

# On contraction coefficients, partial orders and approximation of capacities for quantum channels

Christoph Hirche<sup>1</sup>, Cambyse Rouzé<sup>2</sup>, and Daniel Stilck França<sup>1</sup>

<sup>1</sup>QMATH, Department of Mathematical Sciences, University of Copenhagen,  
Universitetsparken 5, 2100 Copenhagen, Denmark

<sup>2</sup>Department of Mathematics, Technische Universität München, 85748 Garching, Germany

<sup>2</sup>Munich Center for Quantum Science and Technology (MCQST), München, Germany

## Abstract

The data processing inequality is the most basic requirement for any meaningful measure of information. It essentially states that distinguishability measures between states decrease if we apply a quantum channel. It is the centerpiece of many results in information theory and justifies the operational interpretation of most entropic quantities. In this work, we revisit the notion of contraction coefficients of quantum channels, which provide sharper and specialized versions of the data processing inequality. A concept closely related to data processing are partial orders on quantum channels. We discuss several quantum extensions of the well known less noisy ordering and then relate them to contraction coefficients. We further define approximate versions of the partial orders and show how they can give strengthened and conceptually simple proofs of several results on approximating capacities. Moreover, we investigate the relation to other partial orders in the literature and their properties, particularly with regards to tensorization. We then investigate further properties of contraction coefficients and their relation to other properties of quantum channels, such as hypercontractivity. Next, we extend the framework of contraction coefficients to general  $f$ -divergences and prove several structural results. Finally, we consider two important classes of quantum channels, namely Weyl-covariant and bosonic Gaussian channels. For those, we determine new contraction coefficients and relations for various partial orders.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Contraction coefficients and partial orders of quantum channels</b>	<b>3</b>
2.1	Channel preorders in quantum information theory, and related tasks . . . . .	4
2.2	Contraction coefficients and the strong data processing inequality . . . . .	7
<b>3</b>	<b>Properties and Bounds</b>	<b>9</b>
3.1	Tensorization of partial orders . . . . .	9
3.2	Bounds on contraction coefficients . . . . .	11
3.3	Additional examples . . . . .	14
<b>4</b>	<b>Approximate partial orders and capacities</b>	<b>15</b>
4.1	Approximate partial orders . . . . .	16
4.2	Approximate capacities of a quantum channel . . . . .	18
<b>5</b>	<b>Characterizations of Strong data processing</b>	<b>20</b>

5.1	SDPI via hypercontractivity . . . . .	20
<b>6</b>	<b>Generalized contraction coefficients and partial orders</b>	<b>25</b>
6.1	f-divergences . . . . .	25
6.2	Spectral approaches to quantum less-noisy pre-orders . . . . .	28
<b>7</b>	<b>Functional inequalities</b>	<b>30</b>
<b>8</b>	<b>Special classes of channels</b>	<b>33</b>
8.1	Weyl-covariant channels . . . . .	33
8.2	Gaussian channels . . . . .	36
<b>9</b>	<b>Conclusions</b>	<b>42</b>

# 1 Introduction

One of the arguably most fundamental concepts in (quantum) information theory is that of data processing<sup>1</sup>. Many relevant quantities are monotone under the application of a quantum channel. That is what allows us to assign them an operational meaning in terms of distinguishability and, in turn, makes them useful in assessing physical properties. A commonly considered quantity is the relative entropy for which, when considering an arbitrary channel  $\mathcal{N}$ , data processing manifests as

$$D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \leq D(\rho\|\sigma). \quad (1.1)$$

This mathematical statement also gives rise to an operational interpretation of data processing: The relative entropy gives the optimal rate at which one can asymptotically discriminate two quantum states in an asymmetric setting [33, 62]. It is now immediately apparent that applying a quantum channel to the state can never make the discrimination task easier, therefore the relative entropy has to become smaller.

Since the relative entropy acts as a parent quantity for several other entropic quantities, they also inherit its data processing property. An important example is the mutual information: Given a bipartite quantum state  $\rho_{AA'}$  and a quantum channel  $\mathcal{N}: A' \rightarrow B$ , we always have

$$I(A : B) \leq I(A : A'). \quad (1.2)$$

Although data processing is a potent tool, often it is not sufficient to know that a quantity is decreasing when applying a channel, one also needs to quantify by how much. When fixing the reference state  $\sigma$  and varying  $\rho$ , in many cases it is possible to show that  $D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))$  is strictly smaller than  $D(\rho\|\sigma)$  unless  $\rho = \sigma$ . A natural approach is then to consider contraction coefficients,

$$\eta_{\text{Re}}(\mathcal{N}, \sigma) := \sup_{\rho \neq \sigma} \frac{D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma))}{D(\rho\|\sigma)}, \quad (1.3)$$

and we say that  $\mathcal{N}$  satisfies a *strong data processing inequality* at input state  $\sigma$  when  $\eta_{\text{Re}}(\mathcal{N}, \sigma) < 1$ . These contraction coefficients have already found many applications in information theory [67, 70] and were studied in the quantum setting in [46, 34].

Another closely related question is to compare channels based on how much they contract a certain quantity. In this case, comparing the contraction coefficients is only of limited use. A more direct approach is to define partial orders on the set of channels based on how much the channels contract each possible input state individually. While many such orders are known, the most commonly used ones are the less noisy and the more capable order [42], especially in the classical setting.

