalisations of the classical entropic quantities via the intuitive chain and decomposition rules. Accordingly, we have the quantum joint entropy:

$$H(\rho_{AB}) = -\operatorname{tr}(\rho_{AB} \log \rho_{AB}), \tag{1.83}$$

the quantum conditional entropy:

$$H(A|B) = H(\rho_{AB}) - H(\rho_B) \tag{1.84}$$

and the quantum mutual information:

$$I(A : B) = H(\rho_A) + H(\rho_B) - H(\rho_{AB}). \tag{1.85}$$

Similar to the classical entropies, these relationships coincide with the definitions established via the quantum relative entropy [21]

$$D(\rho\|\sigma) = \mathrm{tr}\left[\rho(\log\rho - \log\sigma)\right]. \tag{1.86}$$

Analogous to the classical case, if $\sigma \not\gg \rho$ then $D(\rho\|\sigma) = \infty$.

We have for $\rho_{AB} \in \mathcal{D}(AB)$:

$$H(\rho_{AB}) = \log|\mathrm{supp}(\rho_{AB})| - D\left(\rho_{AB}\left\|\frac{\mathbb{1}_{AB}}{|\mathrm{supp}(\rho_{AB})|}\right.\right), \tag{1.87}$$

$$H(A|B)_\rho = -D(\rho_{AB}\|\mathbb{1}_A \otimes \rho_B) \tag{1.88}$$

$$= -\mathrm{tr}(\rho_{AB}\log\rho_{AB}) + \mathrm{tr}(\rho_B\log\rho_B) \tag{1.89}$$

$$= H(\rho_{AB}) - H(\rho_B), \tag{1.90}$$

$$I(A : B)_\rho = D(\rho_{AB}\|\rho_A \otimes \rho_B) \tag{1.91}$$

$$= \mathrm{tr}(\rho_{AB}\log\rho_{AB}) - \mathrm{tr}(\rho_A\log\rho_A) - \mathrm{tr}(\rho_B\log\rho_B) \tag{1.92}$$

$$= H(\rho_A) + H(\rho_B) - H(\rho_{AB}) \tag{1.93}$$

$$= H(\rho_A) - H(A|B)_\rho. \tag{1.94}$$

In fact we obtain equivalent definitions for the conditional entropy and mutual information by minimising over one or both subsystems. As this is shown for conditional entropy in [28], we provide a similar demonstration for mutual information: if we consider the positive-definiteness of the relative entropy due to Klein's inequality [52], and observe that when $\sigma_B = \rho_B$ then $D(\rho_B\|\sigma_B) = 0$, we can write

$$I(A : B)_\rho = D(\rho_{AB}\|\rho_A \otimes \rho_B) + \inf_{\sigma_B \in \mathcal{D}(B)} D(\rho_B\|\sigma_B) \tag{1.95}$$

$$= \mathrm{tr}(\rho_{AB}\log\rho_{AB} - \rho_{AB}\log(\rho_A \otimes \rho_B)) + \mathrm{tr}(\rho_B\log\rho_B) - \inf_{\sigma_B \in \mathcal{D}(B)}\mathrm{tr}(\rho_B\log\sigma_B) \tag{1.96}$$

$$= \inf_{\sigma_B \in \mathcal{D}(B)}\mathrm{tr}(\rho_{AB}\log\rho_{AB} - \rho_{AB}\log(\rho_A \otimes \sigma_B)) \tag{1.97}$$

$$= \inf_{\sigma_B \in \mathcal{D}(B)} D(\rho_{AB}\|\rho_A \otimes \sigma_B). \tag{1.98}$$

A similar calculation can be used to show that the equivalence also holds when minimised over both subsystems, i.e.

$$I(A : B)_\rho = \inf_{\substack{\sigma_A \in \mathcal{D}(A) \\ \sigma_B \in \mathcal{D}(B)}} D(\rho_{AB}\|\sigma_A \otimes \sigma_B). \tag{1.99}$$

So far we have been discussing these relative entropy or divergence quantities only in terms of Shannon or von Neumann entropies. It is important to point out that this framework allows for a relatively clear and simple definition of all these quantities but in the case of Rényi entropies things are not so straight-forward. Nonetheless, armed with these definitions it is much easier to discuss how to extend this structure to Rényi entropies.

1.4.3 Rényi entropy

Rényi entropies were first proposed in [50] as an alternative definition of entropy, having most of the desired properties as well as allowing the consideration of information of different order.

The Rényi entropy of order $\alpha \in (0,1) \cup (1,\infty)$ of a classical random variable X, with probability mass function $p(x)$, is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_x p(x)^\alpha \right). \tag{1.100}$$

It generalises the Shannon entropy and serves to weigh outcomes with more or less likelihood differently depending on the order α. The Shannon entropy is recovered in the limit $\alpha \to 1$ (for the proof see Appendix B.1). Rényi's original derivation is beyond the scope of this thesis, but it is constructive to note that in his formalism the Shannon entropy corresponds to a linear case of a more general exponential function where the order α acts as the exponent.

Rényi entropic quantities via relative entropy

There have been various quantities with which we could derive definitions of Rényi conditional entropy and Rényi mutual information, the foremost being classical Rényi divergence (Rényi relative entropy) [50]: given two random variables X_1 and X_2 indexed on the same set $\{x\}$ with probability mass functions $p_1(x)$ and $p_2(x)$ respectively, the Rényi divergence is defined

$$D_\alpha(X_1 \| X_2) = \frac{1}{\alpha - 1} \log \sum_x \frac{p_1(x)^\alpha}{p_2(x)^{\alpha-1}}. \tag{1.101}$$

Some specific examples of classical divergences belonging to this family of divergences are

$$D_0(X_1 \| X_2) = -\log \sum_x p_1(x)^0 p_2(x) = -\log \sum_{\{x:p_1(x)>0\}} p_2(x), \tag{1.102}$$

$$D_{1/2}(X_1 \| X_2) = -2 \sum_x \log \sqrt{p_1(x)p_2(x)} \tag{1.103}$$

$$= 2 D_B(X_1, X_2) \quad \textit{(twice the Bhattacharyya distance [53])}, \tag{1.104}$$

$$D_1(X_1 \| X_2) := D(X_1 \| X_2), \tag{1.105}$$

$$D_2(X_1 \| X_2) = \log \sum_x p_1(x) \frac{p_1(x)}{p_2(x)} = \log \left\langle \frac{p_1(x)}{p_2(x)} \right\rangle_{X_1}, \tag{1.106}$$

$$D_\infty(X_1 \| X_2) := \lim_{\alpha \to \infty} D_\alpha(X_1 \| X_2) = \log \sup_x \frac{p_1(x)}{p_2(x)}. \tag{1.107}$$

We would expect the Rényi divergence to induce definitions of the the Rényi entropic quantities in an analogous way to the definition of classical entropies via relative entropy in Section 1.4.1. That is to say, the Rényi conditional entropy and mutual information would be defined

$$H_\alpha(X|Y) = -D_\alpha(p_{XY} \| p_Y) \quad \text{and} \quad I_\alpha(X:Y) = D_\alpha(p_{XY} \| p_X p_Y). \tag{1.108}$$

Since, in the limit $\alpha \to 1$, Eq. (1.101) produces the relative entropy, we expect the above to produce the Shannon and von Neumann quantities in the same limit.

There are some alternate propositions for conditional Rényi entropy from which we could also derive a definition of mutual information.

Alternate definitions

The main alternative to defining mutual information directly with divergence is to define it via the conditional entropy. There are some definitions which have found utility, derived from analogous definitions based on Shannon entropy or generalisations of other related quantities [54]. For example:

$$H_\alpha^{\mathrm{C}}(X|Y) = \sum_y p(y) H_\alpha(X|Y=y) \qquad [55], \qquad (1.109)$$

$$H_\alpha^{\mathrm{JA}}(X|Y) = H_\alpha(XY) - H_\alpha(Y) \qquad [56, 57], \qquad (1.110)$$

$$H_\alpha^{\mathrm{RW}}(X|Y) = \frac{1}{1-\alpha} \max_y \log \sum p(x|y)^\alpha \qquad [58]. \qquad (1.111)$$

Our classical Rényi mutual information could then be defined

$$I_\alpha^*(X:Y) = H_\alpha(X) + H_\alpha(Y) - H_\alpha(XY) \qquad (1.112)$$

$$= H_\alpha(X) - H_\alpha^*(X|Y) = H_\alpha(Y) - H_\alpha^*(Y|X). \qquad (1.113)$$

On closer examination the above does not hold in general. Although similar or related techniques to derive a Rényi measure of information may find use in specific applications, this method is not fruitful for our purposes.

1.4.4 Quantum Rényi entropy

The quantum Rényi entropy is a quantum generalisation of the Rényi entropy and is derived in an analogous way to von Neumann entropy: for a probability density matrix ρ and $\alpha \in (0,1) \cup (1,\infty)$, we define

$$H_\alpha(\rho) = \frac{1}{1-\alpha} \log \mathrm{tr}(\rho^\alpha). \qquad (1.114)$$

There are some particular choices of α which are either mathematically convenient or reflect specific physical situations.

When $\alpha \to 1$ we recover the von Neumann entropy: $H_1(\rho) := H(\rho)$. When $\alpha = 2$ we have the 'collision' entropy: $H_2(\rho) = -\log \mathrm{tr}(\rho^2)$, which characterises the purity of a quantum system (see Eq. (1.32)). The max-entropy could be naturally defined for $\alpha \to 0$ but, due to some mathematical restrictions, we instead use $\alpha = \frac{1}{2}$ (see, for example, Eq. (1.155)). We have

$$H_0(\rho) = \log |\mathrm{supp}(\rho)|, \qquad (1.115)$$

$$H_{\frac{1}{2}}(\rho) = H_{\max}(\rho) = 2\log \mathrm{tr}(\sqrt{\rho}). \qquad (1.116)$$

The last quantity, and perhaps the most useful except for $\alpha \to 1$, is the min-entropy which we obtain when $\alpha \to \infty$:

$$H_\infty(\rho) = H_{\min}(\rho) = -\log \lambda_{\max}[\rho]. \qquad (1.117)$$

1.5 Rényi divergence and related quantities

We now introduce a generalisation of the quantum relative entropy and the classical Rényi divergence – the 'sandwiched' Rényi divergence [28, 29]: for $\rho \in \mathcal{D}(A), \sigma \in \mathcal{P}(A)$ and $\alpha \in (0,1) \cup (1,\infty)$,

$$D_\alpha(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1} \log \mathrm{tr}\left[\left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha\right] & \text{if } \rho \not\perp \sigma \wedge (\sigma \gg \rho \vee \alpha < 1) \\ \infty & \text{else} \end{cases}. \qquad (1.118)$$

From this point we will refer to this quantity as simply 'Rényi divergence'.

It is prudent at this point to formally introduce the following shorthands which are used extensively in the remainder of the thesis. We often treat the relationships $\frac{1}{\alpha} + \frac{1}{\alpha'} = 1$ and $\frac{1}{\alpha} + \frac{1}{\hat{\alpha}} = 2$. Accordingly, we equate

$$\alpha' = \frac{\alpha}{\alpha - 1} \quad \text{and} \quad \hat{\alpha} = \frac{\alpha}{2\alpha - 1}. \tag{1.119}$$

This produces some equivalences that will be useful for later calculations. We have

$$\frac{\alpha'}{\alpha} = \frac{\alpha'(\alpha' - 1)}{\alpha'} = \alpha' - 1, \tag{1.120}$$

$$\frac{\alpha'}{\hat{\alpha}} = \frac{\alpha'(2\alpha - 1)}{\alpha} = \frac{\alpha'}{\alpha}\left(\frac{2\alpha'}{\alpha' - 1} - 1\right) = (\alpha' - 1)\left(\frac{\alpha' + 1}{\alpha' - 1}\right) = \alpha' + 1 \quad \text{and} \tag{1.121}$$

$$-\alpha' = \frac{\alpha}{1 - \alpha} = \left(1 - \frac{\hat{\alpha}}{2\hat{\alpha} - 1}\right)^{-1}\left(\frac{\hat{\alpha}}{2\hat{\alpha} - 1}\right) = \left(\frac{2\hat{\alpha} - 1}{\hat{\alpha} - 1}\right)\left(\frac{\hat{\alpha}}{2\hat{\alpha} - 1}\right) = \frac{\hat{\alpha}}{\hat{\alpha} - 1} = \hat{\alpha}'. \tag{1.122}$$

Applying a similar technique as in Section 1.4.2, we find that equivalence does not extend to Rényi divergence but we can still produce rigorous definitions of the relevant quantum Rényi entropic quantities. The following notation is adapted from the notation introduced in [22]. We define the quantum Rényi entropy as

$$H_\alpha(\rho) = \log |\mathrm{supp}(\rho)|^{\frac{\alpha}{1 - \alpha}} - D_\alpha\left(\rho \,\middle\|\, \frac{\mathbb{1}}{|\mathrm{supp}(x)|}\right). \tag{1.123}$$

We may also derive quantities that generalise the quantum conditional entropy:

$$H_\alpha^\downarrow(A|B)_\rho = -D_\alpha(\rho_{AB} \| \mathbb{1}_A \otimes \rho_B) \quad \text{and} \tag{1.124}$$

$$H_\alpha^\uparrow(A|B)_\rho = -\inf_{\sigma_B \in \mathcal{D}(B)} D_\alpha(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B). \tag{1.125}$$

The ordering $H_\alpha^\downarrow(A|B)_\rho \leq H_\alpha^\uparrow(A|B)_\rho$, obvious from the definition, is indicated by the direction of the superscript arrow. We can safely assume that $\mathbb{1}_A \otimes \sigma_B \gg \rho_{AB}$, since any choice of σ_B where this is not the case would certainly not achieve the infimum. Similarly, we derive generalisations of the quantum mutual information, originally proposed in [27]:

$$I_\alpha^\uparrow(A\,;B)_\rho = \inf_{\sigma_B \in \mathcal{D}(B)} D_\alpha(\rho_{AB} \| \rho_A \otimes \sigma_B) \quad \text{and} \tag{1.126}$$

$$I_\alpha^\downarrow(A:B)_\rho = \inf_{\substack{\sigma_A \in \mathcal{D}(A) \\ \sigma_B \in \mathcal{D}(B)}} D_\alpha(\rho_{AB} \| \sigma_A \otimes \sigma_B). \tag{1.127}$$

Again, the superscript arrows indicate the ordering of each version and we satisfy the support condition as result of the minimisations. The use of ';' in Eq. (1.126) indicates that this quantity is not symmetric between the systems A and B, whereas Eq. (1.127) is.

1.5.1 Generalised Rényi entropic quantities

In order to express our results in a more general setting, we make use of the following notation for the generalised Rényi mutual information [27] and conditional entropy:

$$H_\alpha(\rho_{AB} \| \tau_B) = -D_\alpha(\rho_{AB} \| \mathbb{1}_A \otimes \tau_B), \tag{1.128}$$

$$I_\alpha(\rho_{AB} \| \tau_A) = \inf_{\sigma_B \in \mathcal{D}(B)} D_\alpha(\rho_{AB} \| \tau_A \otimes \sigma_B). \tag{1.129}$$

We can readily verify that the above quantities generalise Eqs. (1.124)-(1.127), i.e.

$$H_\alpha^\downarrow(A|B)_\rho = H_\alpha(\rho_{AB}\|\rho_B), \qquad H_\alpha^\uparrow(A|B)_\rho = \sup_{\sigma_B \in \mathcal{D}(B)} H_\alpha(\rho_{AB}\|\sigma_B), \tag{1.130}$$

$$I_\alpha^\uparrow(A\,;B)_\rho = I_\alpha(\rho_{AB}\|\rho_A), \qquad I_\alpha^\downarrow(A:B)_\rho = \inf_{\sigma_A \in \mathcal{D}(A)} I_\alpha(\rho_{AB}\|\sigma_A). \tag{1.131}$$

1.5.2   Properties of the Rényi divergence

This definition of a quantum Rényi divergence is not the only one proposed, however, it maintains many useful and desirable properties of the quantities it generalises for a wider range Rényi orders, whereas other propositions[†] do not. There are a few properties of the Rényi divergence which we find particularly useful, which we now detail (for a more comprehensive treatment see [28, 29, 38]).

Firstly, it generalises the quantum relative entropy. Indeed, it is recovered by taking the limit $\alpha \to 1$ as in Proposition B.1. It is also monotonically increasing in α [38], i.e for $\alpha > \beta$,

$$D_\alpha(\rho\|\sigma) \geq D_\beta(\rho\|\sigma). \tag{1.132}$$

It is useful to re-express the Rényi divergence as a Schatten norm of the arguments dependent on α:

$$D_\alpha(\rho\|\sigma) = \log \left\| \sigma^{\frac{-1}{2\alpha'}} \rho \sigma^{\frac{-1}{2\alpha'}} \right\|_\alpha^{\alpha'}. \tag{1.133}$$

Its close relation to Schatten norms is mathematically convenient and as such derived quantities exhibit duality relations. For example we have the duality of the Rényi conditional entropy [38, 28]:

Proposition 1.11. *For a pure state $\rho_{ABC} \in \mathcal{D}(ABC)$ and $\alpha > \frac{1}{2}$ such that $\frac{1}{\alpha} + \frac{1}{\hat{\alpha}} = 2$ we have*

$$H_\alpha^\uparrow(A|B)_\rho = -H_{\hat{\alpha}}^\uparrow(A|C)_\rho. \tag{1.134}$$

And the duality of the generalised mutual information [27]:

Proposition 1.12. *For a pure state $\rho_{ABC} \in \mathcal{D}(ABC), \tau_B \in \mathcal{P}(B)$ and $\alpha > \frac{1}{2}$ such that $\frac{1}{\alpha} + \frac{1}{\hat{\alpha}} = 2$ we have*

$$I_\alpha(\rho_{AB}\|\tau_A) = -I_{\hat{\alpha}}(\rho_{AC}\|\tau_A^{-1}). \tag{1.135}$$

Additionally, it satisfies the data-processing inequality [38, 60, 28]:

Proposition 1.13. *If $\alpha \geq \frac{1}{2}$ and $\Phi \in \mathrm{CPTP}(A,B)$, then for $\rho \in \mathcal{D}(A), \sigma \in \mathcal{P}(A)$*

$$D_\alpha(\rho\|\sigma) \geq D_\alpha(\Phi(\rho)\|\Phi(\sigma)). \tag{1.136}$$

In order to extend certain statements about strictly positive matrices to positive semi-definite matrices we can observe the continuity of the Rényi divergence. The following proposition is adapted from [28, Lemma 13] and ensures the Rényi divergence is continuous even when the rank of Y decreases.

Proposition 1.14. *Let $X \in \mathcal{D}(A), Y \in \mathcal{P}(A)$ with $X \neq 0$ and $\alpha \in (0,1) \cup (1,\infty)$. We have*

$$D_\alpha(X\|Y) = \lim_{\varepsilon \to 0^+} \frac{1}{\alpha - 1} \log \mathrm{tr}\left[\left((Y + \varepsilon \mathbb{1}_A)^{\frac{1-\alpha}{2\alpha}} X (Y + \varepsilon \mathbb{1}_A)^{\frac{1-\alpha}{2\alpha}} \right)^\alpha \right] \tag{1.137}$$

[†]The main contender in this context is the Petz quantum Rényi divergence. It is a simpler generalisation but only satisfies many desired properties of a divergence for $\alpha \in [0,2]$. See [42] and [59] for more details.

and the limit exists in the weaker sense in which a real valued sequence which is bounded from below and not bounded from above and which does not have an accumulation point is considered as being convergent to $+\infty$.

Essentially, this indicates that if we can make a statements for strictly positive matrices, e.g. $Y + \varepsilon \mathbb{1}_A \in \mathcal{P}^*(A)$, then we can extend that statement to positive semi-definite matrices by taking $Y + \varepsilon \mathbb{1}_A \to Y$ with $\text{supp}(Y) \subset \mathcal{H}_A$. Note however, that in some cases the quantity will still diverge if $Y \not\geq X$.

These properties naturally extend to any quantity defined using Rényi divergence – properties which coincide with the mathematical and physical interpretation of quantum entropies. That is to say, monotonicity in α reflects the expected behaviour of Rényi entropy when weighing more or less likely outcomes differently and the data-processing inequality reflects that entropy can only ever increase (or correlation decrease) when information is processed (on each system independently).

1.5.3 Dupuis' chain rules

One of the major contributions of this work is to formalise and provide Rényi approximations of the chain and decomposition rules that exist for Shannon and von Neumann entropies and which often are an important tool in the derivation of related results. To this end, Dupuis' chain rule [37] provides a foundation for the methodology and a proof-of-concept for the more general inequalities covered in Chapter 3.

Theorem 1.15. *Let $\frac{\alpha}{\alpha-1} = \frac{\beta}{\beta-1} + \frac{\gamma}{\gamma-1}$ with $\alpha, \beta, \gamma \in (1/2, 1) \cup (1, \infty)$. For $\rho_{ABC} \in \mathcal{D}(ABC)$ and $\tau_C \in \mathcal{P}(C)$, if $(\alpha-1)(\beta-1)(\gamma-1) > 0$,*

$$H_\alpha(\rho_{ABC} \| \tau_C) \geq H_\beta^\uparrow(A|BC)_\rho + H_\gamma(\rho_{BC} \| \tau_C). \tag{1.138}$$

Otherwise, if $(\alpha-1)(\beta-1)(\gamma-1) < 0$,

$$H_\alpha(\rho_{ABC} \| \tau_C) \leq H_\beta^\uparrow(A|BC)_\rho + H_\gamma(\rho_{BC} \| \tau_C). \tag{1.139}$$

This result shows that we may establish chain rules for conditional Rényi entropies for particular choices of Rényi order. Evidently, these chain rules are weaker than the Shannon and von Neumann versions, but still provide a rigorous generalisation – note that by pinching the inequalities we recover an equality when all orders tend to 1. We provide an alternative proof of this theorem and a variation on the bipartite case in Chapter 3.

1.6 Entropic uncertainty relations

Uncertainty relations manifest in many areas of science, and can be applied to situations ranging across the accuracy of radar readings (trajectory vs position), the analysis of sound waves (frequency vs instant) and the measurement of subatomic particles (position vs momentum).

1.6.1 Commutator formulation

In quantum science, uncertainty relations are a way to characterise what are referred to as 'incompatible' observables. Observables in this context are operators whose action yields something which can be observed about the state of a quantum system. Incompatibility can be measured by the commutator $[C, D]$, a function of operators. Only when $[C, D] = 0$ are the observables said to commute, otherwise they are incompatible. Note that observables commute only if they are simultaneously diagonalisable.

As explained in Section 1.3.4, when an observable M acts on a quantum state $|\psi\rangle$, we can recover the probability $P(m)$ that a certain value m will be measured for that observable. Using this notation we write the expected value of the observable

$$\langle\psi|M|\psi\rangle = \langle M\rangle = \sum_m mP(m) \tag{1.140}$$

and its standard deviation $\sigma(M) = \sqrt{\langle M^2\rangle - \langle M\rangle^2}$, which we find by examining the expected difference of the values m from the expected value:

$$(\sigma(M))^2 = \langle(M - \langle M\rangle)^2\rangle \tag{1.141}$$
$$= \langle M^2 - 2M\langle M\rangle + \langle M\rangle^2\rangle \tag{1.142}$$
$$= \langle M^2\rangle - \langle M\rangle^2. \tag{1.143}$$

Taking the square root will bring things back to the same dimensionality as the observable. Standard deviation is a measure of how spread out a distribution is, and as such is connected to the uncertainty in the measurement of that observable.

The Robertson relation [61] is described by the inequality

$$\sigma(C)\sigma(D) \geq \frac{|\langle[C, D]\rangle|}{2}. \tag{1.144}$$

This implies the combined spread of the observables must be greater than a function of their incompatibility, the right-hand side is always positive and only equal to zero for observables that commute on the support of $|\psi\rangle$. This allows the left-hand side to be a trivial bound for some situations where the eigenstates of the observables are orthogonal to $|\psi\rangle$.

In general, for non-commuting observables, if the outcomes of one observable can be predicted with a high level of accuracy (small standard deviation) then the uncertainty in predicting the outcome of the other observable must be proportionately high (large standard deviation).

Arguably the most famous application of this relation is the *Heisenberg Uncertainty Principle* [35] relating position and momentum:

$$\sigma(X)\sigma(P) \geq \frac{\hbar}{2}, \tag{1.145}$$

where X is the position observable, P is the momentum observable and $\hbar$ is the reduced Planck's constant, equal to the expected value of the their commutator [62, 63].

1.6.2 Entropic formulation

To discuss entropic uncertainty relations we consider finite quantum states represented by probability density matrices measured in a choice of ONB (see Section 1.3.4).

Uncertainty in measuring in two orthonormal bases

Since for each $\rho \in \mathcal{D}(A)$ we can recover the random variables X and Z for two ONBs $\mathbb{X}$ and $\mathbb{Z}$ respectively, we can examine the Shannon entropies $H(X)_\rho$ and $H(Z)_\rho$. Note that this is not the entropy associated with the state ρ, or even the post-measurement state, but rather the entropy associated with the measurement of ρ in either basis. We then have the canonical entropic uncertainty relation [36]:

Theorem 1.16. *For a mixed state $\rho_A \in \mathcal{D}(A)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A we have*

$$H(X)_\rho + H(Z)_\rho \geq \log \frac{1}{c} =: q_{\mathrm{MU}}, \quad c = \max_{x,z} |\langle x|z \rangle|^2. \tag{1.146}$$

This indicates that the total uncertainty of measuring two different observables on the same system must be at least as large as a positive constant depending on the measurement bases but not on the measured state.

We may extend this to a result involving conditional entropy via the following rationale: consider a generic Shannon entropic relation

$$\sum_n H(X_n) \geq q, \tag{1.147}$$

with q state independent. For some classical memory Y containing information about the preparation of the state being measured, we can show [20, Section IV.C] that Eq. (1.147) implies $\sum_n H(X_n|Y) \geq q$. We may therefore conclude that

$$H(X|Y)_\rho + H(Z|Y)_\rho \geq q_{\mathrm{MU}}. \tag{1.148}$$

Information exclusion relations

Also known as mutual information uncertainty relations, information exclusion relations are entropic uncertainty relations expressed in terms of mutual information.

The canonical version of this type of relation is the Hall relation [19] which we may derive from Eq. (1.148) with an application of the decomposition rule, Eq. (1.69):

Theorem 1.17. *For a mixed state $\rho_{AB} \in \mathcal{D}(A)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A and $\mathbb{Y}$ on B we have*

$$I(X:Y)_\rho + I(Z:Y)_\rho \leq \log(d^2 c) =: r_{\mathrm{H}}, \tag{1.149}$$

where d is the dimension of $\mathcal{H}_A$.

Note d appears due to both $H(X)$ and $H(Y)$ being bounded above by $\log d$.

1.6.3 Improvements and extensions

Here we detail some of the pre-existing results relevant to the generalised entropic uncertainty relations in Chapter 4. These take the form of improved bounds, tighter inequalities, extensions to Rényi entropies or a combination thereof.

Improved bounds

Berta et al. [43] have improved on Eq. (1.148) by considering the conditional entropy of the state ρ, which quantifies its inherent mixedness:

Theorem 1.18. *For a mixed state $\rho_{AB} \in \mathcal{D}(AB)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A we have*

$$H(X|B)_\rho + H(Z|B)_\rho \geq q_{\mathrm{MU}} + H(A|B)_\rho. \tag{1.150}$$

With this relation as a starting point, Coles and Piani [44] have derived a tighter, state-dependent version:

Theorem 1.19. *For a mixed state $\rho_{AB} \in \mathcal{D}(AB)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A we have*

$$H(X|B)_\rho + H(Z|B)_\rho \geq q(\rho) + H(A|B)_\rho, \tag{1.151}$$

where $q(\rho) := \max\{q(\rho, \mathbb{X}, \mathbb{Z}), q(\rho, \mathbb{Z}, \mathbb{X})\}$ and $q(\rho_A, \mathbb{A}, \mathbb{B}) := -\sum_a p(a) \log \left(\max_b |\langle a|b \rangle|^2 \right)$.

This can be weakened slightly to a state-independent version:

Theorem 1.20. *For a mixed state $\rho_{AB} \in \mathcal{D}(AB)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A we have*

$$H(X|B)_\rho + H(Z|B)_\rho \geq q_{\mathrm{CP}} + H(A|B)_\rho, \tag{1.152}$$

where $q_{\mathrm{CP}} := \min_{\sigma \in \mathcal{D}(AB)} q(\sigma)$.

These relations are then used to produce improved versions of information exclusion relation in Eq. (1.149). Firstly, they proved the conjecture of Grudka et al. [64], who proposed the following:

Proposition 1.21. *For a mixed state $\rho_{AB} \in \mathcal{D}(AB)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A we have*

$$I(X:B)_\rho + I(Z:B)_\rho \leq r_{\mathrm{G}} \quad where \quad r_{\mathrm{G}} := \log \left(d \sum_{d\ largest} \max_{x,z} |\langle x|z \rangle|^2 \right), \tag{1.153}$$

with the sum over the largest d terms of the matrix $\left[|\langle x|z \rangle|^2 \right]$.

Indeed, this is actually a weaker version of the main result of Coles and Piani:

Theorem 1.22. *For a mixed state $\rho_{AB} \in \mathcal{D}(AB)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A we have*

$$I(X:B)_\rho + I(Z:B)_\rho \leq r_{\mathrm{CP}}, \tag{1.154}$$

where $r_{\mathrm{CP}} := \min\{r(\mathbb{X}, \mathbb{Z}), r(\mathbb{Z}, \mathbb{X})\}$ and $r(\mathbb{A}, \mathbb{B}) := \log \left(d \sum_a \max_b |\langle a|b \rangle|^2 \right)$.

Extensions to Rényi entropies

We now consider the known extensions of these relations to Rényi entropies of specific order. In fact, to begin we need look no further than the Maassen-Uffink relation as it was originally proposed [36] in the general form:

Theorem 1.23. *For a mixed state $\rho_A \in \mathcal{D}(A)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A with $\alpha, \hat{\alpha} \geq \frac{1}{2}$ such that $\frac{1}{\alpha} + \frac{1}{\hat{\alpha}} = 2$,*

$$H_\alpha(X) + H_{\hat{\alpha}}(Z) \geq q_{\mathrm{MU}}. \tag{1.155}$$

Note for $\alpha, \beta = 1$ we recover Eq. (1.148).

This not only establishes a strong precedent for Rényi uncertainty relations but reveals that they are a fundamental part of the mathematical structure of this field.

Tomamichel [42] derived the main improvement of this relation, described in terms of quantum Rényi conditional entropy:

Theorem 1.24. *For a mixed state $\rho_{AB} \in \mathcal{D}(AB)$ and measurement bases $\mathbb{X}$ and $\mathbb{Z}$ on A with $\frac{\alpha}{\alpha-1} = \frac{\gamma}{\gamma-1} + \frac{\beta}{\beta-1}$, and $(\alpha-1)(\beta-1)(\gamma-1) < 0$.*

$$H_\gamma^\uparrow(X|B) + H_\beta^\uparrow(Z|B) \geq q_{\mathrm{MU}} + H_\alpha^\uparrow(A|B). \tag{1.156}$$

This result generalises both Eq. (1.150) and Eq. (1.155), respectively, by letting all parameters go to 1 or choosing B trivial and allowing $\alpha \to 0$.

Once adequate Rényi entropy chain and decomposition rules are established, we may adapt the methodology used for the derivations and proofs of these improved relations to demonstrate the results in Chapter 4.

Chapter 2

Interpolation framework

The principle mechanism which we use to compare Rényi divergence of different order is complex interpolation. There is a well-established and sophisticated theory associated with this field of mathematics but in our context we do not need to go into such depth. Our result is essentially an extension of the Hadamard three-line theorem and, although Hadamard's result can be derived via the broader theory, we purposefully restrict our focus. However, in order to understand the motivation behind the choices of interpolation function it is prudent to provide an overview of the interpolation structure and associated norms of Pisier. The statement and proof of the three-line theorem for unbalanced weighted norms on linear operators are given in the last section of this chapter and the specialised results that follow are given in Chapter 3, in the context in which they are used.

2.1 The super-operator $\Gamma_{\sigma,\tau}$

For conciseness in the subsequent chapters, we introduce a more general form of the super-operator notation found in [38]: with $\sigma \in \mathcal{P}^*(B), \tau \in \mathcal{P}^*(A)$, define

$$\Gamma_{\sigma,\tau} : \mathcal{L}(A,B) \longrightarrow \mathcal{L}(A,B) \tag{2.1}$$

$$: \qquad M \longmapsto \sigma^{\frac{1}{2}} M \tau^{\frac{1}{2}}. \tag{2.2}$$

If σ or τ are instead operators on only part of a tensor product we still use the above notation when there is no ambiguity, e.g. for $\sigma_C \in \mathcal{P}(C), \tau_A \in \mathcal{P}(A)$ and $M \in \mathcal{L}(AB, CD)$

$$\Gamma_{\sigma_C,\tau_A}^{\frac{1}{\lambda}}(M) = \left(\sigma_C^{\frac{1}{2\lambda}} \otimes \mathbb{1}_D \right) M \left(\tau_A^{\frac{1}{2\lambda}} \otimes \mathbb{1}_B \right). \tag{2.3}$$

In Eq. (2.1) we may choose $\mathcal{H}_B = \mathcal{H}_A$. Moreover, we write $\Gamma_{\sigma,\sigma}^{\frac{1}{\lambda}} = \Gamma_{\sigma}^{\frac{1}{\lambda}}$.

2.1.1 Unbalanced weighted norms

We define a norm on $\mathcal{L}(A,B)$ in the following way: given $\sigma \in \mathcal{P}^*(B)$ and $\tau \in \mathcal{P}^*(A)$, let

$$\|Y\|_{p,(\sigma,\tau)} = \left\| \sigma^{\frac{1}{2p}} Y \tau^{\frac{1}{2p}} \right\|_p = \left\| \Gamma_{\sigma,\tau}^{\frac{1}{p}}(Y) \right\|_p. \tag{2.4}$$

It is relatively straight-forward to show this is indeed a norm for $1 \leq p \leq \infty$.

By defining the inner product $\langle Y, X \rangle_{\sigma,\tau} = \mathrm{tr}\left(Y^\dagger \sigma^{\frac{1}{2}} X \tau^{\frac{1}{2}} \right)$ we may show the following duality: for $1 \leq p \leq \infty$,

$$\|X\|_{p,(\sigma,\tau)} = \sup_{\|Y\|_{p',(\sigma,\tau)}=1} \left| \langle Y, X \rangle_{\sigma,\tau} \right|, \tag{2.5}$$

where p' is the Hölder conjugate such that $\frac{1}{p} + \frac{1}{p'} = 1$.

Proof of Eq. (2.5). By the definition of the dual norm [65, Section 2.10] we may write

$$\|X\|_{p,(\sigma,\tau)}^{*} = \sup_{Y \in \mathcal{L}(A,B)} \frac{\left|\langle Y, X\rangle_{\sigma,\tau}\right|}{\|Y\|_{p,(\sigma,\tau)}} \tag{2.6}$$

$$= \sup_{Y \in \mathcal{L}(A,B)} \frac{\left|\mathrm{tr}\left(Y^{\dagger}\sigma^{\frac{1}{2}}X\tau^{\frac{1}{2}}\right)\right|}{\left\|\sigma^{\frac{1}{2p}}Y\tau^{\frac{1}{2p}}\right\|_{p}} \tag{2.7}$$

$$= \sup_{Y \in \mathcal{L}(A,B)} \frac{\left|\mathrm{tr}\left((\sigma^{\frac{-1}{2p}}Y\tau^{\frac{-1}{2p}})^{\dagger}\sigma^{\frac{1}{2}}X\tau^{\frac{1}{2}}\right)\right|}{\|Y\|_{p}} \tag{2.8}$$

$$= \sup_{Y \in \mathcal{L}(A,B)} \frac{\left|\mathrm{tr}\left(Y^{\dagger}\sigma^{\frac{-1}{2p}}\sigma^{\frac{1}{2}}X\tau^{\frac{1}{2}}\tau^{\frac{-1}{2p}}\right)\right|}{\|Y\|_{p}} \tag{2.9}$$

$$= \sup_{Y \in \mathcal{L}(A,B)} \frac{\left|\mathrm{tr}\left(Y^{\dagger}\sigma^{\frac{1}{2p'}}X\tau^{\frac{1}{2p'}}\right)\right|}{\|Y\|_{p}} \tag{2.10}$$

$$= \left\|\sigma^{\frac{1}{2p'}}X\tau^{\frac{1}{2p'}}\right\|_{p'} \tag{2.11}$$

$$= \|X\|_{p',(\sigma,\tau)}. \tag{2.12}$$

$\square$

The above implies that a Hölder type inequality holds for these norms, i.e.

$$\left|\langle Y, X\rangle_{\sigma,\tau}\right| \le \|X\|_{p,(\sigma,\tau)}\|Y\|_{p',(\sigma,\tau)}, \quad 1 \le p, p' \le \infty. \tag{2.13}$$

We must take note that in general these weighted norms are not unitarily invariant, however we may state a specialised unitary invariance: for U and V unitary such that $[U,\sigma] = [V,\tau] = 0$,

$$\|UXV\|_{p,(\sigma,\tau)} = \|UX\|_{p,(\sigma,\tau)} = \|XV\|_{p,(\sigma,\tau)} = \|X\|_{p,(\sigma,\tau)}. \tag{2.14}$$

2.1.2 Rényi Divergence in terms of weighted norms

We may establish the following identities for Rényi entropic quantities in terms of super-operator notation:

Lemma 2.1. *For $M \in \mathcal{L}(AB, C)$ such that $M^{\dagger}M = \rho_{AB}$, $\tau_B \in \mathcal{P}^{*}(B)$ and $\alpha \in (0,1) \cup (1,\infty)$:*

$$H_{\alpha}(\rho_A) = -\log \sup_{\sigma_A \in \mathcal{D}^{*}(A)} \left\|\Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\alpha'}}(M)\right\|_{2}^{2\alpha'}, \tag{2.15}$$

$$H_{\alpha}(\rho_{AB}\|\tau_B) = -\log \left\|\Gamma_{\mathbb{1}_C,\tau_B}^{\frac{-1}{\alpha'}}(M)\right\|_{2\alpha}^{2\alpha'}, \tag{2.16}$$

$$I_{\alpha}(\rho_{AB}\|\tau_B) = \log \inf_{\sigma_A \in \mathcal{D}(A)} \left\|\Gamma_{\mathbb{1}_C,\sigma_A \otimes \tau_B}^{\frac{-1}{\alpha'}}(M)\right\|_{2\alpha}^{2\alpha'}. \tag{2.17}$$

Before we present the proof of Lemma 2.1 we find it useful to examine a more general comparison between Rényi divergence and weighted norms.

Lemma 2.2. *Assume $\alpha \in (0,1) \cup (1,\infty)$. With $M \in \mathcal{L}(AB, C)$ such that $M^{\dagger}M = \rho_{AB} \in \mathcal{D}(AB)$, $\sigma_A \in \mathcal{P}(A)$ and $\tau_B \in \mathcal{P}^{*}(B)$, we have*

$$\log \left\|\Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\alpha}-\frac{1}{\lambda}}\left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M)\right)\right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)}^{2\alpha'} = D_{\alpha}\left(\rho_{AB}\left\|\sigma_A^{1-\frac{\alpha'}{\lambda'}} \otimes \tau_B\right.\right). \tag{2.18}$$

Proof.

$$\left\| \Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\alpha}-\frac{1}{\lambda}}(M) \right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)}^2 = \left\| \Gamma_{\sigma_A}^{\frac{1}{\alpha}-\frac{1}{\lambda}}(\rho_{AB}) \right\|_{\alpha,\tau_B} \tag{2.19}$$

$$= \left\| \left(\sigma_A^{\frac{1}{2\alpha}-\frac{1}{2\lambda}} \otimes \tau_B^{\frac{1}{2\alpha}} \right) \rho_{AB} \left(\sigma_A^{\frac{1}{2\alpha}-\frac{1}{2\lambda}} \otimes \tau_B^{\frac{1}{2\alpha}} \right) \right\|_\alpha \tag{2.20}$$

$$= \left\| \left(\sigma_A^{\frac{1}{2\lambda'}-\frac{1}{2\alpha'}} \otimes \tau_B^{\frac{-1}{2\alpha'}} \right) \Gamma_{\tau_B}(\rho_{AB}) \left(\sigma_A^{\frac{1}{2\lambda'}-\frac{1}{2\alpha'}} \otimes \tau_B^{\frac{-1}{2\alpha'}} \right) \right\|_\alpha \tag{2.21}$$

$$\implies \left\| \Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\alpha}-\frac{1}{\lambda}}\left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M) \right) \right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)}^2 = \left\| \left(\sigma_A^{1-\frac{\alpha'}{\lambda'}} \otimes \tau_B \right)^{\frac{-1}{2\alpha'}} \rho_{AB} \left(\sigma_A^{1-\frac{\alpha'}{\lambda'}} \otimes \tau_B \right)^{\frac{-1}{2\alpha'}} \right\|_\alpha \tag{2.22}$$

$$\implies \log \left\| \Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\alpha}-\frac{1}{\lambda}}\left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M) \right) \right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)}^{2\alpha'} = D_\alpha \left(\rho_{AB} \left\| \sigma_A^{1-\frac{\alpha'}{\lambda'}} \otimes \tau_B \right. \right). \tag{2.23}$$

$\square$

We now make use of the above comparison to prove Lemma 2.1.

Proof of Lemma 2.1. By definition we have

$$H_\alpha(\rho_A) = -D_\alpha(\rho_A \| \mathbb{1}_A) \tag{2.24}$$

$$= -\log \|\rho_A\|_\alpha^{\alpha'} \tag{2.25}$$

$$= -\log \sup_{\sigma_A \in \mathcal{D}^*(A)} \left[\mathrm{tr}\left(\rho_A \sigma_A^{\frac{1}{\alpha'}} \right)^{\alpha'} \right] \tag{2.26}$$

$$= -\log \sup_{\sigma_A \in \mathcal{D}^*(A)} \left[\mathrm{tr}\left(\Gamma_{\sigma_A}^{\frac{1}{\alpha'}}(\rho_{AB}) \right)^{\alpha'} \right] \tag{2.27}$$

$$= -\log \sup_{\sigma_A \in \mathcal{D}^*(A)} \left\| \Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\alpha'}}(M) \right\|_2^{2\alpha'}, \tag{2.28}$$

where in the third line we use Lemma B.2. Note also that in the fourth line, the partial trace on B does not affect operators localised on A.

To show Eq. (2.16) we begin with the comparison in Lemma 2.2 and choose $\lambda = \alpha$ in Eq. (2.18):

$$H_\alpha(\rho_{AB}\|\tau_B) = -D_\alpha(\rho_{AB}\| \mathbb{1}_A \otimes \tau_B) \tag{2.29}$$

$$= -\log \left\| \Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M) \right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)}^{2\alpha'} \tag{2.30}$$

$$= -\log \left\| \Gamma_{\mathbb{1}_C,\tau_B}^{\frac{1}{\alpha}-1}(M) \right\|_{2\alpha}^{2\alpha'} \tag{2.31}$$

$$= -\log \left\| \Gamma_{\mathbb{1}_C,\tau_B}^{\frac{-1}{\alpha'}}(M) \right\|_{2\alpha}^{2\alpha'}. \tag{2.32}$$

Moreover, choosing instead $\lambda = 1$ and $\tau_B = \mathbb{1}_B$,

$$H_\alpha(\rho_{AB}\|\sigma_A) = -D_\alpha(\rho_{AB}\|\sigma_A \otimes \mathbb{1}_B) \tag{2.33}$$

$$= -\log \left\| \Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{-1}{\alpha'}}(M) \right\|_{2\alpha}^{2\alpha'}. \tag{2.34}$$

After relabelling in the second case we obtain Eq. (2.16). Similarly, to show Eq. (2.17) we start

with Eq. (2.18), choosing $\lambda = 1$:

$$I_\alpha(\rho_{AB}\|\tau_B) = \log \inf_{\sigma_A \in \mathcal{D}(A)} D(\rho_{AB}\|\sigma_A \otimes \tau_B) \tag{2.35}$$

$$= \log \inf_{\sigma_A \in \mathcal{D}(A)} \left\| \Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{-1}{\alpha'}} \left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M) \right) \right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)}^{2\alpha'} \tag{2.36}$$

$$= \log \inf_{\sigma_A \in \mathcal{D}(A)} \left\| \Gamma_{\mathbb{1}_C,\sigma_A \otimes \tau_B}^{\frac{-1}{\alpha'}}(M) \right\|_{2\alpha}^{2\alpha'}. \tag{2.37}$$

Moreover,

$$I_\alpha(\rho_{AB}\|\sigma_A) = \log \inf_{\tau_B \in \mathcal{D}(B)} \left\| \Gamma_{\mathbb{1}_C,\sigma_A \otimes \tau_B}^{\frac{-1}{\alpha'}}(M) \right\|_{2\alpha}^{2\alpha'}. \tag{2.38}$$

Again, after relabelling, we arrive at Eq. (2.17). $\qquad\square$

2.2 Pisier's norms

Although not used directly in the main results of this thesis, the non-commutative norms of Pisier still inform the interpolation structure and were crucial in understanding the choices which produce the desired comparisons. Their relation to Rényi divergence and related quantities has already been put to use in other areas of quantum information [40, 41].

In this section we cover the basic concepts, some insights which further simplify things in our context of positive operators and, ultimately, how they factor into the interpolation results of Chapter 3.

2.2.1 Two-part norms

Based on the work of Pisier [39] and the subsequent refinements in [40] we can use complex interpolation to derive the following definition of a non-commutative norm on the product of complex Hilbert spaces of finite linear operators:

Theorem 2.3. *For $1 \le q \le p \le \infty$ there is a unique $r \in [1,\infty]$ such that $\frac{1}{q} = \frac{1}{p} + \frac{1}{r}$. With $Y \in \mathcal{L}(AB)$ the identities*

$$\|Y\|_{(p,q)} \equiv \sup_{M,N \in \mathcal{L}(A)} \frac{\|(M \otimes \mathbb{1}_B)Y(N \otimes \mathbb{1}_B)\|_q}{\|M\|_{2r}\|N\|_{2r}} \tag{2.39}$$

and

$$\|Y\|_{(q,p)} \equiv \inf_{\substack{M,N \in \mathcal{L}(A) \\ Y=(M \otimes \mathbb{1}_B)Z(N \otimes \mathbb{1}_B)}} \|M\|_{2r}\|N\|_{2r}\|Z\|_p \tag{2.40}$$

define a norm.

Although the above expressions are required for the more general case, we may take advantage of our specific context to re-express them in a more workable form.

Proposition 2.4. *The expressions in Eq. (2.39) and Eq. (2.40) are equivalent to*

$$\|Y\|_{(p,q)} = \begin{cases} \sup_{\sigma,\tau \in \mathcal{D}^*(A)} \|\Gamma_{\sigma,\tau}^{\frac{-1}{p}}(Y)\|_{q,(\sigma,\tau)} & \text{if } p \ge q \\ \inf_{\sigma,\tau \in \mathcal{D}^*(A)} \|\Gamma_{\sigma,\tau}^{\frac{-1}{p}}(Y)\|_{q,(\sigma,\tau)} & \text{if } p \le q \end{cases}. \tag{2.41}$$

Proof. First we note that for all $M, N \in \mathcal{L}(A)$ we have a left and right polar decomposition, i.e. there exists unitary operators $U, V \in \mathcal{U}(A)$ and positive semi-definite operators $P, Q \in \mathcal{P}(A)$ such

that $M = UP$ and $N = QV$. Since Schatten norms are unitarily invariant we have that Eq. (2.39) becomes

$$\sup_{P,Q\in\mathcal{P}(A)} \frac{\|(P\otimes\mathbb{1}_B)Y(Q\otimes\mathbb{1}_B)\|_q}{\|P\|_{2r}\|Q\|_{2r}} = \sup_{P,Q\in\mathcal{P}(A)} \frac{\left\|\Gamma^2_{P,Q}(Y)\right\|_q}{\|P\|_{2r}\|Q\|_{2r}}. \tag{2.42}$$

In order to achieve the supremum we may consider only P and Q such that $\|P\|_{2r}, \|Q\|_{2r} < \infty$ and, without loss of generality, we can further use linearity and the scalability of Schatten norms to reduce P and Q to operators such that $\|P\|_{2r} = \|Q\|_{2r} = 1$, in which case we can identify them with density operators with full support, i.e.

$$\sup_{\sigma,\tau\in\mathcal{D}^*(A)} \frac{\left\|\Gamma^{\frac{1}{r}}_{\sigma,\tau}(Y)\right\|_q}{\left\|\sigma^{\frac{1}{2r}}\right\|_{2r}\left\|\tau^{\frac{1}{2r}}\right\|_{2r}} = \sup_{\sigma,\tau\in\mathcal{D}^*(A)} \left\|\Gamma^{\frac{1}{r}}_{\sigma,\tau}(Y)\right\|_q. \tag{2.43}$$

Note $\frac{1}{2r} = \frac{1}{2q} - \frac{1}{2p}$, hence

$$\left\|\Gamma^{\frac{1}{r}}_{\sigma,\tau}(Y)\right\|_q = \left\|\Gamma^{\frac{1}{q}}_{\sigma,\tau}(\Gamma^{\frac{-1}{p}}_{\sigma,\tau}(Y))\right\|_q = \left\|\Gamma^{\frac{-1}{p}}_{\sigma,\tau}(Y)\right\|_{q,(\sigma,\tau)}. \tag{2.44}$$

Similarly, if we start with Eq. (2.40) we arrive at the second case of Eq. (2.41). □

This formulation includes some cases of particular interest.

Corollary 2.5.

$$\|Y\|_{(\infty,q)} = \sup_{\sigma,\tau\in\mathcal{D}^*(A)} \|Y\|_{q,(\sigma,\tau)}, \tag{2.45}$$

$$\|Y\|_{(1,q)} = \inf_{\sigma,\tau\in\mathcal{D}^*(A)} \left\|\Gamma^{-1}_{\sigma,\tau}(Y)\right\|_{q,(\sigma,\tau)}, \tag{2.46}$$

$$\|Y\|_{(p,p)} = \|Y\|_p. \tag{2.47}$$

Proof. In Eq. (2.41), replace p with the relevant values. □

An important feature of these norms is that they extend the log-convexity of Schatten norms:

Theorem 2.6. *For $\theta \in (0,1)$, if $0 \le q_0 < q_1 \le \infty$, $0 \le p_0 < p_1 \le \infty$ and*

$$\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}, \quad \frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1} \tag{2.48}$$

then

$$\|Y\|_{(p_\theta,q_\theta)} \le \|Y\|^{1-\theta}_{(p_0,q_0)}\|Y\|^{\theta}_{(p_1,q_1)}. \tag{2.49}$$

This estimate follows from the original formulation of Pisier [39] for more general spaces, but showing this in a self-contained way for our purposes has proven non-trivial. Although it is related to our three-line theorem – Theorem 2.12, and the derived log-convexity results, it is not actually required to demonstrate them. We include it here as an indication of the potential of the general framework, of which we are only utilising a specific part.

2.2.2 The case of positive operators

If we restrict the focus to positive operators we can show that the optimisation is achieved for the same operator, i.e. $\sigma = \tau \in \mathcal{P}^*(A)$. This result and some of the subsequent observations were originally explored in [40]. It follows that we have a special case of Proposition 2.4:

Proposition 2.7. *For $X \in \mathcal{P}(AB)$,*

$$\|X\|_{(p,q)} = \begin{cases} \sup\limits_{\sigma \in \mathcal{D}^*(A)} \left\|\Gamma_\sigma^{\frac{-1}{p}}(X)\right\|_{q,\sigma} & \text{if } p \geq q \\ \inf\limits_{\sigma \in \mathcal{D}^*(A)} \left\|\Gamma_\sigma^{\frac{-1}{p}}(X)\right\|_{q,\sigma} & \text{if } p \leq q \end{cases}. \tag{2.50}$$

Proof. For $X \in \mathcal{P}(A), M \in \mathcal{L}(A)$, Hölder's inequality implies

$$\|MXN^\dagger\|_p^2 = \|(MX^{\frac{1}{2}}X^{\frac{1}{2}}N^\dagger)\|_p^2 \tag{2.51}$$

$$\leq \|MX^{\frac{1}{2}}\|_{2p}^2 \|X^{\frac{1}{2}}N^\dagger\|_{2p}^2 \tag{2.52}$$

$$= \|MXM^\dagger\|_p \|NXN^\dagger\|_p \tag{2.53}$$

$$\|MXN^\dagger\|_p \leq \sqrt{\|MXM^\dagger\|_p \|NXN^\dagger\|_p} \tag{2.54}$$

$$\leq \max\{\|MXM^\dagger\|_p, \|NXN^\dagger\|_p\}. \tag{2.55}$$

Moreover, unitary invariance implies $\|MXM^\dagger\|_p = \||M|X|M|\|_p$.