---

<sup>1</sup>“If you don’t have data processing, you’re toast”, Mark M. Wilde, *ISL Colloquium*, 17/09/2020

The contraction coefficient (1.3) has been extensively studied in the classical setting. In their pioneering work, Ahlswede and Gács [1] discovered a deep relationship between  $\eta_{\text{Re}}(\mathcal{N}, \sigma)$  and several other quantities, among which the maximal correlation and hypercontractive properties of the channel. After this, the notion of entropic contraction coefficients were extended to general  $f$ -divergences and other relevant measures of distinguishability between two probability distributions such as the Dobrushin contraction coefficient (for the trace distance), or that of the  $\chi^2$ -divergence. [18, 15, 55, 17, 28] (see also the more recent contributions [69, 67, 66, 50]).

In the quantum setting, Watanabe introduced a quantum generalization of the less noisy and more capable orders [78]. A noteworthy difference to the classical case is that the definitions in [78] include a regularization, in the sense that the property has to hold not just for  $\mathcal{N}$  but  $\mathcal{N}^{\otimes n}$  for all  $n$ . This choice proved to be useful in the context of capacities and was therefore operationally motivated. However, in many cases it can also be useful to look at the  $n = 1$  case, see e.g. [75]. We go a step further and also consider two additional quantum generalizations: a third version with additional reference system that turns out to have some desirable properties and a fourth where all systems are fully quantum.

In this work we revisit the concepts of contraction coefficients and partial orders in quantum information theory. Most importantly we show how both contraction coefficients and the less noisy order can be understood in very similar terms as either relative entropy or mutual information based tools. From this starting point, we discuss properties such as tensorization, bounds and the relation to other partial orders in the spirit of [50, 69].

We then discuss the application of partial orders in the context of quantum capacities, revisiting the recently introduced concept of approximate degradability and the resulting capacity bounds [73] and introducing approximate versions of the different less noisy and more capable definitions. This leads to weaker requirements for approximating the quantum, private and classical capacities. Moreover, as corollaries we get the approximate degradability capacity bounds via conceptually very simple proofs.

The main drawback of the entropy based partial orders discussed here is that compared to (approximate) degradability, they are rather complicated to verify. It can be very hard to check whether two quantum channels fulfill the conditions of the partial order of choice. To this end, we investigate alternative characterizations and conditions, providing some progress towards this question. We also draw connections to the notions of hypercontractivity and other functional inequalities.

Finally, we discuss both generalizations and special cases of the discussed concepts. First, we investigate generalizations of contraction coefficients and the less noisy order to  $f$ -divergences. Then we present several results for the setting where the channels are from the special classes of either Weyl-covariant channels and bosonic Gaussian channels. In particular we find examples of Weyl-covariant channels that satisfy the less noisy ordering without one being the degraded version of the other. We end this paper with a derivation of precise contraction coefficients for the most important classes of quantum Gaussian channels, namely the attenuator, the amplifier and the additive noise channels. We also consider the case of tensor products of the latter, under the constrained minimum output entropy conjecture.

**Some Notations** : In what follows, we denote by  $\mathcal{H}$  a finite dimensional Hilbert space.  $\mathcal{B}(\mathcal{H})$  is the set of linear operators acting on  $\mathcal{H}$ . The set of quantum states is denoted by  $\mathcal{D}(\mathcal{H})$ , and that of full-rank states by  $\mathcal{D}(\mathcal{H})_+$ . Given two quantum systems  $\mathcal{H}, \mathcal{K}$ , the set of quantum channels  $\mathcal{N} : \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{K})$ , that is of completely positive, trace preserving superoperators, is denoted by  $\text{CPTP}(\mathcal{H}, \mathcal{K})$ . The set of probability mass functions over sets of cardinality  $m \in \mathbb{N}$  is denoted by  $\mathcal{P}_m$ . We also use standard quantum information conventions, denoting quantum systems by capital letters  $A, B$  and classical ones by  $U$  and  $X$ . For other standard definitions such as most entropic quantities and channel capacities we refer to [76, 82].

## 2 Contraction coefficients and partial orders of quantum channels

In this section we will first review the main partial orders used throughout the manuscript: Degradable, less noisy and more capable. In the case of the latter two, we introduce several variants that are all valid generalizations of the corresponding classical formulations. Next, we review contraction coefficients and in particular, the one based on the relative entropy for which we then show an alternative formulation in terms

of mutual information. This provides us with a theme throughout the rest of the manuscript. Finally, we briefly discuss some connections between contraction coefficients and partial orders.

## 2.1 Channel preorders in quantum information theory, and related tasks

Channel preorders can be understood as ways of quantifying a channel's noisiness as compared to another one. However, the notion of noise varies according to the tasks for which a given channel is being used. For this reason, a zoo of preorders can be found already in the classical setting [50]. We recall some of the main preorders one can find in the literature and recall their relations here. We start with the most common partial order: Degradability.

**Degradation preorder:** [19, 6] A channel  $\mathcal{N} \in \text{CPTP}(\mathcal{H}, \mathcal{K}')$  is said to be a *degraded* version of a channel  $\mathcal{M} \in \text{CPTP}(\mathcal{H}, \mathcal{K})$  with same input space, denoted  $\mathcal{M} \succeq_{\text{deg}} \mathcal{N}$  if the following equivalent conditions hold:

- The channels are related as

$$\mathcal{N} = \Theta \circ \mathcal{M} \quad (2.1)$$

for some channel  $\Theta \in \text{CPTP}(\mathcal{K}, \mathcal{K}')$ .