$\square$

This special case gives rise to the following identity:

Proposition 2.8. *Given $X_{AB} \in \mathcal{P}(AB)$, $\|X_{AB}\|_{(p,1)} = \|X_A\|_p$.*

Proof. Using the duality of Schatten norms we can write

$$\|X_{AB}\|_{(p,1)} = \sup_{\sigma \in \mathcal{D}^*(A)} \left\|\Gamma_\sigma^{\frac{-1}{p}}(X_{AB})\right\|_{1,\sigma} \tag{2.56}$$

$$= \sup_{\sigma \in \mathcal{D}^*(A)} \|\sigma^{\frac{1}{2p'}} \operatorname{tr}_B(X_{AB})\sigma^{\frac{1}{2p'}}\|_1 \tag{2.57}$$

$$= \sup_{\sigma \in \mathcal{D}^*(A)} \operatorname{tr} X_A \sigma^{\frac{1}{p'}} \tag{2.58}$$

$$= \|X_A\|_p. \tag{2.59}$$

$\square$

2.2.3 Weighted two-part norms

Based on the definition of weighted norms in section 2.1.1, we may define a weighted version of the two-part norm on $\mathcal{L}(A,B)$, with $\sigma \in \mathcal{P}^*(B)$ and $\tau \in \mathcal{P}^*(A)$:

$$\|X\|_{p,q,(\sigma,\tau)} = \|\Gamma_{\sigma,\tau}^{\frac{1}{q}}(X)\|_{(p,q)}. \tag{2.60}$$

This definition may not be particularly useful in general but does allow for the following identity:

Proposition 2.9. *For $Y \in \mathcal{L}(AB)$ and $\sigma_B, \tau_B \in \mathcal{P}^*(B)$ we have*

$$\|Y\|_{p,q,(\sigma_B,\tau_B)} = \begin{cases} \sup\limits_{\sigma_A,\tau_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A,\tau_A}^{\frac{-1}{p}}(Y)\|_{q,(\sigma_A \otimes \sigma_B, \tau_A \otimes \tau_B)} & \text{if } p \geq q \\ \inf\limits_{\sigma_A,\tau_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A,\tau_A}^{\frac{-1}{p}}(Y)\|_{q,(\sigma_A \otimes \sigma_B, \tau_A \otimes \tau_B)} & \text{if } p \leq q \end{cases}. \tag{2.61}$$

Proof.

$$\left\| \Gamma_{\sigma_A,\tau_A}^{\frac{-1}{p}} \left(\Gamma_{\sigma_B,\tau_B}^{\frac{1}{q}}(Y) \right) \right\|_{q,(\sigma_A,\tau_A)} = \left\| (\sigma_A \otimes \sigma_B)^{\frac{1}{2q}} (\sigma_A \otimes \mathbb{1}_B)^{\frac{-1}{2p}} Y (\sigma_A \otimes \mathbb{1}_B)^{\frac{-1}{2p}} (\tau_A \otimes \tau_B)^{\frac{1}{2q}} \right\|_q$$

(2.62)

$$= \left\| \Gamma_{\sigma_A,\tau_A}^{\frac{-1}{p}}(Y) \right\|_{q,(\sigma_A \otimes \sigma_B, \tau_A \otimes \tau_B)}.$$

(2.63)

Moreover $\|\Gamma_{\sigma_B,\tau_B}^{\frac{1}{q}}(Y)\|_{(p,q)} = \|Y\|_{p,q,(\sigma_B,\tau_B)}$, therefore, taking the infimum or supremum for the relevant comparisons between p and q, we arrive at Eq. (2.61). $\square$

This identity also has a specific form for positive operators:

Corollary 2.10. *With* $1 \le q, p \le \infty$, $X \in \mathcal{P}(AB)$ *and* $\sigma_B \in \mathcal{P}^*(B)$,

$$\|X\|_{(p,q),\sigma_B} = \begin{cases} \sup_{\sigma_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A}^{\frac{-1}{p}}(X)\|_{q,\sigma_A \otimes \sigma_B} & \text{if } p \ge q \\ \inf_{\sigma_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A}^{\frac{-1}{p}}(X)\|_{q,\sigma_A \otimes \sigma_B} & \text{if } p \le q \end{cases}.$$

(2.64)

Proof. This is clear by replacing Y in Eq. (2.61) with $X \in \mathcal{P}(AB)$. $\square$

2.2.4 Three-part norms

The framework of Pisier [39] also allows for a natural extension of the two-part norm. We define a three-part norm on tripartite systems in an analogous way: for $Y \in \mathcal{L}(ABC)$

$$\|Y\|_{(p,q,s)} = \begin{cases} \sup_{\sigma_A, \tau_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A,\tau_A}^{\frac{1}{q}-\frac{1}{p}}(Y)\|_{(q,s;AB)} & \text{if } p \ge q \\ \inf_{\sigma_A, \tau_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A,\tau_A}^{\frac{1}{q}-\frac{1}{p}}(Y)\|_{(q,s;AB)} & \text{if } p \le q \end{cases}.$$

(2.65)

In the above we use the notation $\|Y\|_{(p,q;A)}$ to indicate over which parts of the space the optimisation occurs. This is an important distinction as in general, with $Y \in \mathcal{L}(ABC)$ and, for instance, $p \le q$:

$$\|Y\|_{(p,q;A)} = \inf_{\sigma_A, \tau_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A,\tau_A}^{\frac{-1}{p}}(Y)\|_{q,(\sigma_A,\tau_A)}$$

(2.66)

$$\neq$$

$$\|Y\|_{(p,q;AB)} = \inf_{\sigma_{AB}, \tau_{AB} \in \mathcal{D}^*(AB)} \|\Gamma_{\sigma_{AB},\tau_{AB}}^{\frac{-1}{p}}(Y)\|_{q,(\sigma_{AB},\tau_{AB})}.$$

(2.67)

This definition allows for a reduction similar to Eq. (2.8), originally given in [40]:

Proposition 2.11. *For* $X_{ABC} \in \mathcal{P}(ABC)$,

$$\|X_{ABC}\|_{(p,q,1)} = \|X_{AB}\|_{(p,q)}.$$

(2.68)

Proof. By Eq. (2.8), with $\sigma_A \in \mathcal{P}^*(A)$

$$\|\Gamma_{\sigma_A}^{\frac{1}{q}-\frac{1}{p}}(X_{ABC})\|_{(q,1;AB)} = \|\Gamma_{\sigma_A}^{\frac{1}{q}-\frac{1}{p}}(X_{AB})\|_q$$

(2.69)

$$= \|\Gamma_{\sigma_A}^{\frac{-1}{p}}(X_{AB})\|_{q,\sigma_A}.$$

(2.70)

Choosing $p \geq q$ and optimising over $\sigma_A \in \mathcal{D}(A)$ on both sides we obtain

$$\sup_{\sigma_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A}^{\frac{1}{q}-\frac{1}{p}}(X_{ABC})\|_{(q,1;AB)} = \sup_{\sigma_A \in \mathcal{D}^*(A)} \|\Gamma_{\sigma_A}^{\frac{-1}{p}}(X_{AB})\|_{q,\sigma_A}, \tag{2.71}$$

which is equivalent to Eq. (2.68). Choosing $p \leq q$ we instead take the infimum and arrive again at Eq. (2.68). □

2.2.5 Adapting the general formulation

In the remainder of this thesis, we essentially use the above quantities but ignore the optimisations. Overall, the optimisations prove problematic and unnecessary to reproduce the convenient interpolation structure, i.e. we work with expressions of the form:

$$\left\|\Gamma_{\sigma_A}^{\frac{-1}{p}}(X)\right\|_{q,\sigma_A \otimes \sigma_B} \quad \text{and} \quad \left\|\Gamma_{\tau_{AB}}^{\frac{1}{s}-\frac{1}{q}}\left(\Gamma_{\sigma_A}^{\frac{1}{q}-\frac{1}{p}}(X)\right)\right\|_s. \tag{2.72}$$

Evidently, these expressions do not define a unique value determined by p and q (and s) but rather encompass a family of norm-like expressions which are free in the first (and second) parameter.

When we remove the optimisation, the reductions in the previous sections relate closely to situations which produce particular Rényi entropic quantities. Moreover, the choice of parameters for the relevant comparisons almost uncannily reproduces the expected form of the chain and decomposition rules. We may even take advantage of the structure to derive general divergence inequalities which prove indispensable in establishing the improved versions of the uncertainty relations and information exclusion relations covered in Chapter 4.

2.3 Three-line theorem for unbalanced weighted norms

This result and its specifications provide the main mechanism by which we compare divergences of different order. We present an extension of the three-line theorem from [38] to spaces equipped with unbalanced weighted norms.

Theorem 2.12. *We denote $S := \{z \in \mathbb{C} : 0 \leq \Re(z) \leq 1\}$, the complex strip. Let $F : S \to \mathcal{L}(A,B)$ be a bounded map that is holomorphic on the interior of S and continuous on the boundary. Define*

$$\mathsf{M}_k = \sup_{t \in \mathbb{R}} \|F(k + \mathbf{i}\,t)\|_{p_k,(\sigma,\tau)}, \tag{2.73}$$

then for $1 \leq p_\theta, p_0, p_1 \leq \infty$ such that $\frac{1}{p_\theta} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$,

$$\|F(\theta)\|_{p_\theta,(\sigma,\tau)} \leq \mathsf{M}_0^{1-\theta}\mathsf{M}_1^\theta. \tag{2.74}$$

This is not only applicable to norms on spaces of square matrices but also on spaces of non-square matrices – a notion we take advantage of to consider Rényi orders less than one. Basically, since we are in essence dealing with positive operators we can instead consider their decomposition into the product of another operator and its adjoint.

The proof of Theorem 2.12 relies on Hadamard's three-line theorem, which we now include for reference.

Theorem 2.13 (Hadamard's three-line [66]). *Let $f : S \to \mathbb{C}$ be a bounded function that is holomorphic on the interior of S and continuous on its boundary. For $k \in \{0,1\}$, let*

$$\mathsf{M}_k = \sup_{t \in \mathbb{R}} |f(k + \mathbf{i}\,t)|. \tag{2.75}$$

Then for every $\theta \in [0, 1]$ we have

$$|f(\theta)| \leq \mathsf{M}_0^{1-\theta}\mathsf{M}_1^{\theta}. \tag{2.76}$$

The proof of Theorem 2.12 closely follows the structure of the proof of Theorem 2 in [38], with some adjustments to allow for general spaces of linear operators.

Proof of Theorem 2.12. Let $X \in \mathcal{L}(A, B)$ be such that $\|X\|_{p'_\theta,(\sigma,\tau)} = 1$ and $\|F(\theta)\|_{p_\theta,(\sigma,\tau)} = \left|\langle X^\dagger, F(\theta)\rangle_{\sigma,\tau}\right|$. We have

$$\|X\|_{p'_\theta,(\sigma,\tau)} = \left\|\Gamma_{\sigma,\tau}^{\frac{1}{p'_\theta}}(X)\right\|_{p'_\theta} = 1, \tag{2.77}$$

hence the singular value decomposition (see Proposition 1.3) of $\Gamma_{\sigma,\tau}^{\frac{1}{p'_\theta}}(X)$ has the form $\Gamma_{\sigma,\tau}^{\frac{1}{p'_\theta}}(X) = UD^{\frac{1}{p'_\theta}}V$, where $U \in \mathcal{U}(C, B), V \in \mathcal{U}(A, C)$ are isometries and D is a square diagonal matrix in $\mathcal{L}(C)$ whose singular values are real, positive and sum to 1. Define

$$X(z) = \Gamma_{\sigma,\tau}^{-\left(\frac{1-z}{p'_0} + \frac{z}{p'_1}\right)}\left(UD^{\left(\frac{1-z}{p'_0} + \frac{z}{p'_1}\right)}V\right) = \sigma^{-\left(\frac{1-z}{2p'_0} + \frac{z}{2p'_1}\right)}\left(UD^{\left(\frac{1-z}{p'_0} + \frac{z}{p'_1}\right)}V\right)\tau^{-\left(\frac{1-z}{2p'_0} + \frac{z}{2p'_1}\right)}. \tag{2.78}$$

Observe that the map $z \mapsto X(z)$ is holomorphic and $\frac{1}{p'_\theta} = \frac{1-\theta}{p'_0} + \frac{\theta}{p'_1}$, therefore

$$X(\theta) = \Gamma_{\sigma,\tau}^{-\frac{1}{p'_\theta}}\left(UD^{\frac{1}{p'_\theta}}V\right) = \Gamma_{\sigma,\tau}^{-\frac{1}{p'_\theta}}\left(\Gamma_{\sigma,\tau}^{\frac{1}{p'_\theta}}(X)\right) = X. \tag{2.79}$$

Now define $g(z) = \langle X^\dagger(z), F(z)\rangle_{\sigma,\tau} = \mathrm{tr}\left(X(z)\sigma^{\frac{1}{2}}F(z)\tau^{\frac{1}{2}}\right)$. Hence g satisfies the requirements of Hadamard's three-line theorem (Theorem 2.13), and we may write

$$\|F(\theta)\|_{p_\theta,(\sigma,\tau)} = \left|\langle X^\dagger, F(\theta)\rangle_{\sigma,\tau}\right| \tag{2.80}$$

$$= \left|\mathrm{tr}\left(X(\theta)\sigma^{\frac{1}{2}}F(\theta)\tau^{\frac{1}{2}}\right)\right| \tag{2.81}$$

$$= |g(\theta)| \tag{2.82}$$

$$\leq \sup_{t\in\mathbb{R}}|g(\mathbf{i}\,t)|^{1-\theta}\sup_{t\in\mathbb{R}}|g(1+\mathbf{i}\,t)|^{\theta} \tag{2.83}$$

$$= \sup_{t\in\mathbb{R}}\left|\langle X^\dagger(\mathbf{i}\,t), F(\mathbf{i}\,t)\rangle_{\sigma,\tau}\right|^{1-\theta}\sup_{t\in\mathbb{R}}\left|\langle X^\dagger(1+\mathbf{i}\,t), F(1+\mathbf{i}\,t)\rangle_{\sigma,\tau}\right|^{\theta} \tag{2.84}$$

$$\leq \sup_{t\in\mathbb{R}}\left(\|X(\mathbf{i}\,t)\|_{p'_0,(\sigma,\tau)}\|F(\mathbf{i}\,t)\|_{p_0,(\sigma,\tau)}\right)^{1-\theta}\sup_{t\in\mathbb{R}}\left(\|X(1+\mathbf{i}\,t)\|_{p'_1,(\sigma,\tau)}\|F(1+\mathbf{i}\,t)\|_{p_1,(\sigma,\tau)}\right)^{\theta}, \tag{2.85}$$

where in the last line we used the Hölder type inequality, Eq. (2.13).

By definition, $\|X(\mathbf{i}\,t)\|_{p'_0,(\sigma,\tau)} = \left\|\Gamma_{\sigma,\tau}^{\frac{1}{p'_0}}(X(\mathbf{i}\,t))\right\|_{p'_0}$ and

$$\Gamma_{\sigma,\tau}^{\frac{1}{p'_0}}(X(\mathbf{i}\,t)) = \Gamma_{\sigma,\tau}^{\frac{1}{p'_0}}\left(\Gamma_{\sigma,\tau}^{-\left(\frac{1-\mathbf{i}\,t}{p'_0} + \frac{\mathbf{i}\,t}{p'_1}\right)}\left(UD^{\left(\frac{1-\mathbf{i}\,t}{p'_0} + \frac{\mathbf{i}\,t}{p'_1}\right)}V\right)\right) \tag{2.86}$$

$$= \Gamma_{\sigma,\tau}^{-\left(\frac{-\mathbf{i}\,t}{p'_0} + \frac{\mathbf{i}\,t}{p'_1}\right)}\left(UD^{\left(\frac{1-\mathbf{i}\,t}{p'_0} + \frac{\mathbf{i}\,t}{p'_1}\right)}V\right) \tag{2.87}$$

$$= \sigma^{-\left(\frac{-\mathbf{i}\,t}{2p'_0} + \frac{\mathbf{i}\,t}{2p'_1}\right)}UD^{\left(\frac{-\mathbf{i}\,t}{p'_0} + \frac{\mathbf{i}\,t}{p'_1}\right)}D^{\frac{1}{p'_0}}V\tau^{-\left(\frac{-\mathbf{i}\,t}{2p'_0} + \frac{\mathbf{i}\,t}{2p'_1}\right)}. \tag{2.88}$$

Observe $\sigma^{-\left(\frac{-\mathbf{i}\,t}{2p_0'}+\frac{\mathbf{i}\,t}{2p_1'}\right)}UD^{\left(\frac{-\mathbf{i}\,t}{p_0'}+\frac{\mathbf{i}\,t}{p_1'}\right)}$ and $V\tau^{-\left(\frac{-\mathbf{i}\,t}{2p_0'}+\frac{\mathbf{i}\,t}{2p_1'}\right)}$ are isometries, thus $\|X(\mathbf{i}\,t)\|_{p_0',(\sigma,\tau)}=1$. In a similar way we can show $\|X(1+\mathbf{i}\,t)\|_{p_1',(\sigma,\tau)}=1$.

Hence Eq. (2.85) becomes

$$\|F(\theta)\|_{p_\theta,(\sigma,\tau)} \leq \sup_{t\in\mathbb{R}}\left(\|F(\mathbf{i}\,t)\|_{p_0,(\sigma,\tau)}\right)^{1-\theta}\sup_{t\in\mathbb{R}}\left(\|F(1+\mathbf{i}\,t)\|_{p_1,(\sigma,\tau)}\right)^{\theta}. \tag{2.89}$$

$\square$

We defer the introduction of the refinements of this theorem to Chapter 3, where they can be viewed with the results in which they are used.

Chapter 3

Rényi entropy divergence inequalities

This chapter covers Rényi extensions of the chain rule, decomposition rule and more general divergence equivalences in the form of divergence inequalities. The first section details the inequalities and the relationships between them while the technique of interpolation and more detailed proofs are found in the next section.

As it is closely related to these results, we also include a previously published approach to the proof of the decomposition rule in Appendix A. This alternative approach is based on the method employed by Dupuis in his proof of the Rényi entropy chain rules (see Section 1.5.3).

The effective difference between these two version of the decomposition rule are the valid ranges of the parameters. For the previous result in Appendix A, α may go to zero while the other two parameters must be no less than $\frac{1}{2}$. However, for the new result in this chapter, instead γ can go to zero and α is greater than or equal to $\frac{1}{2}$.

3.1 Main results

First, we have a family of general bipartite divergence inequalities. These more versatile comparisons are one of the main tools used to improve the bounds on the uncertainty relations in Chapter 4.

Theorem 3.1. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0^*$ with $\alpha, \beta \geq \frac{1}{2}$ and $\gamma \in \mathbb{R}$. For $\rho_{AB} \in \mathcal{D}(AB), \sigma_A \in \mathcal{P}^*(A)$ and $\tau_B \in \mathcal{P}(B)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then*

$$-H_\alpha\left(\rho_{AB}\,\|\,\tau_B\right) \leq D_\beta\left(\rho_{AB}\,\|\,\sigma_A \otimes \tau_B\right) + \log\left(\operatorname{tr}\rho_A \sigma_A^{\frac{1}{\gamma'}}\right)^{\gamma'}. \tag{3.1}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$-H_\alpha\left(\rho_{AB}\,\|\,\tau_B\right) \geq D_\beta\left(\rho_{AB}\,\|\,\sigma_A \otimes \tau_B\right) + \log\left(\operatorname{tr}\rho_A \sigma_A^{\frac{1}{\gamma'}}\right)^{\gamma'}. \tag{3.2}$$

Where in both cases $\alpha, \beta < 1 \vee \mathbb{1}_A \otimes \tau_B \gg \rho_{AB}$.

The derivation of these inequalities essentially involves applying the machinery of Theorem 2.12 to a choice of function motivated by the relationship between Rényi divergence and Pisier norms, and performing some analysis on the resulting valid ranges for the Rényi parameters.

There are some things to note about this result, especially when it comes to the nature of σ_A and τ_B. In general the only requirement for the second argument of the Rényi divergence is that it be positive semi-definite, but here we restrict $\sigma_A \in \mathcal{P}^*(A)$. This follows from the application of the interpolation which is only valid for strictly positive matrices. However we can extend the argument to $\tau_B \in \mathcal{P}(B)$ by observing the continuity of the Rényi divergence (see Proposition 1.14). This does not directly apply to the term on the far right of the above inequalities which involve σ_A, but this issue is resolved when further optimisation is performed. Indeed, these terms are

---

*This condition is equivalent to $\alpha' = \beta' + \gamma'$ which may provide a more convenient form for certain calculations. For further details see Remark 3.2.

Table 3.1 : Some particular choices satisfying $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$

$\alpha =$ / $\gamma =$	0	$\frac{1}{2}$	1	2	∞
0	$\beta = 0$	$\beta = \frac{1}{2}$	$\beta = 1$	$\beta = 2$	$\beta = \infty$
$\frac{1}{2}$	$\beta = \infty$	$\beta = 0$	$\beta = 1$	$\beta = \frac{3}{2}$	$\beta = 2$
1	$\beta = 1$	$\beta = 1$	$\beta \in \mathbb{R}$	$\beta = 1$	$\beta = 1$
2	$\beta = \frac{2}{3}$	$\beta = \frac{3}{4}$	$\beta = 1$	$\beta = 0$	$\beta = \frac{1}{2}$
∞	$\beta = \frac{1}{2}$	$\beta = \frac{2}{3}$	$\beta = 1$	$\beta = \infty$	$\beta = 0$

the logarithm applied to the Schatten inner product so, by duality, optimising over $\sigma_A \in \mathcal{D}^*(A)$ produces the Schatten γ-norm.

Even though we may consider any $\tau_B \in \mathcal{P}(B)$, we still include the support conditions. Without these conditions the quantities might diverge to $\pm\infty$ – but we can easily verify that the inequalities would be satisfied in any case, however trivially.

It is informative to consider the behaviour of the condition on the Rényi parameters for particular values. We summarise some important cases in the following remark.

Remark 3.2. *The condition $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha, \beta, \gamma \in \mathbb{R}$ can also be written in two other convenient forms:*

$$\frac{\alpha}{\alpha - 1} = \frac{\beta}{\beta - 1} + \frac{\gamma}{\gamma - 1} \quad \text{and} \tag{3.3}$$

$$\frac{1}{1 - \alpha} = \frac{1}{1 - \beta} + \frac{1}{1 - \gamma} - 1 \tag{3.4}$$

Given this condition the following hold:

$$\alpha = 0 \implies \beta = \frac{\gamma}{1 - 2\gamma}, \quad \gamma = 0 \implies \beta = \alpha,$$

$$\alpha = \frac{1}{2} \implies \beta = \frac{1 - 2\gamma}{2 - 3\gamma}, \quad \gamma = \frac{1}{2} \implies \beta = 2 - \frac{1}{\alpha} = \frac{1}{\hat{\alpha}},$$

$$\alpha = 1 \implies 1 \in \{\gamma, \beta\}, \quad \gamma = 1 \implies 1 \in \{\alpha, \beta\},$$

$$\alpha = 2 \implies \beta = 2 - \gamma, \quad \gamma = 2 \implies \beta = \frac{\alpha - 2}{2\alpha - 3},$$

$$\alpha \to \infty \implies \beta = \frac{1}{\gamma}, \quad \gamma \to \infty \implies \beta = \frac{1}{2 - \alpha}.$$

Combining these statements we may find the situations summarises in Table 3.1

We cannot always achieve every one of the above combinations due to the conditions on the parameters imposed in the results of this thesis. However, they reveal some of the patterns and structure that the condition implies. We note in particular that although one can choose all parameters equal to 1 (hence recovering the Shannon and von Neumann equivalences) one is only obliged to choose at least two parameters equal to 1. Of course, this results in a weaker inequality, whose direction is determined by the other parameter. Interestingly, this also is the only case when $\alpha = \beta$ and $\gamma \neq 0$.

We can derive the more familiar bipartite comparisons directly from the Theorem 3.1 by relabelling, optimising and making certain specifications.

Theorem 3.3. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha, \beta \geq \frac{1}{2}$ and $\gamma \geq 0$.*
For $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_A \in \mathcal{P}(A)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then

$$I_\beta(\rho_{AB}\|\tau_A) \geq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{3.5}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$I_\beta(\rho_{AB}\|\tau_A) \leq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{3.6}$$

Where in both cases $\alpha, \beta < 1 \vee \tau_A \otimes \mathbb{1}_B \gg \rho_{AB}$.

Theorem 3.4. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha, \beta \geq \frac{1}{2}$ and $\gamma \geq 0$.*
For $\rho_{AB} \in \mathcal{D}(AB)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then

$$H_\alpha(\rho_{AB}) \geq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.7}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$H_\alpha(\rho_{AB}) \leq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.8}$$

Note the main difference between the two results above is choosing $\tau_A = \mathbb{1}_A$ in the chain rule.

We may combine these results to produce an extension of the other familiar form of the mutual information decomposition rule. This inequality is somewhat weaker in exchange for not involving the conditional entropy. Naturally, we recover equivalence when all parameters tend to 1.

Corollary 3.5. *For $\rho_{AB} \in \mathcal{D}(AB)$ with $\alpha, \beta \geq 0$, $\gamma, \delta \geq \frac{1}{2}$ such that $\frac{\delta - \gamma}{\delta\gamma - 2\gamma + 1} = \frac{2\beta\alpha - \beta - \alpha}{\beta\alpha}$ we have, when $\delta < \alpha, \beta, \gamma$,*

$$I_\beta^\downarrow(A : B)_\rho \geq H_\alpha(\rho_A) + H_\beta(\rho_B) - H_\delta(\rho_{AB}) \tag{3.9}$$

and, when $\delta > \alpha, \beta, \gamma$,

$$I_\beta^\downarrow(A : B)_\rho \leq H_\alpha(\rho_A) + H_\beta(\rho_B) - H_\delta(\rho_{AB}). \tag{3.10}$$

This can be determined by applying a further optimisation and substituting the conditional entropy in Theorem 3.3 with the appropriate comparison in Theorem 3.4. Although not necessary in the main decomposition rules, here we must minimize over τ_A to produce a statement which is compatible with the bipartite chain rules, hence the specification to $I_\beta^\downarrow(A : B)_\rho$.

We now introduce the tripartite chain rules. Although we provide a new approach to their derivation using our interpolation technique, this is not much more than a reproduction of Dupuis' chain rules (see Theorem 1.15), with a slight reworking of the conditions on the parameters.

Theorem 3.6. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha, \gamma \geq \frac{1}{2}$ and $\beta > \frac{1}{2}$.*
For $\rho_{ABC} \in \mathcal{D}(ABC)$ and $\tau_C \in \mathcal{P}(C)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then

$$H_\alpha(\rho_{ABC}\|\tau_C) \geq H_\beta^\uparrow(A|BC)_\rho + H_\gamma(\rho_{BC}\|\tau_C). \tag{3.11}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$H_\alpha(\rho_{ABC}\|\tau_C) \leq H_\beta^\uparrow(A|BC)_\rho + H_\gamma(\rho_{BC}\|\tau_C). \tag{3.12}$$

Where in both cases $\alpha, \gamma < 1 \vee \mathbb{1}_{AB} \otimes \tau_C \gg \rho_{ABC}$.

A significant difference of this tripartite result from the bipartite ones above is the occurrence of non-commuting operators. The resolution of this complication involves introducing purifications and using the resulting dualities to make comparisons on the subsystems with which we are concerned. As a result, this method is closely aligned to Dupuis', but it nonetheless conveys that this type of derivation is contained within a more general structure which warrants closer attention.

From Theorem 3.6 we may also derive a bipartite chain rule which is distinct from Theorem 3.4.