- For any  $c - q$  state  $\rho_{UAR}$ ,

$$H_{\min}(U|BR)_{(\text{id}_{UR} \otimes \mathcal{M})(\rho_{UAR})} \leq H_{\min}(U|B'R)_{(\text{id}_{UR} \otimes \mathcal{N})(\rho_{UAR})}. \quad (2.2)$$

The first condition is the well-known standard definition of degradability, while the equivalence with the second was recently shown in [9]. We recall that the min conditional entropy is defined as

$$H_{\min}(A|B)_{\rho} = -\inf_{\sigma_B} \inf \{ \lambda \in \mathbb{R} : \rho_{AB} \leq 2^{\lambda} I_A \otimes \sigma_B \}. \quad (2.3)$$

While the first condition is often more useful due to data processing the second will make the close familiarity to the remaining partial orders easily evident.

We will continue by defining what can be considered the basic version of the less noisy and more capable partial orders.

**Less noisy, more capable:** [42, 78] A channel  $\mathcal{M} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$  is said to be *less noisy* than a channel  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_{B'})$  with same input space, denoted  $\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}$ , if for all classical random variable  $U$  and any  $c - q$  state  $\rho_{UA}$

$$I(U : B)_{(\text{id}_U \otimes \mathcal{M})(\rho_{UA})} \geq I(U : B')_{(\text{id}_U \otimes \mathcal{N})(\rho_{UA})}.$$

Moreover,  $\mathcal{M}$  is said to be *more capable* than  $\mathcal{N}$ , denoted  $\mathcal{M} \succeq_{\text{m.c.}} \mathcal{N}$ , if for any ensemble  $\{p, |\psi_x\rangle_A\}$ , and the corresponding  $c - q$  state  $\rho_{XA} = \sum_x p(x) |x\rangle\langle x| \otimes |\psi_x\rangle\langle \psi_x|$

$$I(X : B)_{(\text{id}_X \otimes \mathcal{M})(\rho_{XA})} \geq I(X : B')_{(\text{id}_X \otimes \mathcal{N})(\rho_{XA})}.$$

We recall that the mutual information  $I(A : B)_{\rho}$  of a bipartite state  $\rho_{AB}$  is defined as

$$I(A : B)_{\rho} := H(A)_{\rho} + H(B)_{\rho} - H(AB)_{\rho},$$

where  $H(C)_{\rho} := H(\rho_C) = -\text{Tr}[\rho_C \ln \rho_C]$  denotes the von Neumann entropy of a state  $\rho$  on a subsystem  $C$ . In the classical setting, these definitions are well motivated by their applications to wiretap and broadcast channels [42, 20, 53]. In the quantum setting, the less noisy and more capable relation determine properties and relations of the quantum and private capacities [78]. Furthermore, it is immediately clear that

$$\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N} \quad \Rightarrow \quad \chi(\mathcal{M}) \geq \chi(\mathcal{N}), \quad (2.4)$$

where  $\chi(\mathcal{N})$  is the Holevo quantity of the channel  $\mathcal{N}$ .

Note that we stated the condition here in its most common form via the mutual information. It can however easily be written in terms of the conditional entropy as

$$H(U|B)_{(\text{id}_U \otimes \mathcal{M})(\rho_{UA})} \leq H(U|B')_{(\text{id}_U \otimes \mathcal{N})(\rho_{UA})}, \quad (2.5)$$

and similarly for the more capable relation, which bears noticeable similarity with the alternative definition of degradability. The lack of a reference system here is crucial and we will discuss this in more details later. Indeed, one can also find a different partial order in the literature called *less ambiguous* that is defined as a mix of the above characteristics: min-entropy, but no reference system (see e.g. [10]).

$$H_{\min}(U|B)_{(\text{id}_U \otimes \mathcal{M})(\rho_{UA})} \leq H_{\min}(U|B')_{(\text{id}_U \otimes \mathcal{N})(\rho_{UA})}. \quad (2.6)$$

As expected in the quantum setting, the notions of less noisy and more capable need in general to be regularized in order to lead to an operational interpretation [78]<sup>2</sup>:

**Regularized less noisy, more capable:** With the previous notations, the channel  $\mathcal{M}$  is said to be *regularized less noisy* than  $\mathcal{N}$ , denoted  $\mathcal{M} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{N}$  if for any  $n \in \mathbb{N}$ , any classical register  $U$  and any  $c - q$  state  $\rho_{UA^n}$ ,

$$I(U : B^n)_{(\text{id}_U \otimes \mathcal{M}^{\otimes n})(\rho_{UA^n})} \geq I(U : B'^n)_{(\text{id}_U \otimes \mathcal{N}^{\otimes n})(\rho_{UA^n})}.$$

Moreover,  $\mathcal{M}$  is *regularized more capable* than  $\mathcal{N}$ , denoted  $\mathcal{M} \succeq_{\text{m.c.}}^{\text{reg}} \mathcal{N}$ , if for any  $n \in \mathbb{N}$ , and ensemble  $\{p, |\psi_x\rangle_{A^n}\}$  over  $\mathcal{H}_A^{\otimes n}$ , and the corresponding c-q state  $\rho_{XA^n} = \sum_x p(x) |x\rangle\langle x| \otimes |\psi_x\rangle\langle\psi_x|_{A^n}$

$$I(X : B^n)_{(\text{id}_X \otimes \mathcal{M}^{\otimes n})(\rho_{XA^n})} \geq I(X : B'^n)_{(\text{id}_X \otimes \mathcal{N}^{\otimes n})(\rho_{XA^n})}.$$