Corollary 3.7. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha \geq 0, \beta > \frac{1}{2}$ and $\gamma \geq \frac{1}{2}$.*
For $\rho_{AB} \in \mathcal{D}(AB)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then

$$H_\alpha(\rho_{AB}) \geq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.13}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$H_\alpha(\rho_{AB}) \leq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.14}$$

This is a direct result of choosing C as a trivial subsystem. Note this has slightly different constraints on the parameters, which may be determined by examining the conditions imposed in the proof of Theorem 3.6.

3.2 Proofs

This section covers the technical proofs of the above results. We begin by with an exploration of the relationship between the Rényi parameters then treat the results in more or less logical order. We include a restatement of each theorem inline for ease of reference.

3.2.1 Applying the three-line theorem

Before moving forward with the proofs of the divergence inequalities, we will first look at some motivation for the choice of parameters.

We want to use the interpolation result to find inequalities of a particular form, for example the decomposition rule:

$$-H_\alpha(\rho_{AB}\|\tau_A)_\rho \leq I_\beta(\rho_{AB}\|\tau_A)_\rho - H_\gamma(\rho_B). \tag{3.15}$$

Exponentiating on both sides and keeping in mind that we can express the resulting quantities as Schatten norms to the power of a function of the relevant index we obtain an inequality of the form

$$\|X_{B|A}\|_{p_\alpha}^{\alpha'} \leq \|X_B\|_{p_\gamma}^{\gamma'} \|X_{A:B}\|_{p_\beta}^{\beta'}, \tag{3.16}$$

where $X_{B|A}$, etc. are simply place-holders for the actual operators, used for brevity. We can then put Eq. (3.16) in the form required for Theorem 2.12 by taking both sides to the power of $\frac{1}{\alpha'}$, resulting in

$$\|X_{B|A}\|_{p_\alpha} \leq \|X_B\|_{p_\gamma}^{\frac{\gamma'}{\alpha'}} \|X_{A:B}\|_{p_\beta}^{\frac{\beta'}{\alpha'}}, \tag{3.17}$$

where $1 - \theta = \frac{\gamma'}{\alpha'}$ and $\theta = \frac{\beta'}{\alpha'}$. This implies

$$1 - \frac{\beta'}{\alpha'} = \frac{\gamma'}{\alpha'} \implies \alpha' = \beta' + \gamma'. \tag{3.18}$$

We can find the reverse of the inequality in Eq. (3.16) by negating all the exponents but this does not affect Eq. (3.18). Additionally, the order of the quantities in Eq. (3.16) has no effect, since we can choose a θ in each case that reproduces Eq. (3.18). For example we could rewrite Eq. (3.16) as

$$\|X_{A:B}\|_{p_\beta}^{-\beta'} \leq \|X_{B|A}\|_{p_\alpha}^{-\alpha'} \|X_B\|_{p_\gamma}^{\gamma'}. \tag{3.19}$$

To apply Theorem 2.12 in this case we would choose $1 - \theta = \frac{\alpha'}{\beta'}$ and $\theta = \frac{-\gamma'}{\beta'}$, resulting in $1 + \frac{\gamma'}{\beta'} = \frac{\alpha'}{\beta'}$, which is again Eq. (3.18).

A more in-depth discussion of the implications and restrictions of this condition, which inform the choices in the following results, is deferred to Appendix B.2.

3.2.2 Bipartite divergence inequalities

We first have the general bipartite inequality:

Theorem 3.1. *Let* $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ *with* $\alpha, \beta \geq \frac{1}{2}$ *and* $\gamma \in \mathbb{R}$. *For* $\rho_{AB} \in \mathcal{D}(AB), \sigma_A \in \mathcal{P}^*(A)$ *and* $\tau_B \in \mathcal{P}(B)$, *if* $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ *then*

$$-H_\alpha\left(\rho_{AB} \,\|\, \tau_B\right) \leq D_\beta\left(\rho_{AB} \,\|\, \sigma_A \otimes \tau_B\right) + \log\left(\operatorname{tr} \rho_A \sigma_A^{\frac{1}{\gamma'}}\right)^{\gamma'}. \tag{3.20}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ *then*

$$-H_\alpha\left(\rho_{AB} \,\|\, \tau_B\right) \geq D_\beta\left(\rho_{AB} \,\|\, \sigma_A \otimes \tau_B\right) + \log\left(\operatorname{tr} \rho_A \sigma_A^{\frac{1}{\gamma'}}\right)^{\gamma'}. \tag{3.21}$$

Where in both cases $\alpha, \beta < 1 \vee \mathbb{1}_A \otimes \tau_B \gg \rho_{AB}$.

To facilitate the demonstration of the above theorem, we introduce a refinement of Theorem 2.12, which establishes a general log-convexity result more closely aligned to our particular context.

Lemma 3.8. *For* $Y \in \mathcal{L}(A, B)$ *and* $\sigma_1, \tau_1 \in \mathcal{P}^*(B), \sigma_2, \tau_2 \in \mathcal{P}^*(A)$ *such that* $[\sigma_1, \tau_1] = [\sigma_2, \tau_2] = 0$,

$$\left\|\Gamma_{\sigma_1,\sigma_2}^{f(\theta)}(Y)\right\|_{q_\theta,(\tau_1,\tau_2)} \leq \left\|\Gamma_{\sigma_1,\sigma_2}^{f(0)}(Y)\right\|_{q_0,(\tau_1,\tau_2)}^{1-\theta} \left\|\Gamma_{\sigma_1,\sigma_2}^{f(1)}(Y)\right\|_{q_1,(\tau_1,\tau_2)}^{\theta}, \tag{3.22}$$

where $f : (0, 1) \to \mathbb{R}$ *is a affine function and* $1 \leq q_\theta, q_0, q_1 \leq \infty$ *are related by*

$$\frac{1}{q_\theta} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}. \tag{3.23}$$

The proof of Lemma 3.8 follows from making some specifications in Theorem 2.12.

Proof. Let $\hat{f}$ be the complex continuation of f on S, i.e if $f(x) = ax + b$ for $x, a, b \in \mathbb{R}$, then $\hat{f}(z) = az + b$. Define $F : S \to \mathcal{L}(A, B)$ such that $F : z \longmapsto \Gamma_{\sigma_1,\sigma_2}^{\hat{f}(z)}(Y)$. Accordingly, $f = \Re(\hat{f})$, hence $F(\theta) = \Gamma_{\sigma_1,\sigma_2}^{f(\theta)}(Y)$.

Additionally, we have

$$\sup_{t\in\mathbb{R}} \|F(\mathbf{i}\,t)\|_{q_0} = \sup_{t\in\mathbb{R}} \left\|\Gamma_{\sigma_1,\sigma_2}^{\hat{f}(\mathbf{i}\,t)}(Y)\right\|_{q_0,(\tau_1,\tau_2)} \tag{3.24}$$

$$= \left\|\Gamma_{\sigma_1,\sigma_2}^{f(0)}(Y)\right\|_{q_0,(\tau_1,\tau_2)}, \tag{3.25}$$

where the second equality is a result of the assumption that $\sigma_1^{\mathbf{i}\,t}$ and $\sigma_2^{\mathbf{i}\,t}$ commute with τ_1 and τ_2 respectively, and that they are unitary for all $t \in \mathbb{R}$. Similarly, we have

$$\sup_{t\in\mathbb{R}} \|F(1+\mathbf{i}\,t)\|_{q_1,(\tau_1,\tau_2)} = \left\|\Gamma_{\sigma_1,\sigma_2}^{f(1)}(Y)\right\|_{q_1,(\tau_1,\tau_2)}. \tag{3.26}$$

With the above conditions we have by Theorem 2.12

$$\left\|\Gamma_{\sigma_1,\sigma_2}^{f(\theta)}(Y)\right\|_{q_\theta,(\tau_1,\tau_2)} \le \left\|\Gamma_{\sigma_1,\sigma_2}^{f(0)}(Y)\right\|_{q_0,(\tau_1,\tau_2)}^{1-\theta} \left\|\Gamma_{\sigma_1,\sigma_2}^{f(1)}(Y)\right\|_{q_1,(\tau_1,\tau_2)}^{\theta}. \tag{3.27}$$

$\square$

With the scaffold provided by this log-convexity result, we need only make some informed choices to produce the general divergence inequalities and the comparisons that follow.

Proof of Theorem 3.1. Noting $\frac{\alpha'}{\alpha} = \alpha' - 1$, we may determine

$$\frac{\alpha'}{\alpha} = \frac{\beta'}{\beta} + \gamma' = \beta' + \frac{\gamma'}{\gamma}. \tag{3.28}$$

Accordingly, in Lemma 3.8 we choose

$$\theta = \frac{\gamma'}{\alpha'}, \quad q_\theta = 2\alpha, \quad q_0 = 2\beta, \quad q_1 = 2, \tag{3.29}$$

$$\sigma_1 = \mathbb{1}_C, \quad \sigma_2 = \sigma_A, \quad \tau_1 = \mathbb{1}_C, \quad \tau_2 = \tau_B^2, \tag{3.30}$$

$$Y = \Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M), \quad f(x) = \frac{x-1}{\beta'} + \frac{x}{\gamma'}. \tag{3.31}$$

Hence for $\alpha' > 0$ we can write Eq. (3.22) as

$$\left\|\Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\alpha'}-\frac{1}{\alpha'}}\left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M)\right)\right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)} \le \left\|\Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{-1}{\beta'}}\left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M)\right)\right\|_{2\beta,(\mathbb{1}_C,\tau_B^2)}^{\frac{\beta'}{\alpha'}} \left\|\Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\gamma'}}\left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M)\right)\right\|_{2,(\mathbb{1}_C,\tau_B^2)}^{\frac{\gamma'}{\alpha'}} \tag{3.32}$$

$$\implies \log\left\|\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M)\right\|_{2\alpha,(\mathbb{1}_C,\tau_B^2)}^{\alpha'} \le \log\left\|\Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{-1}{\beta'}}\left(\Gamma_{\mathbb{1}_C,\tau_B}^{-1}(M)\right)\right\|_{2\beta,(\mathbb{1}_C,\tau_B^2)}^{\beta'} + \log\left\|\Gamma_{\mathbb{1}_C,\sigma_A}^{\frac{1}{\gamma'}}(M)\right\|_{2}^{\gamma'}. \tag{3.33}$$

Considering Lemmata 2.1 and 2.2, we can write the above as

$$-H_\alpha\left(\rho_{AB}\,\|\,\tau_B\right) \le D_\beta\left(\rho_{AB}\,\|\,\sigma_A \otimes \tau_B\right) + \log\left(\operatorname{tr}\rho_A\sigma_A^{\frac{1}{\gamma'}}\right)^{\gamma'}. \tag{3.34}$$

Otherwise, if $\alpha' < 0$, the inequality in Eq. (3.33) is reversed.

If instead $\frac{\alpha'}{\gamma'} \in (0,1)$ we consider the alternate form of Eq. (3.28):

$$\gamma' = \frac{\alpha'}{\alpha} - \frac{\beta'}{\beta}, \quad \frac{\gamma'}{\gamma} = \frac{\alpha'}{\alpha} - \beta. \tag{3.35}$$

Taking similar choices in Lemma 3.8, with $\theta = \frac{\alpha'}{\gamma'}$ and $f(x) = \frac{x-1}{\beta'}$, and employing the same process we may write for $\gamma' > 0$:

$$\left\| \Gamma_{\mathbb{1}_C, \sigma_A}^{\frac{1}{\gamma'}} \left(\Gamma_{\mathbb{1}_C, \tau_B}^{-1}(M) \right) \right\|_{2,(\mathbb{1}_C, \tau_B^2)} \leq \left\| \Gamma_{\mathbb{1}_C, \sigma_A}^{\frac{-1}{\beta'}} \left(\Gamma_{\mathbb{1}_C, \tau_B}^{-1}(M) \right) \right\|_{2\beta,(\mathbb{1}_C, \tau_B^2)}^{\frac{-\beta'}{\gamma'}} \left\| \Gamma_{\mathbb{1}_C, \tau_B}^{-1}(M) \right\|_{2\alpha,(\mathbb{1}_C, \tau_B^2)}^{\frac{\alpha'}{\gamma'}},$$

(3.36)

which implies

$$-H_\alpha \left(\rho_{AB} \| \tau_B \right) \geq D_\beta \left(\rho_{AB} \| \sigma_A \otimes \tau_B \right) + \log \left(\operatorname{tr} \rho_A \sigma_A^{\frac{1}{\gamma'}} \right)^{\gamma'}.$$

(3.37)

We again obtain the reverse of this inequality when $\gamma' < 0$.

Finally, we consider $\theta = \frac{\alpha'}{\beta'} \in (0,1)$ and the form of Eq. (3.28):

$$\frac{\beta'}{\beta} = \frac{\alpha'}{\alpha} - \gamma', \quad \beta = \frac{\alpha'}{\alpha} - \frac{\gamma'}{\gamma},$$

(3.38)

so, with $f(x) = \frac{1-x}{\gamma'}$, for $\beta' > 0$ or $\beta' < 0$ we respectively obtain

$$-H_\alpha \left(\rho_{AB} \| \tau_B \right) \geq D_\beta \left(\rho_{AB} \| \sigma_A \otimes \tau_B \right) + \log \left(\operatorname{tr} \rho_A \sigma_A^{\frac{1}{\gamma'}} \right)^{\gamma'}$$

(3.39)

and its reverse.

We now invoke Proposition 1.14. Consider instead $\tau_B \in \mathcal{P}(B)$ – then we may choose in the above $\tau_B^* \in \mathcal{P}^*(B)$ such that $\tau_B^* = \tau_B + \varepsilon \mathbb{1}_B$. Taking $\varepsilon \to 0^+$, we obtain the same statement for positive semi-definite matrices. Note, without the support conditions $\alpha, \beta < 1 \vee \mathbb{1}_A \otimes \tau_B \gg \rho_{AB}$, the quantities on each side may diverge to $+\infty$.

To obtain the conditions on the Rényi parameters we multiply the equation $\alpha' = \beta' + \gamma'$ by $(\alpha - 1)(\beta - 1)(\gamma - 1)$ and apply the observations in Lemma B.3 and Corollary B.5, i.e.

$$\frac{\alpha}{\alpha - 1} = \frac{\beta}{\beta - 1} + \frac{\gamma}{\gamma - 1}$$

(3.40)

$$\implies \alpha(\beta\gamma - \beta - \gamma + 1) = \beta(\alpha\gamma - \alpha - \gamma + 1) + \gamma(\beta\alpha - \alpha - \beta + 1)$$

(3.41)

$$= \alpha\beta\gamma - \alpha\beta - \alpha\gamma + \alpha = \alpha\beta\gamma - \alpha\beta - \beta\gamma + \beta + \alpha\beta\gamma - \alpha\gamma - \beta\gamma + \gamma$$

(3.42)

$$\implies \alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0.$$

(3.43)

Note that this does not cause any trivial satisfactions since $\alpha' \to \pm\infty$ if and only if at least one of β or γ approaches 1.

Taking into account the valid ranges for the choices of the parameters given in Lemma B.3 we have the statement of the lemma. $\qquad\square$

The decomposition rule and chain rule follow directly:

Theorem 3.3. *Let* $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ *with* $\alpha, \beta \geq \frac{1}{2}$ *and* $\gamma \geq 0$.
For $\rho_{AB} \in \mathcal{D}(AB)$ *and* $\tau_A \in \mathcal{P}(A)$, *if* $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ *then*

$$I_\beta(\rho_{AB} \| \tau_A) \geq H_\gamma(\rho_B) - H_\alpha(\rho_{AB} \| \tau_A).$$

(3.44)

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$I_\beta(\rho_{AB} \| \tau_A) \leq H_\gamma(\rho_B) - H_\alpha(\rho_{AB} \| \tau_A). \tag{3.45}$$

Where in both cases $\alpha, \beta < 1 \vee \tau_A \otimes \mathbb{1}_B \gg \rho_{AB}$.

Proof. We begin with a relabelled version of Eq. (3.20), i.e. for $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$,

$$-H_\alpha(\rho_{AB} \| \tau_A) \leq D_\beta(\rho_{AB} \| \tau_A \otimes \sigma_B) + \log\left(\operatorname{tr} \rho_B \sigma_B^{\frac{1}{\gamma'}}\right)^{\gamma'}. \tag{3.46}$$

Since the left-hand side is independent of σ_B we may choose a σ_B such that the first term is ε-close to its infimum, i.e. we have

$$-H_\alpha(\rho_{AB} \| \tau_A) \leq \inf_{\sigma_B \in \mathcal{D}(B)} [D_\alpha(\rho_{AB} \| \tau_A \otimes \sigma_B)] + \log\left(\operatorname{tr} \rho_B \sigma_B^{\frac{1}{\gamma'}}\right)^{\gamma'} + \varepsilon \tag{3.47}$$

$$-H_\alpha(\rho_{AB} \| \tau_A) \leq \inf_{\sigma_B \in \mathcal{D}(B)} [D_\alpha(\rho_{AB} \| \tau_A \otimes \sigma_B)] + \log \sup_{\sigma_B \in \mathcal{D}(B)} \left[\left(\operatorname{tr} \rho_B \sigma_B^{\frac{1}{\gamma'}}\right)^{\gamma'}\right] + \varepsilon \tag{3.48}$$

$$\implies -H_\alpha(\rho_{AB} \| \tau_A) \leq I_\beta(\rho_{AB} \| \tau_A) - H_\gamma(\rho_B) \tag{3.49}$$

$$\implies I_\beta(\rho_{AB} \| \tau_A) \geq H_\alpha(\rho_{AB} \| \tau_A) - H_\gamma(\rho_B), \tag{3.50}$$

where in the third line we let $\varepsilon \to 0$. Observe that the same process applies to the reverse inequality and that in this context we may take $\gamma \to 0^+$, thus we obtain the theorem. $\square$

Theorem 3.4. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha, \beta \geq \frac{1}{2}$ and $\gamma \geq 0$.*
For $\rho_{AB} \in \mathcal{D}(AB)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then

$$H_\alpha(\rho_{AB}) \geq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.51}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$H_\alpha(\rho_{AB}) \leq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.52}$$

Proof. In Theorem 3.1 we relabel the system and choose $\tau_A = \mathbb{1}_A$ to obtain, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$,

$$-H_\alpha(\rho_{AB}) \leq D_\beta(\rho_{AB} \| \mathbb{1}_A \otimes \sigma_B) + \log\left(\operatorname{tr} \rho_B \sigma_B^{\frac{1}{\gamma'}}\right)^{\gamma'}. \tag{3.53}$$

Optimising over σ_B in the same manner as in Eqs. (3.47)-(3.48) the above becomes

$$-H_\alpha(\rho_{AB}) \leq -H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B) \tag{3.54}$$

$$\implies H_\alpha(\rho_{AB}) \geq H_\beta^\uparrow(A|B)_\rho - H_\gamma(\rho_B). \tag{3.55}$$

Similarly, for $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ we have the reverse direction. $\square$

We now include the proof of Corollary 3.5, essentially combining the two above results.

Corollary 3.5. *For $\rho_{AB} \in \mathcal{D}(AB)$ with $\alpha, \beta \geq 0$, $\gamma, \delta \geq \frac{1}{2}$ such that $\frac{\delta - \gamma}{\delta\gamma - 2\gamma + 1} = \frac{2\beta\alpha - \beta - \alpha}{\beta\alpha}$ we have, when $\delta < \alpha, \beta, \gamma$,*

$$I_\beta^\downarrow(A:B)_\rho \geq H_\alpha(\rho_A) + H_\beta(\rho_B) - H_\delta(\rho_{AB}) \tag{3.56}$$

and, when $\delta > \alpha, \beta, \gamma$,

$$I_\beta^\downarrow(A:B)_\rho \leq H_\alpha(\rho_A) + H_\beta(\rho_B) - H_\delta(\rho_{AB}). \tag{3.57}$$

Proof. We begin with a relabelled version of Eq. (3.44), optimising over τ_B: For $\tilde{\alpha}\beta\alpha - 2\beta\alpha - \tilde{\alpha} + \beta + \alpha = 0$ with $\tilde{\alpha}, \beta \geq \frac{1}{2}$, $\alpha \geq 0$ and $\frac{1}{\beta} + \frac{1}{\alpha} \leq 2$ we have

$$\inf_{\tau_B \in \mathcal{D}(B)} I_\beta(\rho_{AB} \| \tau_B) \geq H_\alpha(\rho_A) - \sup_{\tau_B \in \mathcal{D}(B)} H_{\tilde{\alpha}}(\rho_{AB} \| \tau_B) \tag{3.58}$$

$$I_\beta^\downarrow(A:B) \geq H_\alpha(\rho_A) - H_{\tilde{\alpha}}^\uparrow(A|B). \tag{3.59}$$

By Theorem 3.4, for for $\delta\tilde{\alpha}\gamma - 2\tilde{\alpha}\gamma - \delta + \tilde{\alpha} + \gamma = 0$ with $\delta, \tilde{\alpha} \geq \frac{1}{2}$, $\gamma \geq 0$ and $\frac{1}{\tilde{\alpha}} + \frac{1}{\gamma} \leq 2$ we can substitute

$$I_\beta^\downarrow(A:B) \geq H_\alpha(\rho_A) + H_\gamma(\rho_B) - H_\delta(\rho_{AB}). \tag{3.60}$$

where $\frac{\delta - \gamma}{\delta\gamma - 2\gamma + 1} = \frac{2\beta\alpha - \beta - \alpha}{\beta\alpha}$.

We know from Corollary B.5 that both

$$\frac{1}{\beta} + \frac{1}{\alpha} \leq 2 \quad \text{then} \quad \tilde{\alpha} < \alpha, \beta \quad \text{and} \tag{3.61}$$

$$\frac{1}{\tilde{\alpha}} + \frac{1}{\gamma} \leq 2 \quad \text{then} \quad \delta < \tilde{\alpha}, \gamma. \tag{3.62}$$

Similarly, if we begin with Eq. (3.45) and substitute in Eq. (3.52), we arrive at Eq. (3.57) but with

$$\frac{1}{\beta} + \frac{1}{\alpha} \geq 2 \quad \text{then} \quad \tilde{\alpha} > \alpha, \beta \quad \text{and} \tag{3.63}$$

$$\frac{1}{\tilde{\alpha}} + \frac{1}{\gamma} \geq 2 \quad \text{then} \quad \delta > \tilde{\alpha}, \gamma. \tag{3.64}$$

$\square$

3.2.3 The tripartite chain rules

We now include the novel approach to the proof of the tripartite chain rule and the bipartite version which follows.

Theorem 3.6. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha, \gamma \geq \frac{1}{2}$ and $\beta > \frac{1}{2}$.*
For $\rho_{ABC} \in \mathcal{D}(ABC)$ and $\tau_C \in \mathcal{P}(C)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then

$$H_\alpha(\rho_{ABC} \| \tau_C) \geq H_\beta^\uparrow(A|BC)_\rho + H_\gamma(\rho_{BC} \| \tau_C). \tag{3.65}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$H_\alpha(\rho_{ABC} \| \tau_C) \leq H_\beta^\uparrow(A|BC)_\rho + H_\gamma(\rho_{BC} \| \tau_C). \tag{3.66}$$

Where in both cases $\alpha, \gamma < 1 \vee \mathbb{1}_{AB} \otimes \tau_C \gg \rho_{ABC}$.

To demonstrate the above result we first derive a specific form of Theorem 2.12:

Lemma 3.9. *We define the complex strip $S := \{z \in \mathbb{C} : 0 \leq \Re(z) \leq 1\}$. Let $F : S \to \mathcal{L}(A, B)$ be*

a function of the form

$$F(z) = \Gamma_{\sigma_B,\mathbb{1}_A}^{(1-z)\left(\frac{1}{s_0}-\frac{1}{q_0}\right)+z\left(\frac{1}{s_1}-\frac{1}{q_1}\right)}\left(\Gamma_{\mathbb{1}_B,\tau_A}^{(1-z)\left(\frac{1}{q_0}-\frac{1}{p_0}\right)+z\left(\frac{1}{q_1}-\frac{1}{p_1}\right)}(M)\right), \qquad (3.67)$$

where $\sigma_B \in \mathcal{P}^*(B), \tau_A \in \mathcal{P}^*(A)$ and $M \in \mathcal{L}(A,B)$.

   Denote

$$\mathsf{M}_k = \sup_{t \in \mathbb{R}} \|F(k + \mathbf{i}t)\|_{2s_k}. \qquad (3.68)$$

Let $\alpha' = \beta' + \gamma'$ and $\frac{\gamma'}{\alpha'} \in (0,1)$. Then, given $p_\theta, p_1, q_0, q_1 \in \mathbb{R}^+$, $s_\theta, s_0, s_1 \in [1/2, \infty]$ such that

$$\frac{\alpha'}{p_\theta} = \frac{\beta'}{p_0} + \frac{\gamma'}{p_1}, \quad \frac{\alpha'}{q_\theta} = \frac{\beta'}{q_0} + \frac{\gamma'}{q_1}, \quad \frac{\alpha'}{s_\theta} = \frac{\beta'}{s_0} + \frac{\gamma'}{s_1}, \qquad (3.69)$$

we have

$$\left\|\Gamma_{\sigma_B,\mathbb{1}_A}^{\frac{1}{s_\theta}-\frac{1}{q_\theta}}\left(\Gamma_{\mathbb{1}_B,\tau_A}^{\frac{1}{q_\theta}-\frac{1}{p_\theta}}(M)\right)\right\|_{2s_\theta}^{\alpha'} \begin{cases} \le \mathsf{M}_0^{\beta'}\mathsf{M}_1^{\gamma'} & \text{if } \alpha' > 0 \\ \ge \mathsf{M}_0^{\beta'}\mathsf{M}_1^{\gamma'} & \text{if } \alpha' < 0 \end{cases}. \qquad (3.70)$$

Proof. We let $\theta = \frac{\gamma'}{\alpha'} \implies 1-\theta = \frac{\beta'}{\alpha'}$ then by Theorem 2.12 we have $\|F(\theta)\|_{2s_\theta} \le \mathsf{M}_0^{\frac{\beta'}{\alpha'}}\mathsf{M}_1^{\frac{\gamma'}{\alpha'}}$. By evaluating $F(\theta)$ and exponentiating by α' on both sides we arrive at Eq. (3.70). $\qquad \square$

Proof of Theorem 3.6. Noting $\frac{\alpha'}{\alpha} = \alpha' - 1$ and $\frac{\alpha'}{\hat{\alpha}} = \alpha' + 1$ We may determine

$$\frac{\alpha'}{\alpha} = \beta' + \frac{\gamma'}{\gamma} \quad \text{and} \quad \alpha' = \frac{\beta'}{\hat{\beta}} + \frac{\gamma'}{\gamma}. \qquad (3.71)$$

Let $M \in \mathcal{L}(BC, AD)$, $\sigma_D \in \mathcal{P}^*(D)$ and $\tau_C \in \mathcal{P}^*(C)$. In Lemma 3.9 choose $p_\theta, p_0, p_1, q_0, s_\theta = 1$, $q_\theta = \alpha$, $q_1, s_1 = \gamma$ and $s_0 = \hat{\beta}$. Then for $\theta = \frac{\gamma'}{\alpha'}$ and $\alpha' > 0$ we can write

$$\left\|\Gamma_{\sigma_D,\mathbb{1}_{BC}}^{\frac{1}{\alpha'}}\left(\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-1}{\alpha'}}(M)\right)\right\|_2^{\alpha'} \le \sup_{t \in \mathbb{R}}\left\|\Gamma_{\sigma_D,\mathbb{1}_{BC}}^{\frac{1-\mathbf{i}t}{\beta'}}\left(\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-\mathbf{i}t}{\gamma'}}(M)\right)\right\|_{2\hat{\beta}}^{\beta'}\sup_{t \in \mathbb{R}}\left\|\Gamma_{\sigma_D,\mathbb{1}_{BC}}^{\frac{-\mathbf{i}t}{\beta'}}\left(\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-(1+\mathbf{i}t)}{\gamma'}}(M)\right)\right\|_{2\gamma}^{\gamma'}. \qquad (3.72)$$

Let $|\varphi\rangle \in \mathcal{H}_{ABCD}$ be a pure state with Schmidt decompositions (see Proposition 1.4):

$$|\varphi\rangle = \sum_i r_i |i\rangle_{BC} \otimes |i\rangle_{AD} = \sum_i s_i |i\rangle_{ABC} \otimes |i\rangle_D \qquad (3.73)$$

and $\rho_{ABCD} = |\varphi\rangle\langle\varphi|$.