In general the regularized version of the partial orders give a strictly stronger condition than the unregularized one. We will discuss the problem of tensorization in more details in Section 3. Similar to the standard definition we can also here directly observe that

$$\mathcal{M} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{N} \quad \Rightarrow \quad C(\mathcal{M}) \geq C(\mathcal{N}), \quad (2.7)$$

where  $C(\mathcal{N})$  is the classical capacity of the channel  $\mathcal{N}$ . An alternative quantum generalization is to provide the channel input with an additional quantum reference system. We will later see that this also leads to an order that obeys additional desirable properties such as tensorization.

**Completely less noisy, more capable** With the previous notations, the channel  $\mathcal{M}$  is said to be *completely less noisy* than  $\mathcal{N}$ , denoted  $\mathcal{M} \succeq_{\text{l.n.}}^c \mathcal{N}$  if for any classical register  $U$  and any  $c - q$  state  $\rho_{UAR}$ ,

$$I(U : BR)_{(\text{id}_U \otimes \mathcal{M})(\rho_{UAR})} \geq I(U : B'R)_{(\text{id}_U \otimes \mathcal{N})(\rho_{UAR})}.$$

Moreover,  $\mathcal{M}$  is *completely more capable* than  $\mathcal{N}$ , denoted  $\mathcal{M} \succeq_{\text{m.c.}}^c \mathcal{N}$ , if for any ensemble  $\{p, |\psi_x\rangle_{AR}\}$  over  $\mathcal{H}_{AR}$ , and the corresponding c-q state  $\rho_{XAR} = \sum_x p(x) |x\rangle\langle x| \otimes |\psi_x\rangle\langle\psi_x|$ ,

$$I(X : BR)_{(\text{id}_X \otimes \mathcal{M})(\rho_{XAR})} \geq I(X : B'R)_{(\text{id}_X \otimes \mathcal{N})(\rho_{XAR})}.$$

Note that there is no known restriction on the dimension of the  $R$  system, making this condition generally even harder to check than the standard variants. This is a common problem in quantum information theory, see e.g. [5, 36, 16, 22, 84].

At this point it should be briefly noticed that, in the classical less noisy setting, allowing a classical reference system does not change the order itself. Even more, one can easily see that even in the quantum setting a classical reference system would not help.

<sup>2</sup>Strictly speaking, Watanabe defined the notion of regularized less noisy and more capable between a channel and its complementary, which he simply referred to as less noisy and more capable.

**Lemma 2.1.** *If one restricts the reference system in the completely less noisy ordering to be classical the order would become equivalent to the usual less noisy ordering, meaning*

$$\mathcal{M} \succeq_{\text{l.n.}}^c \mathcal{N} \Leftrightarrow \mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}. \quad (2.8)$$

*Proof.* The “ $\Rightarrow$ ” direction is obvious. The “ $\Leftarrow$ ” direction follows by noticing that

$$\begin{aligned} \mathcal{M} \succeq_{\text{l.n.}}^c \mathcal{N} &\Leftrightarrow I(U : B|R)_{(\text{id}_{UR} \otimes \mathcal{M})(\rho_{UAR})} \geq I(U : B'|R)_{(\text{id}_{UR} \otimes \mathcal{N})(\rho_{UAR})} \\ &\Leftrightarrow \sum_r p(r) I(U : B|R=r)_{(\text{id}_{UR} \otimes \mathcal{M})(\rho_{UAR})} \geq \sum_r p(r) I(U : B'|R=r)_{(\text{id}_{UR} \otimes \mathcal{N})(\rho_{UAR})}. \end{aligned}$$

The first equivalence follows by subtracting  $I(U : R)_\rho$  on both sides and the second because  $R$  is assumed to be classical. Now the proof follows by the definition of the less noisy ordering.  $\square$

As quantum conditioning is significantly more complex, the same argument does not hold for quantum reference systems [5, 35].

All previously discussed definitions of quantum extensions of the classical less noisy order in this work have in common that the system correlated to the input is classical. However, one could envision a *fully quantum* extension of the less noisy order where this restriction is lifted.

**Fully quantum less noisy, more capable** A channel  $\mathcal{M} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$  is said to be *fully quantum less noisy* than a channel  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_{B'})$  with same input space, denoted  $\mathcal{M} \succeq_{\text{l.n.}}^{\text{fq}} \mathcal{N}$ , if for all quantum states  $\rho_{AA'}$

$$I(A : B)_{(\text{id}_A \otimes \mathcal{M})(\rho_{AA'})} \geq I(A : B')_{(\text{id}_A \otimes \mathcal{N})(\rho_{AA'})}.$$

Moreover,  $\mathcal{N}$  is said to be *fully quantum more capable* than  $\mathcal{M}$ , denoted  $\mathcal{N} \succeq_{\text{m.c.}}^{\text{fq}} \mathcal{M}$ , if for all pure quantum states  $\Psi_{AA'}$

$$I(A : B)_{(\text{id}_A \otimes \mathcal{M})(\Psi_{AA'})} \geq I(A : B')_{(\text{id}_A \otimes \mathcal{N})(\Psi_{AA'})}.$$

Similar to before we can also here directly observe that

$$\mathcal{M} \succeq_{\text{m.c.}}^{\text{fq}} \mathcal{N} \Rightarrow C_E(\mathcal{M}) \geq C_E(\mathcal{N}), \quad (2.9)$$

where  $C_E(\mathcal{N})$  is the entanglement assisted classical capacity of the channel  $\mathcal{N}$ . In line with the previous discussion, this partial order can also be compared to the *more coherent order* discussed e.g. in [10].