   We choose $M = \sum_i r_i |i\rangle_{AD}\langle i|_{BC}$, so by Proposition 1.10 the above becomes

$$\left\|\Gamma_{\sigma_D,\mathbb{1}_{ABC}}^{\frac{1}{\alpha'}}\left(\Gamma_{\mathbb{1}_D,\tau_C}^{\frac{-1}{\alpha'}}\left(\sum_i s_i |i\rangle_D\langle i|_{ABC}\right)\right)\right\|_2^{\alpha'} \le \left\|\Gamma_{\sigma_D,\mathbb{1}_{BC}}^{\frac{1}{\beta'}}(M)\right\|_{2\hat{\beta}}^{\beta'}\left\|\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-1}{\gamma'}}(M)\right\|_{2\gamma}^{\gamma'}, \qquad (3.74)$$

where we have also used the fact that $\sigma_D^{\mathbf{i}t}$ and $\tau_C^{\mathbf{i}t}$ are unitary for all $t \in \mathbb{R}$.

   Taking the supremum over $\sigma_D \in \mathcal{D}^*(D)$ on both sides we obtain via Lemma B.2:

$$\implies \left\|\Gamma_{\mathbb{1}_D,\tau_C}^{\frac{-1}{\alpha'}}\left(\sum_i s_i |i\rangle_D\langle i|_{ABC}\right)\right\|_{2\alpha}^{\alpha'} \le \sup_{\sigma_D \in \mathcal{D}^*(D)}\left\|\Gamma_{\sigma_D,\mathbb{1}_{BC}}^{\frac{1}{\beta'}}(M)\right\|_{2\hat{\beta}}^{\beta'}\left\|\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-1}{\gamma'}}(M)\right\|_{2\gamma}^{\gamma'}. \qquad (3.75)$$

Further taking the logarithm of both sides and noting the duality of the conditional entropy (see

Proposition 1.11) we have by Lemma 2.1

$$-H_\alpha(\rho_{ABC}\|\tau_C) \leq H_{\hat{\beta}}^\uparrow(A|D)_\rho - H_\gamma(\rho_{BC}\|\tau_C) \tag{3.76}$$

$$\implies H_\alpha(\rho_{ABC}\|\tau_C) \geq H_{\hat{\beta}}^\uparrow(A|BC)_\rho + H_\gamma(\rho_{BC}\|\tau_C). \tag{3.77}$$

If instead $\alpha' < 0$, the inequality in Eq. (3.72) is reversed and we may use a similar optimisation to arrive at

$$H_\alpha(\rho_{ABC}\|\tau_C) \leq H_{\hat{\beta}}^\uparrow(A|BC)_\rho + H(\rho_{BC}\|\tau_C). \tag{3.78}$$

Note that we may rewrite Eq. (3.71) in the forms

$$\frac{\gamma'}{\gamma} = \frac{\alpha'}{\alpha} - \beta' \quad \text{and} \quad \frac{\gamma'}{\gamma} = \alpha' - \frac{\beta'}{\hat{\beta}}. \tag{3.79}$$

Accordingly, if we consider instead $\frac{\alpha'}{\gamma'} \in (0,1)$ and $\gamma' > 0$ we may make the following choices in Lemma 3.9: $p_\theta, p_0, p_1, q_0, s_1 = 1$, $q_\theta, s_\theta = \gamma$, $q_1 = \alpha$ and $s_0 = \hat{\beta}$. This yields

$$\left\|\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-1}{\gamma'}}(M)\right\|_{2\gamma}^{\gamma'} \leq \sup_{t\in\mathbb{R}}\left\|\Gamma_{\sigma_D,\mathbb{1}_{BC}}^{\frac{1-\mathrm{i}t}{\beta'}+\frac{\mathrm{i}t}{\alpha'}}\left(\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-\mathrm{i}t}{\alpha'}}(M)\right)\right\|_{2\hat{\beta}}^{-\beta'} \sup_{t\in\mathbb{R}}\left\|\Gamma_{\sigma_D,\mathbb{1}_{BC}}^{\frac{-\mathrm{i}t}{\beta'}+\frac{1+\mathrm{i}t}{\alpha'}}\left(\Gamma_{\mathbb{1}_{AD},\tau_C}^{\frac{-(1+\mathrm{i}t)}{\alpha'}}(M)\right)\right\|_2^{\alpha'}, \tag{3.80}$$

which, after rearranging, choosing the correct Schmidt decompositions, optimising and further taking the logarithm to then employ the identities in Lemma 2.1 we obtain

$$H_\alpha(\rho_{ABC}\|\tau_C) \leq H_{\hat{\beta}}^\uparrow(A|BC)_\rho + H(\rho_{BC}\|\tau_C). \tag{3.81}$$

Similarly, this inequality is reversed for $\gamma' < 0$.

Finally, we consider $\frac{\alpha'}{\beta'} \in (0,1)$ with $\beta' > 0$ or $\beta' < 0$. Choosing $p_\theta, p_0, p_1, q_\theta, s_1 = 1$, $q_0, s_0 = \gamma$, $q_1 = \alpha$ and $s_\theta = \hat{\beta}$ we again derive, respectively,

$$H_\alpha(\rho_{ABC}\|\tau_C) \leq H_{\hat{\beta}}^\uparrow(A|BC)_\rho + H(\rho_{BC}\|\tau_C) \tag{3.82}$$

and its reverse.

The conditions on the Rényi parameters can be derived in the same way as for Theorem 3.1 and the extension to positive semi-definite matrices follows from Proposition 1.14. $\qquad\square$

We can specialise this to result to a bipartite setting.

Corollary 3.7. *Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha \geq 0, \beta > \frac{1}{2}$ and $\gamma \geq \frac{1}{2}$.*
For $\rho_{AB} \in \mathcal{D}(AB)$, if $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$ then

$$H_\alpha(\rho_{AB}) \geq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.83}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$ then

$$H_\alpha(\rho_{AB}) \leq H_\beta^\uparrow(A|B)_\rho + H_\gamma(\rho_B). \tag{3.84}$$

Proof. We simply choose C to be trivial in Theorem 3.6 and examine the valid ranges in its proof. $\qquad\square$

Chapter 4

Generalised Rényi divergence uncertainty relations

We now arrive at the applications which motivate the inequalities of the previous section. These take the form of some refinements, extensions and improvements of the relations given in Section 1.6. We collect the Maassen-Uffink-like, bipartite uncertainty relations in the first section of this chapter and the information exclusion relations in the second. Again, the detailed proofs of these and their related results are deferred to the final section.

4.1 Bipartite conditional entropy relations

Here we detail generalised Rényi bipartite uncertainty relations, starting with a slightly more general version of Theorem 1.24.

For a summary of the formalism used for measurements and measured states, see Section 1.3.4.

Theorem 4.1. *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)^*$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

For $\alpha, \gamma \geq \frac{1}{2}$ and $\beta > \frac{1}{2}$ such that $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ and $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$, then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \geq H_\alpha(\rho_{AB}\|\tau_B) + q_{\mathrm{MU}}. \tag{4.1}$$

We now give Rényi extensions of the improved uncertainty relations in [44]. These results also constitute versions of the above theorem with an improved bound which is Rényi order dependent. We first have the state-dependent version:

Theorem 4.2. *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

For $\alpha, \gamma \geq \frac{1}{2}, \beta > \frac{1}{2}, \delta \in \mathbb{R}$ if there exists a $\mu \geq \frac{1}{2}$ such that

$$\frac{\alpha - \gamma}{\alpha\gamma - 2\gamma + 1} = \frac{\beta - \delta}{\beta\delta - 2\delta + 1} = \mu \ and \ \frac{1}{\delta} \leq 2 - \frac{1}{\mu} \leq \frac{1}{\gamma}, \tag{4.2}$$

then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta(\rho, \mathbb{X}, \mathbb{Z}) \tag{4.3}$$

and

$$H_\gamma(\mathcal{M}_X(\rho_{AB})\|\tau_B) + H_\beta^\uparrow(Z|B)_\rho \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta(\rho, \mathbb{Z}, \mathbb{X}), \tag{4.4}$$

where $q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = -\log\left(\mathrm{tr}\,\rho_X \sum_x \max_z (c_{x,z})^{\frac{1}{\delta'}} |x\rangle\langle x|\right)^{\delta'}$ and $c_{x,z} = |\langle x|z\rangle|^2$.

Moreover, if $\gamma > \frac{1}{2}$ and there exists a $\tilde{\mu} \geq \frac{1}{2}$ such that

$$\frac{\alpha - \beta}{\alpha\beta - 2\beta + 1} = \frac{\gamma - \delta}{\gamma\delta - 2\delta + 1} = \tilde{\mu} \ and \ \frac{1}{\delta} \leq 2 - \frac{1}{\tilde{\mu}} \leq \frac{1}{\beta}, \tag{4.5}$$

---

*Note we take advantage of the notation $\mathcal{M}_X(\rho_{AB}) = \rho_{XB}$ for brevity and consistency in quantities such as $H_\beta^\uparrow(X|B)_{\mathcal{M}_X(\rho_{AB})} = H_\beta^\uparrow(X|B)_\rho$

then

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma^\uparrow(Z|B)_\rho \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta(\rho), \tag{4.6}$$

where $q_\delta(\rho) = \max\{q_\delta(\rho, \mathbb{X}, \mathbb{Z}), q_\delta(\rho, \mathbb{Z}, \mathbb{X})\}$.

The derivation of this result could be considered an amalgamation of the structures of the proofs of both Theorems 4.1 and 1.19, possible due to the general comparisons of Rényi divergences available from Theorem 3.1.

We may also establish a weaker, state-independent version by considering the 'worst case' state which would achieve the minimum of the bound.

Theorem 4.3. *With the same conditions required for Eq. (4.6), we have*

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma^\uparrow(Z|B)_\rho \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta, \tag{4.7}$$

where

$$q_\delta = \min_{\sigma \in \mathcal{D}(AB)} q_\delta(\sigma) = -\min_{0 \leq p \leq 1} \log \lambda_{\max}\left[\Delta_\delta(p)^{\delta'}\right] \tag{4.8}$$

with $\Delta_\delta(p) = p\sum_x(\max_z c_{x,z})^{\frac{1}{\delta'}}|x\rangle\langle x| + (1-p)\sum_z(\max_x c_{x,z})^{\frac{1}{\delta'}}|z\rangle\langle z|$.

It is relatively straight-forward to show that these results generalise those given in Section 1.6.3. Indeed, when all Rényi parameters go to 1 we recover the relations of Coles and Piani. Otherwise, when $\delta \to 0$ we recover Theorem 4.1.

The relationships between the δ dependent bounds are summarised in the following proposition:

Proposition 4.4. *With $q_\delta(\rho, \mathbb{X}, \mathbb{Z})$ defined as in Theorem 4.2 we have*

$$\lim_{\delta \to 1} q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = q(\rho, \mathbb{X}, \mathbb{Z}), \tag{4.9}$$

$$\lim_{\delta \to 0} q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = q_{\mathrm{MU}}. \tag{4.10}$$

It follows that

$$\lim_{\delta \to 1} q_\delta(\rho) = q(\rho), \tag{4.11}$$

$$\lim_{\delta \to 1} q_\delta = q_{\mathrm{CP}}. \tag{4.12}$$

The proof is mainly an application of l'Hôpital's rule in the same vein as Proposition B.1.

4.2   Information exclusion relations

We may now adapt the above results to derive a Rényi extension of the Hall relation, Theorem 1.17, and further generalise to the improved bounds of Coles and Piani.

Theorem 4.5. *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

For $\alpha, \gamma \geq \frac{1}{2}$ and $\beta > \frac{1}{2}$, satisfying both $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ and $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$, then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$I_\beta^\downarrow(X : B)_\rho + I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq r_{\mathrm{H}} - H_\alpha(\rho_{AB}\|\tau_B). \tag{4.13}$$

The technique employed in the demonstration of this result is almost identical to the technique used for Theorem 1.17. In this case however, the relevant comparisons are possible due to the decomposition and chain rules found in Chapter 3.

Choosing $\alpha \to \infty$, we have the following corollary which summarises the possible choices of parameters which produce an optimal inequality.

Corollary 4.6. *Given the same conditions as Theorem 4.5, for $\frac{1}{2} < \alpha < 2$, we have*

$$I_\alpha^\downarrow(X:B)_\rho + I_{\frac{1}{\alpha}}^\uparrow(B\,;Z)_\rho \leq r_\mathrm{H} - H_\mathrm{min}(A|B)_\rho. \tag{4.14}$$

This brings us to our final main result, the Rényi generalisation of Theorem 1.18. Note that, compared to the relations in the previous section, the bounds here are not order-dependent nor state-dependent, rather we find that the bounds coincide with those of the Shannon and von Neumann situations.

Theorem 4.7 (Improved Rényi information exclusion relations). *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

Given $\alpha, \gamma \geq \frac{1}{2}$, $\beta \in \left[\frac{1}{2}, 2\right]$ with

$$\frac{\alpha - \gamma}{\alpha\gamma - 2\gamma + 1} = \frac{1}{2 - \beta} \ \ and \ \beta\gamma \leq 1, \tag{4.15}$$

then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$I_\beta^\downarrow(X:B)_\rho + I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq r(\mathbb{X}, \mathbb{Z}) - H_\alpha(\rho_{AB}\|\tau_B) \tag{4.16}$$

and

$$I_\gamma(\mathcal{M}_X(\rho_{AB})\|\tau_B) + I_\beta^\downarrow(Z:B)_\rho \leq r(\mathbb{Z}, \mathbb{X}) - H_\alpha(\rho_{AB}\|\tau_B). \tag{4.17}$$

Moreover, if $\gamma \leq 2 - \beta$ and

$$\frac{2\alpha - \alpha\gamma - 1}{\alpha - \gamma} = \frac{1}{2 - \beta}, \quad \frac{2\alpha - \alpha\beta - 1}{\alpha - \beta} = \frac{1}{2 - \gamma} \tag{4.18}$$

then

$$I_\beta^\downarrow(X:B)_\rho + I_\gamma^\downarrow(Z:B)_\rho \leq r_\mathrm{CP} - H_\alpha^\uparrow(A|B)_\rho. \tag{4.19}$$

The proof follows the structure of the derivation of the von Neumann result in Theorem 1.18, now achievable with a reduced version of the general comparison in Theorem 3.1.

Similarly, we find an optimal version of the above by choosing $\alpha \to \infty$.

Corollary 4.8. *Given $\frac{1}{2} \leq \alpha \leq \frac{3}{2}$, then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$*

$$I_\alpha^\downarrow(X:B)_\rho + I_{2-\alpha}(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq r(\mathbb{X}, \mathbb{Z}) - H_\mathrm{min}(A|B)_\rho. \tag{4.20}$$

Specifically,

$$I_{\frac{1}{2}}^\downarrow(X:B)_\rho + I_{\frac{3}{2}}^\downarrow(Z:B)_\rho \leq r_\mathrm{CP} - H_\mathrm{min}(A|B)_\rho. \tag{4.21}$$

4.3 Proofs

We now include the detailed proofs of the above results.

4.3.1 Proofs of the bipartite conditional entropy relations

We first demonstrate the generalised form of the bipartite Rényi uncertainty relation.

Theorem 4.1. *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

For $\alpha, \gamma \geq \frac{1}{2}$ and $\beta > \frac{1}{2}$ such that $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ and $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$, then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \geq H_\alpha(\rho_{AB}\|\tau_B) + q_{\mathrm{MU}}. \tag{4.22}$$

Before we treat the proof of Theorem 4.1 we first introduce a specific form of the Stinespring dilation [67].

Definition 4.9 (Stinespring dilation). *A map $\mathcal{M} \in \mathrm{CPTP}(A, B)$ if and only if there exists an isometry $U \in \mathcal{L}(A, BC)$ such that $\mathcal{M}(\rho) = \mathrm{tr}_C(U\rho U^\dagger)$ for all $\rho \in \mathcal{D}(A)$.*

Proof of Theorem 4.1. Let $\mathcal{S}_Z \in \mathrm{CPTP}(A, ZZ')$ be a Stinespring dilation of $\mathcal{M}_Z$ such that

$$\mathcal{S}_Z(\rho_A) = \sum_{z, z'} \langle z| \rho_A |z'\rangle |z\rangle\langle z'| \otimes |z\rangle\langle z'|. \tag{4.23}$$

We use the same argument as the proof of [42, Theorem 7.6], to arrive at

$$H_\beta^\uparrow(X|B)_\rho \geq H_\beta^\uparrow(Z|Z'B)_{\mathcal{S}_Z(\rho)} + q_{\mathrm{MU}}. \tag{4.24}$$

The two main components of this argument are the comparisons: for $\beta \geq \frac{1}{2}$,

$$H_\beta^\uparrow(Z|Z'B)_{\mathcal{S}_Z(\rho)} \leq -\inf_{\sigma_{Z'B} \in \mathcal{D}(Z'B)} D_\beta\left(\mathcal{M}_X(\rho_{AB})\|\mathcal{M}_X\left(\mathcal{S}_Z^{-1}(\mathbb{1}_Z \otimes \sigma_{Z'B})\right)\right) \quad \text{and} \quad (4.25)$$

$$\mathcal{M}_X\left(\mathcal{S}_Z^{-1}(\mathbb{1}_Z \otimes \sigma_{Z'B})\right) = \sum_{x, z} |\langle x|z\rangle|^2 |x\rangle\langle x| \otimes \langle z'|\sigma_{Z'B}|z'\rangle \leq c\,\mathbb{1}_X \otimes \sigma_B. \tag{4.26}$$

The first comparison is a result of the data-processing inequality (see Proposition 1.13) and for the second we maximise $|\langle x|z\rangle|$ over x and z.

Substituting Eq. (4.26) into Eq. (4.25) yields Eq. (4.24).

Let $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ with $\alpha, \gamma \geq \frac{1}{2}$, $\beta > \frac{1}{2}$ and $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$. Then by Theorem 3.6 we can write for all $\tau_B \in \mathcal{D}(B)$,

$$H_\beta^\uparrow(Z|Z'B)_\rho \geq H_\alpha(\mathcal{S}_Z(\rho_{AB})\|\tau_B) - H_\gamma(\mathrm{tr}_Z(\mathcal{S}_Z(\rho_{AB}))\|\tau_B). \tag{4.27}$$

Substituting Eq. (4.27) into Eq. (4.24) we have

$$H_\beta^\uparrow(X|B)_\rho \geq H_\alpha(\mathcal{S}_Z(\rho_{AB})\|\tau_B) - H_\gamma(\mathrm{tr}_Z(\mathcal{S}_Z(\rho_{AB}))\|\tau_B) + q_{\mathrm{MU}}. \tag{4.28}$$

Using the fact that the marginals on ZB and $Z'B$ of the state $\mathcal{S}_Z(\rho_{AB})$ are equivalent and that the conditional entropies are invariant under local isometries we obtain Eq. (4.22). $\qquad\square$

The following proof Theorem 4.2 has the same broad strokes as the previous proof, the main difference being a tighter comparison when applying the measurement in $\mathbb{X}$.

Theorem 4.2. *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

For $\alpha, \gamma \geq \frac{1}{2}, \beta > \frac{1}{2}, \delta \in \mathbb{R}$ if there exists a $\mu \geq \frac{1}{2}$ such that

$$\frac{\alpha - \gamma}{\alpha\gamma - 2\gamma + 1} = \frac{\beta - \delta}{\beta\delta - 2\delta + 1} = \mu \ and \ \frac{1}{\delta} \leq 2 - \frac{1}{\mu} \leq \frac{1}{\gamma}, \tag{4.29}$$

then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta(\rho, \mathbb{X}, \mathbb{Z}) \tag{4.30}$$

and

$$H_\gamma(\mathcal{M}_X(\rho_{AB})\|\tau_B) + H_\beta^\uparrow(Z|B)_\rho \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta(\rho, \mathbb{Z}, \mathbb{X}), \tag{4.31}$$

where $q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = -\log \left(\operatorname{tr} \rho_X \sum_x \max_z (c_{x,z})^{\frac{1}{\delta'}} |x\rangle\langle x| \right)^{\delta'}$ and $c_{x,z} = |\langle x|z\rangle|^2$.
Moreover, if $\gamma > \frac{1}{2}$ and there exists a $\tilde{\mu} \geq \frac{1}{2}$ such that

$$\frac{\alpha - \beta}{\alpha\beta - 2\beta + 1} = \frac{\gamma - \delta}{\gamma\delta - 2\delta + 1} = \tilde{\mu} \ and \ \frac{1}{\delta} \leq 2 - \frac{1}{\tilde{\mu}} \leq \frac{1}{\beta}, \tag{4.32}$$

then

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma^\uparrow(Z|B)_\rho \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta(\rho), \tag{4.33}$$

where $q_\delta(\rho) = \max\{q_\delta(\rho, \mathbb{X}, \mathbb{Z}), q_\delta(\rho, \mathbb{Z}, \mathbb{X})\}$.

We use a similar argument to the proof of Theorem 4.1, but following the structure given in the proof of [44, Theorem 2].

Proof. Let $\mathcal{S}_Z \in \mathrm{CPTP}(A, ZZ')$ be a Stinespring dilation of $\mathcal{M}_Z$ such that

$$\mathcal{S}_Z(\rho_A) = \sum_{z,z'} \langle z| \rho_A |z'\rangle |z\rangle\langle z'| \otimes |z\rangle\langle z'|. \tag{4.34}$$

By the data-processing inequality:

$$H_\mu^\uparrow(Z|Z'B)_{\mathcal{S}_Z(\rho)} \leq - \inf_{\sigma_{Z'B} \in \mathcal{D}(Z'B)} D_\mu(\rho_{AB}\|\mathcal{S}_Z^{-1}(\mathbb{1}_Z \otimes \sigma_{Z'B})) \tag{4.35}$$

$$\leq - \inf_{\sigma_{Z'B} \in \mathcal{D}(Z'B)} D_\mu \left(\mathcal{M}_X(\rho_{AB})\|\mathcal{M}_X \left(\mathcal{S}_Z^{-1}(\mathbb{1}_Z \otimes \sigma_{Z'B}) \right) \right). \tag{4.36}$$

We may compare

$$\mathcal{M}_X \left(\mathcal{S}_Z^{-1}(\mathbb{1}_Z \otimes \sigma_{Z'B}) \right) = \sum_{x,z} |\langle x|z\rangle|^2 |x\rangle\langle x| \otimes \langle z'| \sigma_{Z'B} |z'\rangle \tag{4.37}$$

$$\leq \underbrace{\sum_x \max_z (c_{x,z}) |x\rangle\langle x|}_{\omega_X} \otimes \sigma_B. \tag{4.38}$$

Therefore,

$$H_\mu^\uparrow(Z|Z'B)_{\mathcal{S}_Z(\rho)} \leq \sup_{\sigma_B \in \mathcal{D}(B)} -D_\mu(\mathcal{M}_X(\rho_{AB})\|\omega_X \otimes \sigma_B). \tag{4.39}$$

By Theorem 3.1 we conclude for $\beta \geq \frac{1}{2}, \mu > 0, \delta \in \mathbb{R}$ with $\beta\mu\delta - 2\mu\delta - \beta + \mu + \delta = 0$ and $\frac{1}{\mu} + \frac{1}{\delta} \leq 2$,

$$- H_\beta(\mathcal{M}_X(\rho_{AB})\|\sigma_B) \leq D_\mu(\mathcal{M}_X(\rho_{AB})\|\omega_X \otimes \sigma_B) + \log \operatorname{tr} \left(\rho_X \omega_X^{\frac{1}{\delta'}} \right)^{\delta'}. \tag{4.40}$$

Substituting this into Eq. (4.39) we obtain

$$H_\mu^\uparrow(Z|Z'B)_{\mathcal{S}_Z(\rho)} \leq \sup_{\sigma_B \in \mathcal{D}(B)} H_\beta(\mathcal{M}_X(\rho_{AB})\|\sigma_B) + \log \operatorname{tr}\left(\rho_X \omega_X^{\frac{1}{\delta'}}\right)^{\delta'}. \tag{4.41}$$

By Theorem 3.6: for $\alpha, \mu, \gamma \geq \frac{1}{2}$ such that $\alpha\mu\gamma - 2\mu\gamma - \alpha + \mu + \gamma = 0$ and $\frac{1}{\mu} + \frac{1}{\gamma} \geq 2$

$$H_\alpha(\mathcal{S}_Z(\rho_{AB})\|\tau_B) - H_\gamma(\operatorname{tr}_Z(\mathcal{S}_Z(\rho_{AB}))\|\tau_B) \leq H_\mu^\uparrow(Z|Z'B)_{\mathcal{S}_Z(\rho)} \tag{4.42}$$

for all $\tau_B \in \mathcal{D}(B)$. Hence

$$H_\alpha(\mathcal{S}_Z(\rho_{AB})\|\tau_B) - H_\gamma(\operatorname{tr}_Z(\mathcal{S}_Z(\rho_{AB})\|\tau_B) \leq H_\beta^\uparrow(X|B)_\rho + \log \operatorname{tr}\left(\rho_X \omega_X^{\frac{1}{\delta'}}\right)^{\delta'}. \tag{4.43}$$

As in the proof of Theorem 4.1, the marginals on ZB and ZB' of the state $\mathcal{S}_Z(\rho_{AB})$ are equivalent and the conditional entropies are invariant under local isometries, therefore we obtain Eq. (4.30).