The relations between the different partial orders are summarized in Figure 1 and we defer their proofs to Section 3.

As customary we also define the anti-orders, i.e. anti-degradable, anti-less noisy and so on, as above but with the roles of  $\mathcal{N}$  and  $\mathcal{M}$  interchanged. This will become relevant when we discuss capacities where we often consider  $\mathcal{M}$  to be the complementary channel of  $\mathcal{N}$ .

Before ending this section we state some elementary facts about the channels dominated by a fixed channel  $\mathcal{M}$ . To that end, given a channel  $\mathcal{M}$ , define the *less noisy domination region*

$$\mathcal{L}_{\mathcal{M}} := \{\mathcal{N} : \mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}\}$$

as well as the *degradation region* of  $\mathcal{M}$

$$\mathcal{D}_{\mathcal{M}} := \{\mathcal{N} : \mathcal{M} \succeq_{\text{deg}} \mathcal{N}\}.$$

**Proposition 2.2.** *Given the channel  $\mathcal{M}$ , its less noisy domination region  $\mathcal{L}_{\mathcal{M}}$  and its degradation region  $\mathcal{D}_{\mathcal{M}}$  are non-empty, closed, convex and for any channel  $\Lambda$*

$$\mathcal{N} \in \mathcal{L}_{\mathcal{M}} \Rightarrow \Lambda \circ \mathcal{N} \in \mathcal{L}_{\mathcal{M}} \quad \text{and} \quad \mathcal{N} \in \mathcal{D}_{\mathcal{M}} \Rightarrow \Lambda \circ \mathcal{N} \in \mathcal{D}_{\mathcal{M}}.$$

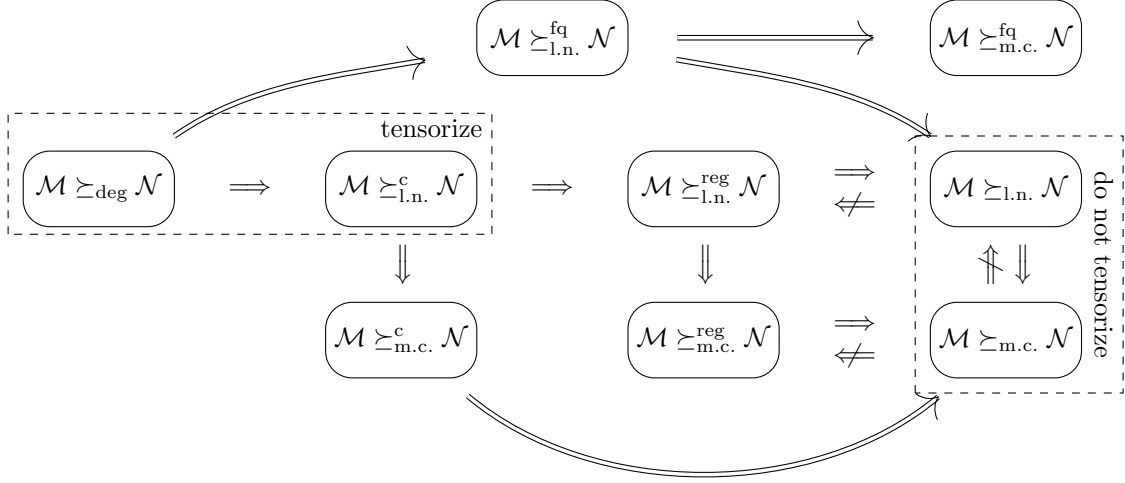


Figure 1: Summary of relations of the partial orders defined in Section 2.1.

*Proof.* Non-emptiness is obvious since  $\mathcal{M} \in \mathcal{D}_{\mathcal{M}}$  and  $\mathcal{M} \in \mathcal{L}_{\mathcal{M}}$ . The closure of the sets is proved as follows: given two states  $\rho, \sigma$  consider a sequence of channels  $\mathcal{N}_k \in \mathcal{L}_{\mathcal{N}}$  such that  $\mathcal{N}_k \rightarrow \mathcal{N}$  (say with respect to the  $2 \rightarrow 2$  norm). Then we also have that  $\mathcal{N}_k(\rho) \rightarrow \mathcal{N}(\rho)$  and  $\mathcal{N}_k(\sigma) \rightarrow \mathcal{N}(\sigma)$  (say in Hilbert-Schmidt norm). Therefore

$$D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma)) \leq \liminf_{k \rightarrow \infty} D(\mathcal{N}_k(\rho) \parallel \mathcal{N}_k(\sigma)) \leq D(\mathcal{M}(\rho) \parallel \mathcal{M}(\sigma)),$$

where the first inequality arises from lower semicontinuity of Umegaki's relative entropy, and the second one holds because  $\mathcal{N}_k \in \mathcal{L}_{\mathcal{M}}$  for all  $k$ . Therefore the set  $\mathcal{L}_{\mathcal{M}}(\rho, \sigma) := \{\mathcal{N} : D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma)) \leq D(\mathcal{M}(\rho) \parallel \mathcal{M}(\sigma))\}$  is closed. Since  $\mathcal{L}_{\mathcal{M}} = \bigcap_{\rho, \sigma} \mathcal{L}_{\mathcal{M}}(\rho, \sigma)$ ,  $\mathcal{L}_{\mathcal{M}}$  itself is closed as an intersection of closed sets. The closure of  $\mathcal{D}_{\mathcal{M}}$  follows similarly: consider a sequence of channels  $\mathcal{N}_k$  such that  $\mathcal{N}_k \rightarrow \mathcal{N}$ . Since for each  $k$ ,  $\mathcal{M}_k = \Lambda_k \circ \mathcal{M}$  for some  $\Lambda_k$  belonging to the compact set of quantum channels, there exists a subsequence  $\Lambda_{k_m}$  that converges by sequential compactness:  $\Lambda_{k_m} \rightarrow \Lambda$ . Hence  $\mathcal{N} \in \mathcal{D}_{\mathcal{M}}$  since  $\mathcal{N}_{k_m} = \Lambda_{k_m} \circ \mathcal{M} \rightarrow \Lambda \circ \mathcal{M}$ , proving closedness of  $\mathcal{D}_{\mathcal{M}}$ . The convexity of the sets is obvious from the convexity of the relative entropy for  $\mathcal{L}_{\mathcal{M}}$  and the linearity of the defining condition for degradability. Invariance under output channels is obvious by DPI for  $\mathcal{L}_{\mathcal{M}}$  and using the direct definition for  $\mathcal{D}_{\mathcal{M}}$ .  $\square$

## 2.2 Contraction coefficients and the strong data processing inequality

The central objects of this section are contraction coefficients and strong data processing constants. Given a fixed state  $\sigma \in \mathcal{D}(\mathcal{H}_A)$  and a quantum channel  $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \rightarrow \mathcal{B}(\mathcal{K}_B)$ , define the *relative entropy strong data processing inequality (SDPI) constants* [1, 34]: given  $\sigma \in \mathcal{H}_A$

$$\eta_{\text{Re}}(\mathcal{N}, \sigma) := \sup_{\substack{\rho \in \mathcal{D}(\mathcal{H}_A) \\ 0 < D(\rho \parallel \sigma) < \infty}} \frac{D(\mathcal{N}(\rho) \parallel \mathcal{N}(\sigma))}{D(\rho \parallel \sigma)}, \quad (2.10)$$

and the contraction coefficient is then  $\eta_{\text{Re}}(\mathcal{N}) := \sup_{\sigma \in \mathcal{D}(\mathcal{H})} \eta_{\text{Re}}(\mathcal{N}, \sigma)$ .

It turns out that these coefficients and the notion of (regularized/completely) less noisiness  $\succeq_{\text{l.n.}}^{(\text{reg/c})}$  can be understood on a very similar footing, allowing us to treat them with the same tools by expressing the partial orders in terms of relative entropies and the contraction coefficients in terms of mutual information. The key argument here is the following Proposition that is an extension of [78, Proposition 4]:

**Proposition 2.3.** *Given two channels  $\mathcal{M} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$  and  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_{B'})$ ,  $n \in \mathbb{N}$ ,  $\sigma_A \in \mathcal{D}(\mathcal{H}_A)$  and  $\eta \geq 0$ , the following are equivalent:*

(i) For all c-q state  $\rho_{UA}$  with marginal  $\sigma_A = \text{Tr}_U(\rho_{UA})$ , where  $U$  is an arbitrary classical system,

$$\eta I(U : B)_{(\text{id}_U \otimes \mathcal{M})(\rho_{UA})} \geq I(U : B')_{(\text{id}_U \otimes \mathcal{N})(\rho_{UA})}.$$

(ii) For any state  $\rho_A \in \mathcal{D}(\mathcal{H}_A)$  with  $\text{supp}(\rho_A) \subseteq \text{supp}(\sigma_A)$ ,

$$\eta D(\mathcal{M}(\rho_A) \| \mathcal{M}(\sigma_A)) \geq D(\mathcal{N}(\rho_A) \| \mathcal{N}(\sigma_A)). \quad (2.11)$$

*Proof.* We postpone the proof to Section 6, where a more general equivalence between  $f$ -divergences and the corresponding mutual information quantities is proven in Proposition 6.2.  $\square$

Note that the above Proposition also applies to our regularized quantities by choosing  $\mathcal{M}' = \mathcal{M}^{\otimes n}$  and to the complete ones by choosing  $\mathcal{M}' = \mathcal{M} \otimes \text{id}$ . A direct consequence is that,  $\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}$  if and only if Equation (2.11) holds for  $\eta = 1$ , whereas  $\mathcal{M} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{N}$  if and only if Equation (2.11) holds for  $\mathcal{M}^{\otimes n}$  and all  $n \in \mathbb{N}$  and  $\eta = 1$ . Finally,  $\mathcal{M} \succeq_{\text{l.n.}}^c \mathcal{N}$  if and only if Equation (2.11) holds for  $\mathcal{M} \otimes \text{id}$  and  $\eta = 1$ .