For the conditions on the parameters note that we can express μ in terms of β and δ:

$$\beta\mu\delta - 2\mu\delta - \beta + \mu + \delta = 0 \tag{4.44}$$

$$\implies \mu(\beta\delta - 2\delta + 1) = \beta - \delta \tag{4.45}$$

$$\implies \mu = \frac{\beta - \delta}{\beta\delta - 2\delta + 1}, \tag{4.46}$$

and similarly for α and γ.

Moreover the conditions determining the direction of the inequalities can be combined via μ:

$$\frac{1}{\mu} + \frac{1}{\delta} \leq 2 \tag{4.47}$$

$$\implies \frac{1}{\delta} \leq 2 - \frac{1}{\mu} \leq \frac{1}{\gamma}. \tag{4.48}$$

$\square$

This leads us to the state-independent version:

Theorem 4.3. *With the same conditions required for Eq. (4.33), we have*

$$H_\beta^\uparrow(X|B)_\rho + H_\gamma^\uparrow(Z|B)_\rho \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta, \tag{4.49}$$

where

$$q_\delta = \min_{\sigma \in \mathcal{D}(AB)} q_\delta(\sigma) = -\min_{0 \leq p \leq 1} \log \lambda_{\max}\left[\Delta_\delta(p)^{\delta'}\right] \tag{4.50}$$

with $\Delta_\delta(p) = p\sum_x (\max_z c_{x,z})^{\frac{1}{\delta'}} |x\rangle\langle x| + (1-p)\sum_z (\max_x c_{x,z})^{\frac{1}{\delta'}} |z\rangle\langle z|$.

Proof. We may rewrite

$$q_\delta(\rho) = \max_{0 \leq p \leq 1}\left(-\log\left[\operatorname{tr}\rho_A \Delta_\delta(p)\right]^{\delta'}\right) \tag{4.51}$$

$$\implies \min_{\sigma \in \mathcal{D}(AB)} q_\delta(\sigma) = -\log \max_{\sigma \in \mathcal{D}(AB)} \min_{0 \leq p \leq 1}\left[\operatorname{tr}\sigma_A \Delta_\delta(p)\right]^{\delta'}. \tag{4.52}$$

By the linearity in the arguments we may use the minimax theorem [68] to swap the optimisations, hence we obtain Eq. (4.50). $\square$

We conclude this section with the summary of the δ-dependent bounds:

Proposition 4.4. *With $q_\delta(\rho, \mathbb{X}, \mathbb{Z})$ defined as in Theorem 4.2 we have*

$$\lim_{\delta \to 1} q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = q(\rho, \mathbb{X}, \mathbb{Z}), \tag{4.53}$$

$$\lim_{\delta \to 0} q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = q_{\mathrm{MU}}. \tag{4.54}$$

It follows that

$$\lim_{\delta \to 1} q_\delta(\rho) = q(\rho), \tag{4.55}$$

$$\lim_{\delta \to 1} q_\delta = q_{\mathrm{CP}}. \tag{4.56}$$

Proof. Let $\{\lambda_x\}_x$ be the eigenvalues of ρ_X, i.e.

$$\rho_X = \sum_x \lambda_x |x\rangle\langle x|. \tag{4.57}$$

Note therefore, that ρ_X and ω_X are diagonal in the same basis and we can write

$$q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = -\log \operatorname{tr}\left(\rho_X \sum_x \left(\max_z c_{x,z}\right)^{\frac{1}{\delta'}} |x\rangle\langle x|\right)^{\delta'} = -\delta' \log \sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{\delta'}}. \tag{4.58}$$

We use l'Hôpital's rule, choosing

$$f(y) = \log \sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}} \implies \lim_{y \to \infty} f(y) = 0, \tag{4.59}$$

$$g(y) = \frac{1}{y} \implies \lim_{y \to \infty} g(y) = 0. \tag{4.60}$$

We have

$$f'(y) = \frac{d}{dy}\left(\log \sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}}\right) \tag{4.61}$$

$$= \frac{\sum_x \lambda_x \frac{d}{dy}\left(\left(\max_z c_{x,z}\right)^{\frac{1}{y}}\right)}{\sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}}} \tag{4.62}$$

$$= \frac{-\sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}} \log \max_z c_{x,z}}{y^2 \sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}}} \tag{4.63}$$

and $g'(y) = \frac{-1}{y^2}$. Hence

$$\lim_{y \to \infty} \frac{f(y)}{g(y)} = \lim_{y \to \infty} \frac{f'(y)}{g'(y)} = \lim_{y \to \infty} \frac{\sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}} \log \max_z c_{x,z}}{\sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}}} = \sum_x \lambda_x \log \max_z c_{x,z}. \tag{4.64}$$

Observe that $\delta' \to \infty \implies \delta \to 1$, therefore

$$\lim_{\delta \to 1} q_\delta(\rho, \mathbb{X}, \mathbb{Z}) = \sum_x \lambda_x \log(1/\max_z c_{x,z}). \tag{4.65}$$

Moreover, $\lim_{y \to 0} f(y) = \infty$ and $\lim_{y \to 0} g(y) = \infty$. So, again by l'Hôpital's rule,

$$\lim_{y \to 0} \frac{f(y)}{g(y)} = \lim_{y \to 0} \frac{\sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}} \log \max_z c_{x,z}}{\sum_x \lambda_x \left(\max_z c_{x,z}\right)^{\frac{1}{y}}} = \log \max_{x,z} c_{x,z} = -q_{\mathrm{MU}}. \tag{4.66}$$

The last two statements of the proposition are evident from Eq. (4.53) and the definitions of the relevant quantities. □

4.3.2 Proofs of the Rényi information exclusion relations

We begin with the unimproved Rényi information exclusion relations, which more directly generalise the Hall relation.

Theorem 4.5. *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

Let $\alpha, \gamma \geq \frac{1}{2}$ and $\beta > \frac{1}{2}$, satisfying both $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ and $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$, then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$I_\beta^\downarrow(X : B)_\rho + I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq r_\mathrm{H} - H_\alpha(\rho_{AB}\|\tau_B). \tag{4.67}$$

Proof of Theorem 4.5. Starting with Eq. (4.22), and choosing parameters which satisfy the conditions, we can write

$$H_{\tilde{\beta}}^\uparrow(X|B)_\rho + H_{\bar{\gamma}}(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\mathrm{MU}. \tag{4.68}$$

For each conditional entropy on the left-hand side we can derive the following inequalities from Eq. (3.6):

$$H_{\tilde{\beta}}^\uparrow(X|B)_\rho \leq H_{\tilde{\beta}}(\rho_X) - I_\beta^\downarrow(X : B)_\rho, \tag{4.69}$$

$$H_{\bar{\gamma}}(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq H_{\tilde{\gamma}}(\rho_Z) - I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B). \tag{4.70}$$

We can then write

$$-I_\beta^\downarrow(X : B)_\rho - I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \geq H_\alpha(\rho_{AB}\|\tau_B) - H_{\tilde{\beta}}(\rho_X) - H_{\tilde{\gamma}}(\rho_Z) + q_\mathrm{MU} \tag{4.71}$$

$$\implies I_\beta^\downarrow(X : B)_\rho + I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq H_{\tilde{\beta}}(\rho_X) + H_{\tilde{\gamma}}(\rho_Z) - H_\alpha(\rho_{AB}\|\tau_B) - q_\mathrm{MU} \tag{4.72}$$

$$\leq \log(d^2 c) - H_\alpha(\rho_{AB}\|\tau_B). \tag{4.73}$$

The last line is due to $H_\alpha(\rho) \leq \log d$ for all α.

We can optimise the parameters when $\tilde{\beta}, \tilde{\gamma} = 0$, which implies $\bar{\beta} = \beta$ and $\bar{\gamma} = \gamma$. We therefore have the familiar condition, $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$. Finally, note that for all $0 \leq \tilde{\beta} \leq \frac{1}{2} \implies \frac{1}{\tilde{\beta}} \geq 2$, hence we also have the condition $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$. □

The following optimal case then follows:

Corollary 4.6. *Given the same conditions as Theorem 4.5, for $\frac{1}{2} < \alpha < 2$, we have*

$$I_\alpha^\downarrow(X : B)_\rho + I_{\frac{1}{\alpha}}^\uparrow(B ; Z)_\rho \leq r_\mathrm{H} - H_\mathrm{min}(A|B)_\rho. \tag{4.74}$$

Proof. Taking the limit

$$\lim_{\alpha \to \infty} \beta\gamma - \frac{2\beta\gamma}{\alpha} - 1 + \frac{\beta}{\alpha} + \frac{\gamma}{\alpha} = \beta\gamma - 1, \tag{4.75}$$

we may conclude $\beta = \frac{1}{\gamma}$. We then let $\tau_B = \rho_B$ to have a tighter inequality, obtaining Eq. (4.74). □

This brings us to our final theorem, determining the improved Rényi information exclusion relations.

Theorem 4.7. *Let $\mathcal{M}_X \in \mathrm{CPTP}(A, X)$ and $\mathcal{M}_Z \in \mathrm{CPTP}(A, Z)$ be two incompatible measurement maps, defined by the ONBs $\mathbb{X}$ of $\mathcal{H}_X$ and $\mathbb{Z}$ of $\mathcal{H}_Z$.*

Given $\alpha, \gamma \geq \frac{1}{2}$, $\beta \in [1/2, 2]$ with

$$\frac{\alpha - \gamma}{\alpha\gamma - 2\gamma + 1} = \frac{1}{2 - \beta} \ \text{ and } \ \beta\gamma \leq 1, \tag{4.76}$$

then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$

$$I_\beta^\downarrow(X : B)_\rho + I_\gamma(\mathcal{M}_Z(\rho_{AB}) \| \tau_B) \leq r(\mathbb{X}, \mathbb{Z}) - H_\alpha(\rho_{AB} \| \tau_B) \tag{4.77}$$

and

$$I_\gamma(\mathcal{M}_X(\rho_{AB}) \| \tau_B) + I_\beta^\downarrow(Z : B)_\rho \leq r(\mathbb{Z}, \mathbb{X}) - H_\alpha(\rho_{AB} \| \tau_B). \tag{4.78}$$

Moreover, if $\gamma \leq 2 - \beta$ and

$$\frac{2\alpha - \alpha\gamma - 1}{\alpha - \gamma} = \frac{1}{2 - \beta}, \quad \frac{2\alpha - \alpha\beta - 1}{\alpha - \beta} = \frac{1}{2 - \gamma} \tag{4.79}$$

then

$$I_\beta^\downarrow(X : B)_\rho + I_\gamma^\downarrow(Z : B)_\rho \leq r_{\mathrm{CP}} - H_\alpha^\uparrow(A|B)_\rho. \tag{4.80}$$

Before the proof of this result, we first show how to derive a state-independent bound analogous to that in [44].

Lemma 4.10. *For $\alpha \geq \frac{1}{2}$, $\delta \in (-\infty, \alpha)$ such that $\frac{\alpha - \delta}{\alpha\delta - 2\delta + 1} \geq \frac{1}{2}$ and $\frac{(\alpha - 1)(\delta - 1)}{\alpha - \delta} + \frac{1}{\delta} \geq 1$,*

$$H_\alpha(\rho_X) - q_\delta(\rho, \mathbb{X}, \mathbb{Z}) \leq \log \sum_x \max_z c_{x,z}. \tag{4.81}$$

In particular

$$H_{\frac{1}{2}}(\rho_X) - q_{\mathrm{MU}} \leq \log \sum_x \max_z c_{x,z}. \tag{4.82}$$

Remark 4.11. *The conditions on α and δ in the above lemma are derived from certain conditions on a β parameter that appears in the proof but is not required in the final statement.*

Indeed, including this parameter provides an arguably nicer form of the condition: For $\alpha \geq \beta, \delta$ such that $\beta \geq \frac{1}{2}$ and $\frac{1}{\beta} + \frac{1}{\delta} \geq 2$. However, in order to present the lemma as a more self-contained statement, the β has been omitted.

Proof. Note that the trace is a CPTP, thus by the data-processing inequality (see Proposition 1.13) for all $\beta \geq \frac{1}{2}$,

$$D_\beta(\rho_A \| \sigma_A) \geq D_\beta(\mathrm{tr}\,\rho_A \| \mathrm{tr}\,\sigma_A) \tag{4.83}$$

$$= \log \left[(\mathrm{tr}\,\sigma_A)^{\frac{-1}{2\beta'}} (\mathrm{tr}\,\sigma_A)^{\frac{-1}{2\beta'}} \right]^{\beta'} \tag{4.84}$$

$$= -\log \mathrm{tr}\,\sigma_A. \tag{4.85}$$

To derive Eq. (4.81) we consider a similar interpolation as in Theorem 3.1. By choosing B to be trivial and given $\alpha, \beta \geq \frac{1}{2}, \delta \in \mathbb{R}$ such that $\alpha\beta\delta - 2\beta\delta - \alpha + \beta + \delta = 0$ and $\frac{1}{\beta} + \frac{1}{\delta} \geq 2$ we may conclude

$$H_\alpha(\rho_X) + \log \mathrm{tr}\left(\rho_X \omega_X^{\frac{1}{\delta'}}\right)^{\delta'} \leq -D_\beta(\rho_X \| \omega_X). \tag{4.86}$$

Let $\omega_X = \sum_x \max_z(c_{x,z}) |x\rangle\langle x|$, then by Eq. (4.85) we can bound the left-hand side

$$H_\alpha(\rho_X) + \log \operatorname{tr}\left(\rho_X \omega_X^{\frac{1}{\delta'}}\right)^{\delta'} \leq \log \operatorname{tr} \omega_X = \log \sum_x \max_z(c_{x,z}), \tag{4.87}$$

hence we obtain Eq. (4.81).

As in the proof of Theorem 4.1, we have $\frac{\alpha-\delta}{\alpha\delta-2\delta+1} = \beta$. Since we can choose any $\beta \geq \frac{1}{2}$ we have the condition in the lemma. Moreover

$$\frac{1}{\beta} + \frac{1}{\delta} \geq 2 \tag{4.88}$$

$$\implies \frac{\alpha\delta - 2\delta + 1}{\alpha - \delta} + \frac{1}{\delta} \geq 2 \tag{4.89}$$

$$\implies \frac{\alpha\delta - 2\delta + 1 + \alpha - \delta - \alpha + \delta}{\alpha - \delta} + \frac{1}{\delta} \geq 2 \tag{4.90}$$

$$\implies \frac{\alpha\delta - \delta - \alpha + 1}{\alpha - \delta} + 1 + \frac{1}{\delta} \geq 2 \tag{4.91}$$

$$\implies \frac{(\alpha - 1)(\delta - 1)}{\alpha - \delta} + \frac{1}{\delta} \geq 1. \tag{4.92}$$

To achieve the optimal situation in Eq. (4.82) we note the Rényi divergence is monotonically increasing in its parameter (see Eq. (1.132)) hence the choice which provides the tightest inequality in Eq. (4.85) is $\beta = \frac{1}{2}$. Eq. (B.23) requires $\alpha \geq \beta$ so, again through monotonicity, $\alpha = \frac{1}{2}$ yields the tightest inequality in Eq. (4.81). Substituting these values into our conditions we have

$$\delta = \frac{\alpha - \beta}{\alpha\beta - 2\beta + 1} = \frac{\frac{1}{2} - \frac{1}{2}}{\frac{1}{4} - 1 + 1} = 0 \tag{4.93}$$

and

$$\lim_{\delta \to 0^+} \frac{1}{\beta} + \frac{1}{\delta} = \lim_{\delta \to 0^+} 2 + \frac{1}{\delta} = \infty \geq 2. \tag{4.94}$$

$\square$

We now have all the tools required to prove Theorem 4.7.

Proof of Theorem 4.7. We begin with Eq. (4.22) and apply Theorem 3.3 to the measured conditional entropies:

$$H_{\bar{\beta}}^{\uparrow}(X|B)_\rho + H_{\bar{\gamma}}(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \geq H_\alpha(\rho_{AB}\|\tau_B) + q_\delta(\rho, \mathbb{X}, \mathbb{Z}) \tag{4.95}$$

$$\implies I_{\bar{\beta}}^{\downarrow}(X:B)_\rho + I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq H_{\tilde{\beta}}(\rho_X) + H_{\tilde{\gamma}}(\rho_Z) - H_\alpha(\rho_{AB}\|\tau_B) - q_\delta(\rho, \mathbb{X}, \mathbb{Z}), \tag{4.96}$$

where $\bar{\beta}, \beta, \bar{\gamma}, \gamma \geq \frac{1}{2}$ and $\tilde{\beta}, \tilde{\gamma} \geq 0$ such that $\bar{\beta}\beta\tilde{\beta} - 2\beta\tilde{\beta} - \bar{\beta} + \beta + \tilde{\beta} = 0$, $\bar{\gamma}\gamma\tilde{\gamma} - 2\gamma\tilde{\gamma} - \bar{\gamma} + \gamma + \tilde{\gamma} = 0$ and $\frac{1}{\bar{\beta}} + \frac{1}{\tilde{\beta}}, \frac{1}{\gamma} + \frac{1}{\tilde{\gamma}} \geq 2$.

We choose $\tilde{\beta} = \frac{1}{2}, \tilde{\gamma} \to 0$ and $\delta \to 0$. Hence we can use Eq. (4.82) and the fact that $H_{\tilde{\gamma}}(\rho_Z) \leq \log d$ for all $\tilde{\gamma} > 0$ to determine

$$I_{\bar{\beta}}^{\downarrow}(X:B)_\rho + I_\gamma(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq \log\left(d \sum_x \max_z c_{x,z}\right) - H_\alpha(\rho_{AB}\|\tau_B). \tag{4.97}$$

From the above choices we deduce

$$\bar{\beta}\beta\tilde{\beta} - 2\beta\tilde{\beta} - \bar{\beta} + \beta + \tilde{\beta} = 0 \tag{4.98}$$

$$\frac{\bar{\beta}\beta}{2} - \beta - \bar{\beta} + \beta + \frac{1}{2} = 0 \tag{4.99}$$

$$\bar{\beta}\beta - 2\bar{\beta} = -1 \tag{4.100}$$

$$\bar{\beta} = \frac{1}{2-\beta}, \tag{4.101}$$

$$\lim_{\tilde{\gamma}\to 0} \bar{\gamma}\gamma\tilde{\gamma} - 2\gamma\tilde{\gamma} - \bar{\gamma} + \gamma + \tilde{\gamma} = 0 \tag{4.102}$$

$$\bar{\gamma} - \gamma = 0 \tag{4.103}$$

$$\bar{\gamma} = \gamma. \tag{4.104}$$

Hence we can write

$$\frac{\alpha - \bar{\gamma}}{\alpha\bar{\gamma} - 2\bar{\gamma} + 1} = \frac{\bar{\beta} - \delta}{\bar{\beta}\delta - 2\delta + 1} = \mu \tag{4.105}$$

$$\implies \frac{\alpha - \gamma}{\alpha\gamma - 2\gamma + 1} = \frac{1}{2-\beta} = \mu. \tag{4.106}$$

Moreover,

$$\lim_{\delta\to 0^-} \frac{1}{\delta} \leq 2 - \frac{1}{\mu} \leq \frac{1}{\bar{\gamma}} = \frac{1}{\gamma} \tag{4.107}$$

$$\implies 2 - 2 + \beta \leq \frac{1}{\gamma} \tag{4.108}$$

$$\implies \beta\gamma \leq 1. \tag{4.109}$$

Noting $\mu \geq \frac{1}{2}$ we have the conditions on the first part of the theorem.

For the second part of the theorem, we perform the same procedure but instead start with $q_\delta(\rho, \mathbb{Z}, \mathbb{X})$ and choose both $\tilde{\beta}, \tilde{\gamma} = \frac{1}{2}$. This yields

$$\frac{\alpha - \bar{\gamma}}{\alpha\bar{\gamma} - 2\bar{\gamma} + 1} = \frac{\alpha - \frac{1}{2-\gamma}}{\alpha\frac{1}{2-\gamma} - 2\frac{1}{2-\gamma} + 1} \tag{4.110}$$

$$= \frac{2\alpha - \alpha\gamma - 1}{\alpha - 2 + 2 - \gamma} \tag{4.111}$$

$$= \frac{2\alpha - \alpha\gamma - 1}{\alpha - \gamma} \tag{4.112}$$

and

$$2 - \frac{1}{\mu} \leq \frac{1}{\bar{\gamma}} \tag{4.113}$$

$$\implies \beta \leq 2 - \gamma \tag{4.114}$$

$$\implies \beta + \gamma \leq 2. \tag{4.115}$$

Then, by choosing the minimum over the order of the measurements, we obtain Eq. (4.80). □

This leaves us with the proof of the optimal version of the above relations.

Corollary 4.8. *Given $\frac{1}{2} \leq \alpha \leq \frac{3}{2}$, then for all $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_B \in \mathcal{P}(B)$*

$$I_\alpha^\downarrow(X : B)_\rho + I_{2-\alpha}(\mathcal{M}_Z(\rho_{AB})\|\tau_B) \leq r(\mathbb{X}, \mathbb{Z}) - H_{\min}(A|B)_\rho. \tag{4.116}$$

Specifically,

$$I_{\frac{1}{2}}^\downarrow(X : B)_\rho + I_{\frac{3}{2}}^\downarrow(Z : B)_\rho \leq r_{\mathrm{CP}} - H_{\min}(A|B)_\rho. \tag{4.117}$$

Proof. Taking the limit $\alpha \to \infty$ in Eq. (4.76) we obtain

$$\frac{1}{\gamma} = \frac{1}{2 - \beta} \geq \frac{1}{2} \tag{4.118}$$

$$\implies 2 \geq \gamma = 2 - \beta \geq 0. \tag{4.119}$$

To obtain the second inequality we choose $\alpha = \frac{1}{2}$ or $\alpha = \frac{3}{2}$. Noting that, by minimising the Z measurement term, this inequality no longer depends on the order of the measurement and we may take the minimum bound. $\qquad\square$

Chapter 5

Discussion

5.1 Immediate observations

There are a couple of notions which follow from the main results.

5.1.1 Specialising to classical Rényi entropies

We have the option to consider classical states as a density operator whose eigenvectors form an ONB and are considered as a 'classical register' (see Section 1.3.3). In this case the eigenvalues of this density operator represent the values of the probability mass function of a classical random variable.

It is known that the quantum Rényi entropy $H_\alpha(\rho_X)$ of such a classical state is exactly the classical Rényi entropy $H_\alpha(X)$. By choosing the arguments as classical states in the Rényi divergence we recover the classical Rényi divergence (see Eq. (1.101)) and the associated conditional entropy and mutual information would then be strong candidates for classical Rényi versions of the well-known Shannon entropic quantities. Moreover, the chain and decomposition rules and any subsequent uncertainty relations would be equally applicable, providing useful tools for research in classical information theory.

5.1.2 Monotonicity in α

Each of the possible inequalities given by the families of inequalities in Chapters 3 and 4 additionally allows for a whole spectrum of weaker ancillary inequalities via the monotonicity of the Rényi divergence.

Considering this fact, we introduce a great deal more freedom in choosing specific parameters beyond those stipulated in the conditions of each theorem. This significantly widens the applicability of each result and provides an extra level of generality.

5.2 Advantages and drawbacks

Clearly, the divergence equations presented not only give us some insight into the fundamental relationships between these quantities but also provide important tools in the derivation of Rényi versions of the routine comparisons for Shannon and von Neumann quantities and beyond. The structure afforded may help in consolidating the theory of generalised Rényi quantities and provide some possible candidates for further study. The results themselves are quite general, which is usually desirable given that they are more flexible and therefore applicable in more situations. However, this level of generality has the unwanted consequence of obfuscating the actual utility of the results.

5.2.1 Place within the broader theory

Smooth entropies

Smooth entropies are defined as optimisations of Rényi conditional entropies, but in particular we are concerned with the min- and max-entropies,

$$H_\infty^\uparrow(A|B)_\rho = H_{\min}(A|B)_\rho \quad \text{and} \quad H_{\frac{1}{2}}^\uparrow(A|B)_\rho = H_{\max}(A|B)_\rho. \tag{5.1}$$

These particular entropies are of interest as they can be calculated through semi-definite programs which are generally more efficient than direct computation. The smooth min- and max-entropies are then considered as optimisations over states τ that are ε-close in the *purified distance* to the given state ρ.

We define the *smoothing ball* $\mathcal{B}^\varepsilon(A, \rho) := \{\tau \in \mathcal{D}(A) : P(\tau, \rho) \leq \varepsilon\}$, where

$$P(\tau, \rho) = \sqrt{1 - \left(\mathrm{tr} \left| \sqrt{\rho}\sqrt{\tau} \right| + \sqrt{(1 - \mathrm{tr}\,\rho)(1 - \mathrm{tr}\,\tau)} \right)^2}. \tag{5.2}$$

Then the smooth entropies are defined

$$H^\varepsilon_{\min}(A|B)_\rho = \max_{\bar\rho_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})} H_{\min}(A|B)_{\bar\rho}, \tag{5.3}$$

$$H^\varepsilon_{\max}(A|B)_\rho = \min_{\bar\rho_{AB} \in \mathcal{B}^\varepsilon(\rho_{AB})} H_{\max}(A|B)_{\bar\rho}. \tag{5.4}$$

These 'smoothed' entropies are also calculable by semi-definite programs and exhibit similar duality relations as the Rényi entropies they are based on. Of particular note is the asymptotic equipartition property which essentially states for normalised smooth min- or max-entropies on a number of copies of the same state, the limit as the number of copies approaches infinity is the von Neumann entropy of the original state, i.e for $\rho_{AB} \in \mathcal{D}(AB)$ and $\rho_{A^n B^n} = \rho_{AB}^{\otimes n} = \underbrace{\rho_{AB} \otimes \cdots \otimes \rho_{AB}}_{n \text{ times}} \in \mathcal{D}(A^n B^n)$,

$$\lim_{n \to \infty} \frac{1}{n} H^\varepsilon_{\min}(A^n|B^n)_{\rho^{\otimes n}} = \lim_{n \to \infty} \frac{1}{n} H^\varepsilon_{\max}(A^n|B^n)_{\rho^{\otimes n}} = H(A|B)_\rho. \tag{5.5}$$

See [46] and [42, Chapter 6] for a detailed treatment of these quantities, their properties and applications.

Some chain and decomposition rules similar to those covered in Chapter 3 have been established for smooth min and max-entropies (see [69, 70]) but it is unknown whether these rules can be extended to smooth entropies of general Rényi order. This work may provide some tools or scaffolding to that end and aid in establishing a more coherent theory of smooth entropies.

Conditional mutual information

The relationships between the established divergence inequalities would indicate that there is a more general structure available for tripartite and, in turn, multipartite systems.