The next result is another direct consequence of the above Proposition and provides us with an alternative way to express  $\eta_{\text{Re}}(\mathcal{N}, \sigma)$  in terms of the mutual information.

**Lemma 2.4.**

$$\eta_{\text{Re}}(\mathcal{N}, \sigma) = \sup_U \sup_{\substack{\rho_{UA} \\ 0 < I(U:A)_{\rho_{UA}} < \infty \\ \text{Tr}_U(\rho_{UA}) = \sigma}} \frac{I(U : B)_{(\text{id}_U \otimes \mathcal{N})(\rho_{UA})}}{I(U : A)_{\rho_{UA}}}. \quad (2.12)$$

*Proof.* The proof follows directly from Proposition 2.3.  $\square$

The above result suggests the introduction of the following so-called *less noisy domination factor* [50]: given two channels  $\mathcal{M} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$  and  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_{B'})$ , and a state  $\sigma \in \mathcal{D}(\mathcal{H}_A)$ :

$$\eta_{\text{Re}}(\mathcal{M}, \mathcal{N}, \sigma) := \sup_{\substack{\rho \in \mathcal{D}(\mathcal{H}_A) \\ 0 < D(\mathcal{M}(\rho) \| \mathcal{M}(\sigma)) < \infty}} \frac{D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma))}{D(\mathcal{M}(\rho) \| \mathcal{M}(\sigma))} \equiv \sup_U \sup_{\substack{\rho_{UA} \\ 0 < I(U:B)_{\rho_{UA}} < \infty}} \frac{I(U : B')}{I(U : B)}.$$

and  $\eta_{\text{Re}}(\mathcal{M}, \mathcal{N}) := \sup_{\sigma} \eta(\mathcal{M}, \mathcal{N}, \sigma)$ . In the case when  $\mathcal{M} = \text{id}$ , we retrieve the usual SDPI constants of  $\mathcal{N}$ . In particular, from Proposition 2.3 we directly have

$$\eta_{\text{Re}}(\mathcal{M}, \mathcal{N}) \leq 1 \quad \Leftrightarrow \quad \mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}.$$

One can of course also define complete and regularized versions of these coefficients. However, it can be easily seen that the *complete* contraction coefficient  $\eta_{\text{Re}}^c(\mathcal{N})$  is equal to 1 for any channel  $\mathcal{N}$ , which limits its usefulness severely.

Finally, we want to end this section by pointing out some relations between the less noisy order in relation to erasure channels and contraction coefficients which will serve as an example of the close relation of the two concepts and reappear throughout this work. It is immediately clear from the above discussion that we have

$$\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N} \implies \eta_{\text{Re}}(\mathcal{M}, \sigma) \leq \eta_{\text{Re}}(\mathcal{N}, \sigma) \quad \forall \sigma \quad (2.13)$$

$$\implies \eta_{\text{Re}}(\mathcal{M}) \leq \eta_{\text{Re}}(\mathcal{N}) \quad (2.14)$$

Now consider  $\mathcal{M}$  to be the quantum erasure channel  $\mathcal{M}_{\eta_e}^{\text{er}}$  of erasure parameter  $\eta_e$ :  $\mathcal{H}_B = \mathcal{H}_A \oplus \mathbb{C}$  and for any  $X \in \mathcal{B}(\mathcal{H}_A)$ ,

$$\mathcal{M}_{\eta_e}^{\text{er}}(X) := (1 - \eta_e)X + \eta_e|e\rangle\langle e|, \quad (2.15)$$

where  $|e\rangle \perp \mathcal{H}_A$ .

The following is a direct consequence of Proposition 2.3:

**Proposition 2.5.** *Let  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$  and  $\eta_e \in [0, 1]$ . Then the following are equivalent:*



$$(i) \mathcal{M}_{\eta_e}^{\text{er}} \succeq_{\text{l.n.}} \mathcal{N}.$$

$$(ii) \eta_{\text{Re}}(\mathcal{N}) \leq \eta_{\text{Re}}(\mathcal{M}_{\eta_e}^{\text{re}}) = (1 - \eta_e).$$

*Proof.* First, a simple calculation gives the following (see [82, Proposition 21.6.1]):

$$I(U : B)_{(\text{id}_U \otimes \mathcal{M}_{\eta_e}^{\text{er}})(\rho_{UA})} = (1 - \eta_e) I(U : A)_{\rho_{UA}}.$$

The result then follows directly from Proposition 2.3.  $\square$

Thus, we conclude that a bound on the relative entropy contraction coefficient is equivalent to a less noisy relation for the corresponding channel when compared to the erasure channel. In the next section we will discuss further properties of the concepts introduced in this section.

### 3 Properties and Bounds

In this section we will discuss properties of partial orders and contraction coefficients. First we will discuss tensorization properties of partial orders, then we discuss bounds on contraction coefficients and finally we illustrate our results with several simple examples.

#### 3.1 Tensorization of partial orders

In the classical setting, it is a well-known fact that the notion of less noisiness *tensorizes* (see e.g. [67, Proposition 16], [72, Proposition 5]): Given channels  $\mathcal{M}_1 \succeq_{\text{l.n.}} \mathcal{N}_1$  and  $\mathcal{M}_2 \succeq_{\text{l.n.}} \mathcal{N}_2$ ,

$$\mathcal{M}_1 \otimes \mathcal{M}_2 \succeq_{\text{l.n.}} \mathcal{N}_1 \otimes \mathcal{N}_2.$$

In particular, we have as expected that the notions of less noisy and regularized less noisy are equivalent. This is false in general in the quantum case, as we will also discuss in more details further down. Often tensorization properties can be recovered when allowing for an additional reference system. The same is true in the case of the less noisy ordering and we show in the following that the completely less noisy ordering tensorizes.