The following diagram summarises what we have established and the 'gaps' in this structure:

$$\text{General tripartite divergence inequality} \xdashrightarrow{\inf_{\sigma_{BC} \in \mathcal{D}^*(BC)}} I^*_\beta(B; A|C)_\rho \gtrless H_\gamma(\rho_{BC}\|\tau_C) - H_\alpha(\rho_{ABC}\|\tau_{AC})$$

C trivial $\Big\downarrow$ $\qquad\qquad\qquad \tau_{AC} = \tau_C \Big\downarrow$

$$-H_\alpha\left(\rho_{AB}\|\tau_A\right) \lessgtr D_\beta\left(\rho_{AB}\|\sigma_B \otimes \tau_B\right) + \log\left(\mathrm{tr}\,\rho_B \sigma_B^{\frac{1}{\gamma'}}\right)^{\gamma'} \qquad H_\alpha(\rho_{ABC}\|\tau_C) \gtrless H^\uparrow_\beta(A|BC)_\rho + H_\gamma(\rho_{BC}\|\tau_C)$$

$\inf_{\sigma_B \in \mathcal{D}^*(B)}$ $\Big\downarrow$ $\qquad\qquad\qquad C \text{ trivial} \qquad\qquad\qquad\qquad C \text{ trivial} \Big\downarrow$

$$I_\beta(\rho_{AB}\|\tau_A) \gtrless H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A) \xrightarrow[\tau_A = \mathbb{1}_A]{} H_\alpha(\rho_{AB}) \gtrless H^\uparrow_\beta(A|B)_\rho + H_\gamma(\rho_B)$$

The dashed lines are proposed relationships and the dotted lines are relationships formed by the composition of the explicit transformations.

In essence, this examination implies that there is a general tripartite comparison of Rényi quantities from which all inequalities of this type follow. Perhaps of more immediate interest is the indication that there are decomposition/chain rules for some Rényi divergence based conditional mutual information (denoted $I_\beta^*(B; A|C)_\rho$ in the diagram). The (von Neumann) quantum conditional mutual information has found many uses in quantum information theory, hence it is currently an important open problem to derive an operationally significant Rényi conditional mutual information which is useful in applications. There are several candidates for a quantity of this type (see [47] and references therein), but one that is compatible with this pre-existing structure would be a strong choice for further study.

5.2.2 Applicability

Alternative generalised entropies

There are some other possible candidates for a generalised divergence from which entropic quantities could be similarly derived. Notable are the Tsallis entropies [71] and the Petz divergence [59]. However, these quantities do not have the structure which allows us to derive comparisons via interpolation. That said, the Petz divergence is still closely related to the Rényi divergence, sharing many of the same properties and there exist duality relations involving both quantities [42].

Although the quantum Rényi divergence has already been established as one of the forerunners in the possible choices for the basis of a generalised quantum Rényi entropy framework, having found many applications and utility in quantum information theory, the structure revealed in this work further reinforces the compatibility and usefulness of this definition.

Generality vs. usability

The comparisons in Chapter 3 are, in isolation, relatively simple expressions. However, in applications they are often used in conjunction with one another, as evidenced by the proofs in Chapter 4.

The results of composing the conditions on the parameters for each individual comparison are rather unwieldy and generally take the form of obscure polynomial expressions. An immediate solution is to make choices which yield the tightest versions of the relevant inequality. This notion has already been applied in this thesis, albeit the final conditions still often being difficult to interpret.

Clearly, there is a trade-off in how general we can formulate these types of statements and their accessibility for a broader audience. For the moment, it seems the possible loss of a small amount of information at the gain of usability and consistency is worth the cost.

Overall, the analysis of the relationship between, and valid ranges of, the parameters given a particular set of comparisons has proved time-consuming and beyond the scope of this work. Moreover, any attempts to do so have been unfruitful in establishing a consistent underlying structure, given one exists.

5.3 Future work

We now suggest some possible directions for research or study which follow from the material covered in this thesis.

5.3.1 Numerical simulations

At the moment we have only the abstract representation of most of the relationships detailed in this thesis. By running numeric simulations and producing visualisations thereof we may gain some insights into the relationships between the Rényi orders, dimension, degree of entanglement and the relative tightness of each comparisons. An analysis of these trends and correlations may illuminate

possible avenues for further refinements and improvements. These simulations are relatively easy to perform when calculating the divergence of known matrices but the problem becomes more involved when considering the necessary optimisations – fortunately there are packages and/or libraries available for most high-level programming languages that allow for efficient optimisation over convex sets. The nature of the models require exponentially more processing power for calculations involving higher dimensions, so obtaining sufficiently broad samples may still be a significant undertaking, requiring dedicated time and resources.

5.3.2   Multiple measurements

An interesting direction to take these general uncertainty relations would be to formulate statements involving more than two measurements. Of course, given uncertainty relations of the form

$$H_{\alpha_1}(X_1) + H_{\alpha_2}(X_2) \geq q_{1,2} \quad \text{and} \quad H_{\alpha_3}(X_3) + H_{\alpha_4}(X_4) \geq q_{3,4}, \tag{5.6}$$

we may immediately derive

$$H_{\alpha_1}(X_1) + H_{\alpha_2}(X_2) + H_{\alpha_3}(X_3) + H_{\alpha_4}(X_4) \geq q_{\max}, \tag{5.7}$$

where $q_{\max} = \max_{\{i,j,k,l\}=\{1,2,3,4\}} q_{i,j} + q_{k,l}$. However, this does not yield any advantage compared to the uncertainty relations from which it was derived.

We would instead be investigating inequalities of the form

$$\sum_n H_{\alpha_n}(X_n|B)_\rho \geq q, \tag{5.8}$$

such that the bound $q \geq q_{\max}$. Given the Rényi decomposition rule and general comparisons we could then derive

$$\sum_n I_{\alpha_n}(X_n : B)_\rho \leq r \tag{5.9}$$

with r dependent on q.

For a more complete review of multiple measurement uncertainty relations see [20, Section III.G] and the references therein.

5.3.3   POVMs and tripartite uncertainty relations

In Chapter 4 we only consider measurements on ONBs as it provides a simpler scaffold. Indeed, most of the results in Section 1.6.2 have at least one generalisation to POVMs. It is currently an open question whether these generalisations extend to general Rényi uncertainty relations.

A feature of these POVM-based von Neumann relations is that they are often expressed on tripartite systems. It is then likely any Rényi version would also require the consideration of tripartite systems. For a more detailed treatment of these situations see [44] and [42].

5.3.4   Adapting Pisier norms

As stated in Section 2.2, the full generality of Pisier's norms has not been used in the interpolation employed in Chapter 3. This does not mean that the close relationship between these norms and Rényi divergence based quantities does not merit further thought. Indeed, the general framework of Pisier contains some useful results which could provide important insights into the nature and structure of these quantities.

The main issue with this pursuit is the ability to make use of this theory without having to rely too heavily on the unnecessarily abstract mathematics it is based on. There is currently work

underway to adapt this theory in a cohesive way, accessible from a quantum information theory perspective, and to investigate the implications of such a framework.

Appendix A

An alternative proof of the decomposition rules

We include a different approach[*] to the decomposition rules which follows the method employed in [37] in establishing Theorem 1.15. In a similar vein to the differences between the two versions of the bipartite chain rule (Theorem 3.4 and Corollary 3.7), this approach produces slightly different valid ranges for the parameters compared to Theorem 3.3. Although for certain applications the ranges on this alternative result may be preferable, for our purposes in the main body of the thesis they are more constrained than desirable.

Reproduced from A. McKinlay and M. Tomamichel, "Decomposition rules for quantum Rényi mutual information with an application to information exclusion relations," *Journal of Mathematical Physics*, vol. 61, no. 7, p. 072202, 2020., with the permission of AIP Publishing.

Theorem A.1. *Let* $\alpha\beta\gamma - 2\beta\gamma - \alpha + \beta + \gamma = 0$ *with* $\alpha \geq 0$, $\beta > \frac{1}{2}$ *and* $\gamma \geq \frac{1}{2}$. *Then, for* $\rho_{AB} \in \mathcal{D}(AB)$ *and* $\tau \in \mathcal{P}(A)$, *if* $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$

$$I_\beta(\rho_{AB}\|\tau_A) \geq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.1}$$

Otherwise, if $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$

$$I_\beta(\rho_{AB}\|\tau_A) \leq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.2}$$

This method draws on several components, which we break up into the following sections.

A.1 Operator-vector correspondence

We may establish a correspondence between vectors on a composite Hilbert space and operators mapping from one subspace to another. Given bases $\{|e_i\rangle\}_i$ and $\{|f_i\rangle\}_i$ for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, define

$$\mathrm{Op}_{A\to B}(|e_i\rangle \otimes |f_i\rangle) = |f_i\rangle\langle e_i|. \tag{A.3}$$

Accordingly, for $|\psi\rangle \in \mathcal{H}_{AB}$ such that $|\psi\rangle = \sum_{i,j} \lambda_{ij} |e_i\rangle \otimes |f_j\rangle$ we have $\mathrm{Op}_{A\to B}(|\psi\rangle) = \sum_{i,j} \lambda_{ij} |f_j\rangle\langle e_i|$. There are some useful properties of this correspondence, for proofs see [9, Section 1.1].

Lemma A.2. *Let* $|\psi\rangle \in \mathcal{H}_{AB}$, $M \in \mathcal{L}(A)$ *and* $N \in \mathcal{L}(B)$. *Then*

$$\mathrm{Op}_{A\to B}[(M \otimes N)|\psi\rangle] = N\mathrm{Op}_{A\to B}(|\psi\rangle)M^\top. \tag{A.4}$$

Lemma A.3. *Let* $|\psi\rangle \in \mathcal{H}_{AB}$. *Then*

$$\| |\psi\rangle \|_2 = \sqrt{\langle\psi|\psi\rangle} = \left(\mathrm{tr}\left[\mathrm{Op}_{A\to B}(|\psi\rangle)^\dagger \mathrm{Op}_{A\to B}(|\psi\rangle)\right]\right)^{\frac{1}{2}} = \|\mathrm{Op}_{A\to B}(|\psi\rangle)\|_2. \tag{A.5}$$

---

[*]This approach was used to produce the results in [72] which in turn were applied to the derivation of information exclusion relations. These relations have subsequently been improved in this thesis by the more amenable ranges found in Theorem 3.3.

Lemma A.4. *Let $|\psi\rangle \in \mathcal{H}_{AB}$, $\rho_{AB} = |\psi\rangle\langle\psi|$. Then*

$$\rho_A = \mathrm{Op}_{A\to B}(|\psi\rangle)^\dagger \mathrm{Op}_{A\to B}(|\psi\rangle) \quad and \quad \rho_B = \mathrm{Op}_{A\to B}(|\psi\rangle)\mathrm{Op}_{A\to B}(|\psi\rangle)^\dagger. \tag{A.6}$$

A.2 Identities in terms of operator-vector correspondence

We begin the main part of the demonstration by establishing the following identities for the relevant entropic quantities:

Lemma A.5. *For a pure state $|\varphi\rangle \in \mathcal{H}_{ABC}$ with $|\varphi\rangle\langle\varphi| = \rho_{ABC}$, let $X = X_{B\to AC} = \mathrm{Op}_{B\to AC}(|\varphi\rangle)$. Given $(\tau_A \otimes \mathbb{1}_B) \gg \rho_{AB}$ and $\alpha \in (0,1) \cup (1,\infty)$ we have*

$$H_\alpha(\rho_{AB}\|\tau_A) = -\log \sup_{\sigma_C \in \mathcal{D}(C)} \left\| \left(\tau_A^{-1} \otimes \sigma_C\right)^{\frac{1}{2\alpha'}} X \right\|_2^{2\alpha'}, \tag{A.7}$$

$$H_\alpha(\rho_B) = -\log \|X\|_{2\alpha}^{2\alpha'}. \tag{A.8}$$

If in addition $\alpha > \frac{1}{2}$,

$$I_\alpha(\rho_{AB}\|\tau_A) = \log \sup_{\sigma_C \in \mathcal{D}(C)} \left\| \left(\tau_A^{-1} \otimes \sigma_C\right)^{\frac{1}{2\alpha'}} X \right\|_{2\hat{\alpha}}^{2\alpha'}. \tag{A.9}$$

The proof of Lemma A.5 relies on the operator-vector correspondence.

Proof of Eq. (A.7). From equation (19) in [28], we can write

$$H_\alpha(\rho_{AB}\|\tau_A) = -\log \sup_{\sigma_C \in \mathcal{D}(C)} \langle\varphi| \tau_A^{\frac{-1}{\alpha'}} \otimes \mathbb{1}_B \otimes \sigma_C^{\frac{1}{\alpha'}} |\varphi\rangle^{\alpha'}. \tag{A.10}$$

Also, using Lemma A.2 we have

$$\mathrm{Op}_{B\to AC}\left(\tau_A^{\frac{-1}{2\alpha'}} \otimes \mathbb{1}_B \otimes \sigma_C^{\frac{1}{2\alpha'}} |\varphi\rangle\right) = \left(\tau_A^{\frac{-1}{2\alpha'}} \otimes \sigma_C^{\frac{1}{2\alpha'}}\right) X. \tag{A.11}$$

From this we can deduce, using Lemma A.3,

$$\langle\varphi| \tau_A^{\frac{-1}{\alpha'}} \otimes \mathbb{1}_B \otimes \sigma_C^{\frac{1}{\alpha'}} |\varphi\rangle^{\alpha'} = \left\| \tau_A^{\frac{-1}{2\alpha'}} \otimes \mathbb{1}_B \otimes \sigma_C^{\frac{1}{2\alpha'}} |\varphi\rangle \right\|_2^{2\alpha'} \tag{A.12}$$

$$= \left\| \mathrm{Op}_{B\to AC}\left(\tau_A^{\frac{-1}{2\alpha'}} \otimes \mathbb{1}_B \otimes \sigma_C^{\frac{1}{2\alpha'}} |\varphi\rangle\right) \right\|_2^{2\alpha'} \tag{A.13}$$

$$= \left\| \left(\tau_A^{\frac{-1}{2\alpha'}} \otimes \sigma_C^{\frac{1}{2\alpha'}}\right) X \right\|_2^{2\alpha'}. \tag{A.14}$$

$\square$

Proof of Eq. (A.8). From Lemma A.4 we can see that $\rho_B = X^\dagger X$, hence we have

$$H_\alpha(B)_\rho = -\log \left\| X^\dagger X \right\|_\alpha^{\alpha'} = -\log \|X\|_{2\alpha}^{2\alpha'}, \tag{A.15}$$

where we have also used the identity, Proposition 1.7. $\square$

The proof of Eq. (A.9) relies on the duality result in Proposition 1.12.

Proof of Eq. (A.9). Using operator-vector correspondence we can re-express the generalised Rényi mutual information using operator norms, i.e. for $X_{C \to AB} = \mathrm{Op}_{C \to AB}(|\varphi\rangle)$,

$$I_\alpha(\rho_{AB}\|\tau_A) = \frac{1}{\alpha - 1} \log \inf_{\sigma_B \in \mathcal{D}(A)} \mathrm{tr}\left(\left[(\tau_A \otimes \sigma_B)^{\frac{-1}{2\alpha'}} \rho_{AB} (\tau_A \otimes \sigma_B)^{\frac{-1}{2\alpha'}}\right]^\alpha\right) \tag{A.16}$$

$$= \log \inf_{\sigma_B \in \mathcal{D}(A)} \left\|(\tau_A \otimes \sigma_B)^{\frac{-1}{2\alpha'}} X_{C \to AB} X_{C \to AB}^\dagger (\tau_A \otimes \sigma_B)^{\frac{-1}{2\alpha'}}\right\|_\alpha^{\alpha'} \tag{A.17}$$

$$= \log \inf_{\sigma_B \in \mathcal{D}(A)} \left\|(\tau_A \otimes \sigma_B)^{\frac{-1}{2\alpha'}} X_{C \to AB}\right\|_{2\alpha}^{2\alpha'}. \tag{A.18}$$

By the duality of the generalised mutual information and Eq. (A.18), we can write

$$I_\alpha(\rho_{AB}\|\tau_A) = -I_{\hat\alpha}(\rho_{AC}\|\tau_A^{-1}) \tag{A.19}$$

$$= -\log \inf_{\omega_C \in \mathcal{D}(C)} \left\|\left(\tau_A^{\frac{1}{2\hat\alpha'}} \otimes \omega_C^{\frac{-1}{2\hat\alpha'}}\right) X\right\|_{2\hat\alpha}^{2\hat\alpha'} \tag{A.20}$$

$$= \log \sup_{\omega_C \in \mathcal{D}(C)} \left\|\left(\tau_A^{\frac{-1}{2\alpha'}} \otimes \omega_C^{\frac{1}{2\alpha'}}\right) X\right\|_{2\hat\alpha}^{2\alpha'}. \tag{A.21}$$

where in the last line we used the fact that $\hat\alpha' = -\alpha'$. □

A.3 Interpolating for valid choices of θ

Theorem A.1 can be proved directly from the following propositions which make use of the above identities.

Proposition A.6. *Let α, β, γ be such that $\alpha' = \beta' + \gamma'$, $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_A \in \mathcal{P}^*(A)$. Then the following holds:*

For $\alpha \in (1, 2), \beta, \gamma \in (1, \infty)$, we find

$$I_\beta(\rho_{AB}\|\tau_A) \geq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.22}$$

For $\alpha \in [2/3, 1), \beta \in (1/2, 1), \gamma \in [1/2, 1)$, we find

$$I_\beta(\rho_{AB}\|\tau_A) \leq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.23}$$

Proof. Choose $F(z) = (\tau_A^{-1} \otimes \sigma_C)^{\frac{z}{2\beta'}} X$, $\theta = \frac{\beta'}{\alpha'}$, $p_0 = 2\gamma$, $p_1 = 2\hat\beta$. With these choices we can determine $\theta = \frac{1}{\alpha'}(\alpha' - \gamma') = 1 - \frac{\gamma'}{\alpha'}$, hence $1 - \theta = \frac{\gamma'}{\alpha'}$.

We can also calculate the appropriate value of p_θ to use Theorem 2.12:

$$\frac{1}{p_\theta} = \frac{\gamma'}{2\alpha'\gamma} + \frac{\beta'}{2\alpha'\hat\beta} \implies \frac{2\alpha'}{p_\theta} = \gamma' - 1 + \beta' + 1 = \gamma' + \beta', \tag{A.24}$$

thus we can conclude that $p_\theta = 2$.

We can therefore calculate that

$$\|F(\theta)\|_{p_\theta} = \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\alpha'}} X\right\|_2. \tag{A.25}$$

Additionally,

$$\|F(\mathbf{i}\,t)\|_{p_0} = \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{\mathbf{i}\,t}{2\beta'}} X\right\|_{2\gamma} \tag{A.26}$$

and

$$\|F(1+\mathbf{i}\,t)\|_{p_1} = \left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1+\mathbf{i}\,t}{2\beta'}}X\right\|_{2\hat{\beta}}. \tag{A.27}$$

Since $(\tau_A^{-1}\otimes\sigma_C)^{\frac{\mathbf{i}\,t}{2\beta'}}$ is unitary for all $t\in\mathbb{R}$ we can write

$$\mathsf{M}_0 = \|X\|_{2\gamma} \quad\text{and}\quad \mathsf{M}_1 = \left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\beta'}}X\right\|_{2\hat{\beta}}. \tag{A.28}$$

Applying Theorem 2.12 we have

$$\left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\alpha'}}X\right\|_2 \le \|X\|_{2\gamma}^{\frac{\gamma'}{\alpha'}}\left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\beta'}}X\right\|_{2\hat{\beta}}^{\frac{\beta'}{\alpha'}}. \tag{A.29}$$

First, consider $\alpha' > 0$. Maximising over σ_C on both sides we have

$$\sup_{\sigma_C\in\mathcal{D}(C)}\left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\alpha'}}X\right\|_2^{2\alpha'} \le \|X\|_{2\gamma}^{2\gamma'}\sup_{\sigma_C\in\mathcal{D}(C)}\left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\beta'}}X\right\|_{2\hat{\beta}}^{2\beta'}. \tag{A.30}$$

Using Lemma A.5, we can rewrite this as

$$-H_\alpha(\rho_{AB}\|\tau_A) \le -H_\gamma(\rho_B) + I_\beta(\rho_{AB}\|\tau_A) \tag{A.31}$$
$$\implies I_\beta(\rho_{AB}\|\tau_A) \ge H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.32}$$

If instead $\alpha' < 0$, we obtain

$$\sup_{\sigma_C\in\mathcal{D}(C)}\left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\alpha'}}X\right\|_2^{2\alpha'} \ge \|X\|_{2\gamma}^{2\gamma'}\sup_{\sigma_C\in\mathcal{D}(C)}\left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\beta'}}X\right\|_{2\hat{\beta}}^{2\beta'}, \tag{A.33}$$

giving us Eq. (A.23). The valid ranges can be determined using Lemma B.3. $\qquad\square$

Proposition A.7. *Let α,β,γ be such that $\alpha' = \beta' + \gamma'$, $\rho_{AB}\in\mathcal{D}(AB)$ and $\tau_A\in\mathcal{P}^*(A)$. Then the following holds:*

For $\alpha\in(0,1), \beta\in(1/2,1), \gamma\in(1,\infty)$, we find

$$I_\beta(\rho_{AB}\|\tau_A) \ge H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.34}$$

For $\gamma\in[1/2,1), \beta\in(1,2), \alpha\in(1,\infty)$, we find

$$I_\beta(\rho_{AB}\|\tau_A) \le H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.35}$$

Proof. Choose $F(z) = (\tau_A^{-1}\otimes\sigma_C)^{\frac{z}{2\alpha'}}X$, $\quad\theta = \frac{\alpha'}{\beta'}$, $\quad p_0 = 2\gamma$, $\quad p_1 = 2$. We have, as before, $1-\theta = \frac{-\gamma'}{\beta'}$ and through a similar calculation we can conclude that $p_\theta = 2\hat{\beta}$.

We have

$$\|F(\theta)\|_{p_\theta} = \left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\beta'}}X\right\|_{2\hat{\beta}}, \quad \|F(\mathbf{i}\,t)\|_{p_0} = \left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{\mathbf{i}\,t}{2\alpha'}}X\right\|_{2\gamma} \tag{A.36}$$

$$\text{and } \|F(1+\mathbf{i}\,t)\|_{p_1} = \left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1+\mathbf{i}\,t}{2\alpha'}}X\right\|_2, \tag{A.37}$$

hence $\mathsf{M}_0 = \|X\|_{2\gamma}$ and $\mathsf{M}_1 = \left\|(\tau_A^{-1}\otimes\sigma_C)^{\frac{1}{2\alpha'}}X\right\|_2$.

Applying Theorem 2.12 we have

$$\left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\beta'}} X\right\|_{2\hat{\beta}} \leq \|X\|_{2\gamma}^{\frac{-\gamma'}{\beta'}} \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\alpha'}} X\right\|_{2}^{\frac{\alpha'}{\beta'}}. \tag{A.38}$$

First, we consider the case where $\beta' > 0$. It follows that

$$\left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\beta'}} X\right\|_{2\hat{\beta}}^{2\beta'} \leq \|X\|_{2\gamma}^{-2\gamma'} \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\alpha'}} X\right\|_{2}^{2\alpha'}. \tag{A.39}$$

As in Proposition A.6, we can maximise over σ_C on both sides. Hence we arrive at Eq. (A.35). Repeating the same process with the assumption $\beta' < 0$ yields Eq. (A.34). We can again refer to Lemma B.3 to determine the valid ranges. $\qquad\square$

Proposition A.8. *Let α, β, γ be such that $\alpha' = \beta' + \gamma'$, $\rho_{AB} \in \mathcal{D}(AB)$ and $\tau_A \in \mathcal{P}^*(A)$. Then the following holds:*

For $\alpha \in (0,1), \gamma \in [1/2, 1), \beta \in (1, \infty)$, we find

$$I_\beta(\rho_{AB}\|\tau_A) \geq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.40}$$

For $\beta \in (1/2, 1), \gamma \in (1, 2), \alpha \in (1, \infty)$, we find

$$I_\beta(\rho_{AB}\|\tau_A) \leq H_\gamma(\rho_B) - H_\alpha(\rho_{AB}\|\tau_A). \tag{A.41}$$

Proof. Choose $F(z) = (\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\beta'} - z\frac{\gamma'}{2\alpha'\beta'}} X$, $\quad \theta = \frac{\alpha'}{\gamma'}$, $\quad p_0 = 2\hat{\beta}$, $\quad p_1 = 2$. As above, $1 - \theta = \frac{-\beta'}{\gamma'}$ and $p_\theta = 2\gamma$.

We have

$$\|F(\theta)\|_{p_\theta} = \|X\|_{2\gamma}, \quad \|F(\mathbf{i}\,t)\|_{p_0} = \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\beta'} - \frac{\mathbf{i}\,t\gamma'}{2\alpha'\beta'}} X\right\|_{2\hat{\beta}} \tag{A.42}$$

$$\text{and } \|F(1 + \mathbf{i}\,t)\|_{p_1} = \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\alpha'} - \frac{\mathbf{i}\,t\gamma'}{2\alpha'\beta'}} X\right\|_{2}, \tag{A.43}$$

hence $\mathsf{M}_0 = \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\beta'}} X\right\|_{2\hat{\beta}}$ and $\mathsf{M}_1 = \left\|(\tau_A^{-1} \otimes \sigma_C)^{\frac{1}{2\alpha'}} X\right\|_{2}$. Applying Theorem 2.12 and performing the same procedure as in Propositions A.6 and A.7, for both $\gamma' > 0$ and $\gamma' < 0$ we obtain Eqs. (A.40) and (A.41). For the valid ranges, we have a similar situation as in Proposition A.7 but with symmetry in β and γ. $\qquad\square$

A.4 Consolidation and analysis

We may now prove Theorem A.1:

Proof of Theorem A.1. We first combine the three above propositions and examine the valid ranges. We have from Lemma B.3 that the propositions cover all possible permutations of the parameters, and hence all valid values of α, β and γ.

For the forward inequality, i.e. Eqs. (A.22), (A.34) and (A.40) we can see that either $(\alpha, \beta, \gamma > 1)$, $(\alpha, \gamma < 1, \beta > 1)$ or $(\alpha, \beta < 1, \gamma > 1)$, which all satisfy $\frac{1}{\beta} + \frac{1}{\gamma} \leq 2$.

For the reverse inequality, i.e. Eqs. (A.23), (A.35) and (A.41) we have either $(\alpha, \beta, \gamma < 1)$, $(\alpha, \gamma > 1, \beta > 1)$ or $(\alpha, \beta > 1, \gamma < 1)$, which all satisfy $\frac{1}{\beta} + \frac{1}{\gamma} \geq 2$. We again obtain the final conditions and the extension to positive semi-definite matrices in the same manner as Theorem 3.1. $\qquad\square$

Appendix B

Other useful results

B.1 Background results

Proposition B.1.

$$\lim_{\alpha \to 1} H_\alpha(X) = H(X) \tag{B.1}$$

Proof. We have

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_x p(x)^\alpha \right). \tag{B.2}$$

We choose:

$$f(\alpha) = \log \left(\sum_x p(x)^\alpha \right), \quad g(\alpha) = 1 - \alpha. \tag{B.3}$$

Observe that $\lim_{\alpha \to 1} f(\alpha) = \log \sum_x p(x) = \log 1 = 0$ and $\lim_{\alpha \to 1} g(\alpha) = 0$. Hence, we may use l'Hôpital's rule to assert $\lim_{\alpha \to 1} \frac{f(\alpha)}{g(\alpha)} = \lim_{\alpha \to 1} \frac{f'(\alpha)}{g'(\alpha)}$. We compute

$$f'(\alpha) = \frac{d}{d\alpha} \left[\log \left(\sum_x p(x)^\alpha \right) \right] = \frac{\sum_x \frac{d}{d\alpha} p(x)^\alpha}{\sum_x p(x)^\alpha} = \frac{\sum_x p(x)^\alpha \log p(x)}{\sum_x p(x)^\alpha}. \tag{B.4}$$

and $g'(\alpha) = -1$. Therefore we may write

$$\lim_{\alpha \to 1} \frac{f(\alpha)}{g(\alpha)} = \lim_{\alpha \to 1} -\frac{\sum_x p(x)^\alpha \log p(x)}{\sum_x p(x)^\alpha} = -\sum_x p(x) \log p(x). \tag{B.5}$$

$\square$

B.2 Ancillary results

We make use of a modified version of Lemma 12 from [28].

Lemma B.2. *Let $p \in \mathbb{R}^+ \setminus \{0, 1\}$ and $p' \in \mathbb{R} \setminus [0, 1]$ be such that $\frac{1}{p} + \frac{1}{p'} = 1$. Then for $M \in \mathcal{L}(A, B)$ such that $M^\dagger M = X \in \mathcal{P}(A)$,*

$$\|M\|_{2p}^{2p'} = \|X\|_p^{p'} = \sup_{\sigma_A \in \mathcal{D}^*(A)} \left(\operatorname{tr} X \sigma_A^{\frac{1}{p'}} \right)^{p'} = \sup_{\sigma_A \in \mathcal{D}^*(A)} \left\| \Gamma_{\mathbb{1}_B, \sigma_A}^{\frac{1}{p'}}(M) \right\|_2^{2p'}. \tag{B.6}$$

Proof. We restrict to density operators with full support which guarantees $\sigma_A \gg X$, hence by Lemma 12 from [28]

$$\|X\|_p^{p'} = \begin{cases} \left(\sup_{\sigma_A \in \mathcal{D}^*(A)} \operatorname{tr} X \sigma_A^{\frac{1}{p'}} \right)^{p'} & \text{if } p > 1 \\ \left(\inf_{\sigma_A \in \mathcal{D}^*(A)} \operatorname{tr} X \sigma_A^{\frac{1}{p'}} \right)^{p'} & \text{if } p < 1 \end{cases}. \tag{B.7}$$

Note that when $p > 1 \implies p' > 1$ and otherwise when $p < 1 \implies p' < 0$. Hence when we take the optimisation outside the exponent we obtain Eq. (B.6). The other equalities are evident from the definition of the Schatten norm (see Section 1.2.3). $\square$

It is prudent to explicitly state the available choices we have for the parameters when performing the interpolation in Theorems 3.1, 3.6 and related results. This is especially important when we are constrained to a particular direction of the given inequality.

Lemma B.3. *If $\alpha, \beta \geq \frac{1}{2}, \gamma \in \mathbb{R}$ and are related by*

$$\frac{\alpha}{\alpha - 1} = \frac{\beta}{\beta - 1} + \frac{\gamma}{\gamma - 1} \tag{B.8}$$

and assuming, without loss of generality, that $\beta \geq \gamma$, then the following are true and cover all possible cases up to symmetry:

If $0 < \dfrac{\beta'}{\alpha'} < 1$ then either

    Case 1. $\alpha, \beta, \gamma > 1, \quad \alpha < \gamma \leq \beta \quad and \quad \alpha \in (1, 2), \beta, \gamma \in (1, \infty)$ (B.9)

or

    Case 2. $\alpha, \beta, \gamma < 1, \quad \gamma \leq \beta < \alpha \quad and \quad \alpha, \beta \in [1/2, 1), \gamma \in [0, 1)$ (B.10)

or

    Case 3. $\alpha, \beta > 1, \gamma \leq 0, \quad \gamma < \alpha \leq \beta \quad and \quad \alpha, \beta \in (1, \infty), \gamma \in (-\infty, 0].$ (B.11)

If $0 < \dfrac{\alpha'}{\gamma'} < 1$ then

    Case 4. $\alpha, \gamma < 1, \beta > 1, \quad \alpha < \gamma < \beta \quad and \quad \alpha \in [1/2, 1), \gamma \in (2/3, 1), \beta \in (1, \infty).$ (B.12)

If $0 < \dfrac{\alpha'}{\beta'} < 1$ then

    Case 5. $\alpha, \beta > 1, \gamma < 1, \quad \gamma < \beta < \alpha \quad and \quad \alpha, \beta \in (1, \infty), \gamma \in [0, 1)$ (B.13)

or

    Case 6. $\alpha, \beta < 1, \gamma \leq 0, \quad \gamma < \alpha \leq \beta \quad and \quad \alpha, \beta \in [1/2, 1), \gamma \in (-\infty, 0].$ (B.14)

Proof. We first consider the cases where $\gamma > 0$ then move on to $\gamma \leq 0$.

Given $\gamma > 0$, we investigate the possible cases or, more specifically, the cases missing from the lemma. Given three independent binary options there are 8 possible permutations. Of the four that are missing the following: $(\alpha, \gamma > 1, \beta < 1)$ and $(\alpha, \beta < 1, \gamma > 1)$, contradict the assumption that $\beta > \gamma$. The remaining two: $(\alpha > 1, \beta, \gamma < 1)$ and $(\alpha < 1, \beta, \gamma > 1)$, never satisfy Eq. (B.8). We can now explore the implications of each of the assumptions.

Consider $0 < \frac{\beta'}{\alpha'} < 1$. It is evident that $(\alpha - 1)(\beta - 1) > 0$, a condition which now excludes **Case 4**. However, we can examine the two situations where this condition is satisfied:

$$0 < \frac{\beta'}{\alpha'} < 1 \implies \begin{cases} \alpha < \beta & \text{if} \quad \alpha, \beta > 1 \\ \alpha > \beta & \text{if} \quad \alpha, \beta < 1 \end{cases}. \tag{B.15}$$

It is clear that **Case 5** does not satisfy these implications but that **Cases 1** and **2** do depending on the sign of $\alpha - 1$.

For **Case 1**, we can calculate that $\lim\limits_{\eta \to 1^+} \dfrac{\eta}{\eta - 1} = \infty$ and $\lim\limits_{\eta \to \infty} \dfrac{\eta}{\eta - 1} = 1$.

Since α, β and γ are related by Eq. (B.8), we have

$$\alpha \longrightarrow 1 \implies \beta, \gamma \longrightarrow 1 \quad \text{and} \tag{B.16}$$

$$\alpha \longrightarrow 2 \implies \beta, \gamma \longrightarrow \infty, \tag{B.17}$$

i.e. $1 < \alpha < 2$ and $1 < \beta, \gamma < \infty$.

Moreover, for **Case 2**, another simple calculation shows that $\max\limits_{1/2 \leq \eta < 1} \dfrac{\eta}{\eta - 1} = -1$ and $\lim\limits_{\eta \to 1^-} \dfrac{\eta}{\eta - 1} = -\infty$. Hence, $\alpha = \frac{2}{3} \implies \beta, \gamma = \frac{1}{2}$, i.e. $\frac{2}{3} \leq \alpha < 1$ and $\frac{1}{2} \leq \beta, \gamma < 1$.

If instead $0 < \frac{\alpha'}{\beta'} < 1$, we still have the condition $(\alpha - 1)(\beta - 1) > 0$ but in the second part of the argument the inequalities are reversed, i.e

$$0 < \frac{\alpha'}{\beta'} < 1 \implies \begin{cases} \alpha > \beta & \text{if} \quad \alpha, \beta > 1 \\ \alpha < \beta & \text{if} \quad \alpha, \beta < 1 \end{cases}. \tag{B.18}$$

This overall excludes **Cases 1**, **2** and **4** but satisfies **Case 5**.

In this situation we again have $\alpha \to 1 \implies \beta, \gamma \to 1$ and for fixed γ we can write $\lim\limits_{\alpha \to \infty} \beta = \dfrac{1}{\gamma}$. Given that $\gamma > 0$, this implies $1 < \beta < \infty$.

Lastly, we have $0 < \frac{\alpha'}{\gamma'} < 1$, which implies $(\alpha - 1)(\gamma - 1) > 0$, excluding **Case 5** Similarly, we have following situations:

$$0 < \frac{\alpha'}{\gamma'} < 1 \implies \begin{cases} \alpha > \gamma & \text{if} \quad \alpha, \gamma > 1 \\ \alpha < \gamma & \text{if} \quad \alpha, \gamma < 1 \end{cases}, \tag{B.19}$$

which exclude **Cases 1** and **2**. So **Case 4** is the only remaining case which is satisfied.

We again have $\alpha \to 1 \implies \beta, \gamma \to 1$ and for fixed β, $\lim\limits_{\alpha \to 1/2} \gamma = \dfrac{2\beta - 1}{3\beta - 2}$ and $\lim\limits_{\beta \to \infty} \dfrac{2\beta - 1}{3\beta - 2} = \dfrac{2}{3}$. Hence $2/3 < \gamma < 1$ and $1 < \beta < \infty$.

We now treat the cases where $\gamma \leq 0$. Firstly, we eliminate the possibility that $\gamma \leq 0$ when $\frac{\alpha'}{\gamma'} \in (0, 1)$. For this to be satisfied we require α' and γ' to have the same sign. Since $\gamma \leq 0 \implies \gamma' \in [0, 1)$ we require $\alpha > 1 \implies \alpha' > 1$. The second requirement is $\alpha' \leq \gamma'$, however this is never satisfied given the valid ranges of α' and γ'.

We now consider **Case 3**. When $\frac{\beta'}{\alpha'} \in (0, 1)$ we have $\frac{\gamma'}{\alpha'} \in (0, 1)$ so, as above, we require the same sign but the second requirement, $\gamma' \leq \alpha'$, is always satisfied. This condition also implies that $\beta > 1$ in order to satisfy Eq. (B.8). The valid ranges are then evident as we have no extra restriction on γ and β, and when $\gamma = 0 \implies \alpha = \beta$.

Finally, for **Case 6**, we have $\frac{\alpha'}{\beta'} \in (0, 1)$ and require α' and β' to have the same sign and $|\alpha'| \leq |\beta'|$. If $\alpha, \beta > 1$, Eq. (B.8) determines that $\alpha' \geq \beta'$ so this case cannot be used. Otherwise, when $\alpha, \beta < 1$ we instead have $\alpha' \geq \beta' \implies |\alpha'| \leq |\beta'|$. The ranges can then be similarly determined. $\qquad\square$

We may use the following conditions to summarise the above cases.

Corollary B.4. *Given the assumptions in Lemma B.3 we have that*

$$\alpha < \gamma < \beta \implies (\alpha - 1)(\beta - 1)(\gamma^2 - \gamma) > 0 \quad \text{and} \tag{B.20}$$

$$\gamma < \beta < \alpha \implies (\alpha - 1)(\beta - 1)(\gamma^2 - \gamma) < 0. \tag{B.21}$$

Proof. This is evident from examining each case of Lemma B.3. $\qquad\square$

Corollary B.5. *Given the conditions in Lemma B.3 we have*

$$\frac{1}{\beta} + \frac{1}{\gamma} \leq 2 \quad \text{if} \quad \alpha \leq \beta, \tag{B.22}$$

$$\frac{1}{\beta} + \frac{1}{\gamma} \geq 2 \quad \text{if} \quad \alpha \geq \beta. \tag{B.23}$$

Proof. Assume without loss of generality that $\frac{\alpha'}{\beta'} \geq 0$, then

$$\frac{\gamma' + \beta'}{\beta'} = \frac{\gamma'}{\beta'} + 1 \geq 0 \tag{B.24}$$

$$\implies \frac{1}{\beta'} + \frac{1}{\gamma'} \begin{cases} \geq 0 & \text{if } \gamma' \geq 0 \\ \leq 0 & \text{if } \gamma' \leq 0 \end{cases} \tag{B.25}$$

$$\implies \frac{1}{\beta} + \frac{1}{\gamma} \begin{cases} \leq 2 & \text{if } \gamma' \geq 0 \\ \geq 2 & \text{if } \gamma' \leq 0 \end{cases}. \tag{B.26}$$

We can perform a symmetric argument where $\frac{\alpha'}{\gamma'} \geq 0$. Comparing these with the cases in Lemma B.3 we have our comparisons. $\qquad \square$

Bibliography

[1] L. Boltzmann, "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen," in *Sitzungsberichte der Akademie der Wissenschaften zu Wien*, vol. 66, 1872, pp. 275–370.

[2] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[3] N. Datta, C. Hirche, and A. Winter, "Convexity and operational interpretation of the quantum information bottleneck function," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1157–1161.

[4] G. D. Tourassi, E. D. Frederick, M. K. Markey, and C. E. Floyd Jr., "Application of the mutual information criterion for feature selection in computer-aided diagnosis," *Medical Physics*, vol. 28, no. 12, pp. 2394–2402, 2001. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.1418724

[5] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, Mar. 2003. [Online]. Available: http://dl.acm.org/citation.cfm?id=944919.944968

[6] K. W. Church and P. Hanks, "Word association norms, mutual information, and lexicography," in *27th Annual Meeting of the Association for Computational Linguistics*. Vancouver, British Columbia, Canada: Association for Computational Linguistics, Jun. 1989, pp. 76–83. [Online]. Available: https://www.aclweb.org/anthology/P89-1010

[7] A. Rényi, "On measures of information and entropy," in *Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. Berkeley, California, USA: University of California Press, 1961, pp. 547–561.

[8] F. Dupuis, "The decoupling approach to quantum information theory," Ph.D. dissertation, Université de Montréal, Apr. 2009. [Online]. Available: http://arxiv.org/abs/1004.1641

[9] J. Watrous, *The Theory of Quantum Information*. Cambridge University Press, 2018. [Online]. Available: https://books.google.com.au/books?id=GRNSDwAAQBAJ

[10] M. M. Wilde, *Quantum Information Theory*. Cambridge University Press, Jun. 2013.

[11] A. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 269–273, Jan. 1998. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=651037

[12] B. Schumacher and M. Westmoreland, "Sending classical information via noisy quantum channels," *Physical Review A*, vol. 56, no. 1, pp. 131–138, Jul. 1997. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.56.131

[13] A. S. Holevo, "Bounds for the quantity of information transmitted by a quantum communication channel," *Problems of Information Transmission*, vol. 9, no. 3, pp. 177–183, 1973.

[14] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted classical capacity of noisy quantum channels," *Physical Review Letters*, vol. 83, no. 15, p. 3081–3084, Oct. 1999. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.83.3081

[15] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, Oct. 2002. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1035117

[16] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse Shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2926–2959, 2014.

[17] M. Berta, M. Christandl, and R. Renner, "The quantum reverse Shannon theorem based on one-shot information theory," *Communications in Mathematical Physics*, vol. 306, no. 3, pp. 579–615, Aug. 2011. [Online]. Available: http://www.springerlink.com/index/10.1007/s00220-011-1309-7

[18] F. G. S. L. Brandão and M. Horodecki, "An area law for entanglement from exponential decay of correlations," *Nature Physics*, vol. 9, no. 11, pp. 721–726, Sep. 2013. [Online]. Available: http://www.nature.com/doifinder/10.1038/nphys2747

[19] M. J. Hall, "Information exclusion principle for complementary observables," *Physical Review Letters*, vol. 74, no. 17, pp. 3307–3311, 1995. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/10058168

[20] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, "Entropic uncertainty relations and their applications," *Reviews of Modern Physics*, vol. 89, no. 1, p. 015002, Feb. 2017. [Online]. Available: http://link.aps.org/doi/10.1103/RevModPhys.89.015002

[21] H. Umegaki, "Conditional expectation in an operator algebra," *Kodai Math. Sem. Rep.*, vol. 14, pp. 59–85, 1962.

[22] M. Tomamichel, M. Berta, and M. Hayashi, "Relating different quantum generalizations of the conditional Rényi entropy," *Journal of Mathematical Physics*, vol. 55, no. 8, p. 082206, Aug. 2014. [Online]. Available: http://arxiv.org/abs/1311.3887http://scitation.aip.org/content/aip/journal/jmp/55/8/10.1063/1.4892761

[23] M. Iwamoto and J. Shikata, "Information theoretic security for encryption based on conditional Rényi entropies," in *Proc. ICITS 2013*, ser. Lecture Notes in Computer Science, C. Padró, Ed. Springer, 2014, vol. 8317, pp. 103–121. [Online]. Available: http://eprint.iacr.org/2013/440http://link.springer.com/10.1007/978-3-319-04268-8{_}7

[24] A. Teixeira, A. Matos, and L. Antunes, "Conditional Rényi entropies," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4273–4277, Jul. 2012. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6191351

[25] S. Verdú, "Alpha-mutual information," in *Proc. IEEE ITA 2015*, 2015. [Online]. Available: https://www.princeton.edu/{~}verdu/reprints/VERDU-ITA2015.pdf

[26] M. Tomamichel and M. Hayashi, "Operational interpretation of Rényi information measures via composite hypothesis testing against product and Markov distributions," *IEEE Transactions on Information Theory*, 2017.

[27] M. Hayashi and M. Tomamichel, "Correlation detection and an operational interpretation of the Rényi mutual information," *Journal of Mathematical Physics*, vol. 57, no. 10, p. 102201, 2016.

[28] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel, "On quantum Rényi entropies: A new generalization and some properties," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122203, Jun. 2013. [Online]. Available: http://link.aip.org/link/JMAPAQ/v54/i12/p122203/s1{&}Agg=doi

[29] 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*, vol. 331, no. 2, pp. 593–622, Jul. 2014. [Online]. Available: http://arxiv.org/abs/1306.1586http://link.springer.com/10.1007/s00220-014-2122-x

[30] M. Mosonyi and F. Hiai, "On the quantum Rényi relative entropies and related capacity formulas," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2474–2487, Apr. 2011. [Online]. Available: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=5730573

[31] M. Mosonyi and T. Ogawa, "Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies," *Communications in Mathematical Physics*, vol. 334, no. 3, pp. 1617–1648, Mar. 2015. [Online]. Available: http://arxiv.org/abs/1309.3228http://link.springer.com/10.1007/s00220-014-2248-x

[32] ——, "Strong converse exponent for classical-quantum channel coding," *Communications in Mathematical Physics*, vol. 355, no. 1, pp. 373–426, Oct. 2017. [Online]. Available: http://arxiv.org/abs/1409.3562http://link.springer.com/10.1007/s00220-017-2928-4

[33] T. Cooney, M. Mosonyi, and M. M. Wilde, "Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication," *Communications in Mathematical Physics*, vol. 344, no. 3, pp. 797–829, Jun. 2016. [Online]. Available: https://doi.org/10.1007/s00220-016-2645-4

[34] M. Tomamichel, M. M. Wilde, and A. Winter, "Strong converse rates for quantum communication," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 715–727, 2017.

[35] W. Heisenberg, "Über den Anschaulichen Inhalt der Quantentheoretischen Kinematik und Mechanik," *Zeitschrift für Physik*, vol. 43, no. 3-4, pp. 172–198, Mar. 1927.

[36] H. Maassen and J. Uffink, "Generalized entropic uncertainty relations," *Physical Review Letters*, vol. 60, no. 12, pp. 1103–1106, Mar. 1988. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.60.1103

[37] F. Dupuis, "Chain rules for quantum Rényi entropies," *Journal of Mathematical Physics*, vol. 56, no. 2, p. 022203, 2015.

[38] S. Beigi, "Sandwiched Rényi divergence satisfies data processing inequality," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122202, Jun. 2013. [Online]. Available: http://arxiv.org/abs/1306.5920http://scitation.aip.org/content/aip/journal/jmp/54/12/10.1063/1.4838855

[39] G. Pisier, *Non-commutative vector valued L_p-spaces and completely p-summing map*, ser. Astérisque. Société mathématique de France, 1998, no. 247. [Online]. Available: http://www.numdam.org/item/AST_1998__247__R1_0

[40] I. Devetak, M. Junge, C. King, and M. B. Ruskai, "Multiplicativity of completely bounded p-norms implies a new additivity result," *Communications in Mathematical Physics*, vol. 266, pp. 37–63, 2006.

[41] P. Delgosha and S. Beigi, "Impossibility of local state transformation via hypercontractivity," *Communications in Mathematical Physics*, vol. 332, no. 1, pp. 449–476, Nov 2014. [Online]. Available: https://doi.org/10.1007/s00220-014-2105-y

[42] M. Tomamichel, *Quantum Information Processing with Finite Resources — Mathematical Foundations*, ser. SpringerBriefs in Mathematical Physics. Springer International Publishing, 2016, vol. 5. [Online]. Available: http://link.springer.com/10.1007/978-3-319-21891-5

[43] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, "The uncertainty principle in the presence of quantum memory," *Nature Physics*, vol. 6, p. 659, 2010. [Online]. Available: http://dx.doi.org/10.1038/nphys1734

[44] P. J. Coles and M. Piani, "Improved entropic uncertainty relations and information exclusion relations," *Physical Review A*, vol. 89, no. 2, p. 022112, Feb. 2014. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevA.89.022112

[45] R. Renner, "Security of quantum key distribution," *International Journal of Quantum Information*, vol. 06, no. 01, pp. 1–127, Feb. 2008. [Online]. Available: http://www.worldscientific.com/doi/abs/10.1142/S0219749908003256http://arxiv.org/abs/quant-ph/0512258

[46] M. Tomamichel, "A framework for non-asymptotic quantum information theory," Ph.D. dissertation, ETH Zurich, Mar. 2012. [Online]. Available: http://arxiv.org/abs/1203.2142

[47] M. Berta, K. Seshadreesan, and M. Wilde, "Rényi generalizations of the conditional quantum mutual information," *Journal of Mathematical Physics*, vol. 56, no. 2, p. 022205, 2015. [Online]. Available: http://arxiv.org/abs/1403.6102

[48] G. Golub and C. Van Loan, *Matrix Computations*, ser. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 2013. [Online]. Available: https://books.google.com.au/books?id=X5YfsuCWpxMC

[49] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information: 10th anniversary edition.* New York, NY, USA: Cambridge University Press, 2011.

[50] A. Rényi, "On measures of entropy and information," in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, ser. Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, 1961, Conference Proceedings, pp. 547–561. [Online]. Available: https://projecteuclid.org/euclid.bsmsp/1200512181

[51] T. Cover and J. Thomas, *Elements of information theory*, ser. Wiley series in telecommunications. Wiley, 1991. [Online]. Available: https://books.google.com.au/books?id=CX9QAAAAMAAJ

[52] O. Klein, "Zur quantenmechanischen Begründung des zweiten Hauptsatzes der Wärmelehre," *Zeitschrift für Physik*, vol. 72, no. 11, pp. 767–775, Nov 1931. [Online]. Available: https://doi.org/10.1007/BF01341997

[53] A. Bhattacharyya, "On a measure of divergence between two statistical populations defined by their probability distributions," *Bull. Calcutta math. Soc.*, vol. 35, pp. 99–109, 1943.

[54] M. Iwamoto and J. Shikata, "Revisiting conditional Rényi entropies and generalizing Shannon's bounds in information theoretically secure encryption," Cryptology ePrint Archive 440/2013, Tech. Rep., 2013.

[55] C. Cachin, "Entropy measures and unconditional security in cryptography," Ph.D. dissertation, ETH Zurich, 1997.

[56] P. Jizba and T. Arimitsu, "Generalized statistics: Yet another generalization," *Physica A: Statistical Mechanics and its Applications*, vol. 340, 09 2004.

[57] ——, "The world according to Rényi: Thermodynamics of fractal systems," *AIP Conference Proceedings*, vol. 597, no. 1, pp. 341–348, 2001. [Online]. Available: https://aip.scitation.org/doi/abs/10.1063/1.1427481

[58] R. Renner and S. Wolf, "Simple and tight bounds for information reconciliation and privacy amplification," in *Advances in Cryptology - ASIACRYPT 2005*, B. Roy, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 199–216.

[59] D. Petz, "Quasi-entropies for finite quantum systems," *Reports on Mathematical Physics*, vol. 23, no. 1, pp. 57–65, Feb. 1986. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/0034487786900674

[60] R. L. Frank and E. H. Lieb, "Monotonicity of a relative Rényi entropy," *Journal of Mathematical Physics*, vol. 54, no. 12, p. 122201, Dec. 2013. [Online]. Available: http://dx.doi.org/10.1063/1.4838835

[61] H. P. Robertson, "The uncertainty principle," *Physical Review*, vol. 34, no. 1, pp. 163–164, 1929.

[62] E. H. Kennard, "Zur Quantenmechanik einfacher Bewegungstypen," *Zeitschrift für Physik*, vol. 44, no. 4-5, pp. 326–352, Apr. 1927. [Online]. Available: http://link.springer.com/10.1007/BF01391200

[63] H. Weyl, *Gruppentheorie und Quantenmechanik.* Leipzig: Hirzel, 1928.

[64] A. Grudka, M. Horodecki, P. Horodecki, R. Horodecki, W. Kłobus, and L. Pankowski, "Conjectured strong complementary-correlations tradeoff," *Physical Review A*, vol. 88, no. 3, p. 032106, 2013. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.88.032106

[65] E. Kreyszig, *Introductory Functional Analysis with Applications*, ser. Wiley Classics Library. Wiley, 1989. [Online]. Available: https://books.google.com.au/books?id=nZmpQgAACAAJ

[66] M. Reed and B. Simon, *II: Fourier Analysis, Self-Adjointness*, ser. Methods of Modern Mathematical Physics. Elsevier Science, 1975. [Online]. Available: https://books.google.com.au/books?id=Kz7s7bgVe8gC

[67] W. Stinespring, "Positive functions on C^*-algebras," *Proceedings of the American Mathematical Society*, vol. 6, 04 1955.

[68] J. v. Neumann, "Zur Theorie der Gesellschaftsspiele," *Mathematische Annalen*, vol. 100, no. 1, pp. 295–320, Dec. 1928. [Online]. Available: https://doi.org/10.1007/BF01448847

[69] A. Vitanov, F. Dupuis, M. Tomamichel, and R. Renner, "Chain rules for smooth min- and max-entropies," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2603–2612, May 2013. [Online]. Available: http://arxiv.org/abs/1205.5231http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6408179

[70] N. Ciganovic, N. J. Beaudry, and R. Renner, "Smooth max-information as one-shot generalization for mutual information," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1573–1581, Mar. 2014. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6687244

[71] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, Jul. 1988. [Online]. Available: http://link.springer.com/10.1007/BF01016429

[72] A. McKinlay and M. Tomamichel, "Decomposition rules for quantum Rényi mutual information with an application to information exclusion relations," *Journal of Mathematical Physics*, vol. 61, no. 7, p. 072202, 2020. [Online]. Available: https://doi.org/10.1063/1.5143862