**Lemma 3.1** (The complete less noisy preorder tensorizes). *For any four channels with  $\mathcal{M}_1 \succeq_{\text{l.n.}}^c \mathcal{N}_1$  and  $\mathcal{M}_2 \succeq_{\text{l.n.}}^c \mathcal{N}_2$ , it follows that*

$$\mathcal{M}_1 \otimes \mathcal{M}_2 \succeq_{\text{l.n.}}^c \mathcal{N}_1 \otimes \mathcal{N}_2.$$

*Proof.* Assume that  $\mathcal{M}_1 \succeq_{\text{l.n.}}^c \mathcal{N}_1$  and  $\mathcal{M}_2 \succeq_{\text{l.n.}}^c \mathcal{N}_2$ . Therefore, for any two tripartite states  $\rho_{RA_1A_2}, \sigma_{RA_1A_2}$ , we have

$$\begin{aligned} D((\text{id}_R \otimes \mathcal{M}_1 \otimes \mathcal{M}_2)(\rho_{RA_1A_2}) \| (\text{id}_R \otimes \mathcal{M}_1 \otimes \mathcal{M}_2)(\sigma_{RA_1A_2})) \\ \geq D((\text{id}_R \otimes \mathcal{N}_1 \otimes \mathcal{M}_2)(\rho_{RA_1A_2}) \| (\text{id}_R \otimes \mathcal{N}_1 \otimes \mathcal{M}_2)(\sigma_{RA_1A_2})) \\ \geq D((\text{id}_R \otimes \mathcal{N}_1 \otimes \mathcal{N}_2)(\rho_{RA_1A_2}) \| (\text{id}_R \otimes \mathcal{N}_1 \otimes \mathcal{N}_2)(\sigma_{RA_1A_2})) \end{aligned}$$

where we successively used that  $\mathcal{M}_1 \succeq_{\text{l.n.}}^c \mathcal{N}_1$  for the reference system  $RB_2$  and  $\mathcal{M}_2 \succeq_{\text{l.n.}}^c \mathcal{N}_2$  for the reference system  $RA_1$ .  $\square$

The previous result shows in particular that the notion of complete less noisiness is more stringent than the one of (regularized) less noisiness, but potentially less than that of degradation. The relations are also summarized in Figure 1.

**Proposition 3.2.** *Given two channels  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{H}_B)$  and  $\mathcal{M} \in \text{CPTP}(\mathcal{H}_A, \mathcal{H}_{B'})$*

$$\mathcal{N} \succeq_{\text{deg}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^c \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}} \mathcal{M}, \quad (3.1)$$

$$\mathcal{N} \succeq_{\text{deg}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\text{fq}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}} \mathcal{M}. \quad (3.2)$$

*Proof.* We first prove Equation (3.1): The last implication is obvious and was already discussed in Section 2.1. The first implication follows by the data processing inequality: Assume that  $\mathcal{N} \succeq_{\text{deg}} \mathcal{M}$ . Therefore, by definition there exists a channel  $\Theta$  such that  $\mathcal{M} = \Theta \circ \mathcal{N}$ , which also implies that for any reference system  $R$ ,  $\text{id}_R \otimes \mathcal{M} = \text{id}_R \otimes \Theta \circ \mathcal{N}$ . Then for any two states  $\rho_{AR}, \sigma_{AR}$

$$\begin{aligned} D((\text{id}_R \otimes \mathcal{M})(\rho_{AR}) \| (\text{id}_R \otimes \mathcal{M})(\sigma_{AR})) &= D((\text{id}_R \otimes \Theta \circ \mathcal{N})(\rho_{AR}) \| (\text{id}_R \otimes \Theta \circ \mathcal{N})(\sigma_{AR})) \\ &\leq D((\text{id}_R \otimes \mathcal{N})(\rho_{AR}) \| (\text{id}_R \otimes \mathcal{N})(\sigma_{AR})). \end{aligned}$$

where the inequality follows by DPI for the channel  $\text{id} \otimes \Theta$ . Therefore  $\mathcal{N} \succeq_{\text{l.n.}}^c \mathcal{M}$ . Finally, that  $\mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{M}$  is a simple consequence of Lemma 3.1 and the characterization of  $\succeq_{\text{l.n.}}^{\text{reg}}$  in Proposition 2.3. Equation (3.2) follows by data processing and because classical-quantum states are a special case of general quantum states.  $\square$

We will now briefly discuss whether the less noisy ordering tensorizes. It does in the classical case [67, Proposition 16], [72, Proposition 5]. In the quantum setting tensorization does not hold. Namely, we can find  $\mathcal{M}_1, \mathcal{M}_2, \mathcal{N}_1, \mathcal{N}_2$  such that  $\mathcal{M}_1 \succeq_{\text{l.n.}} \mathcal{M}_2$ ,  $\mathcal{N}_1 \succeq_{\text{l.n.}} \mathcal{N}_2$  but  $\mathcal{M}_1 \otimes \mathcal{M}_2 \not\succeq_{\text{l.n.}} \mathcal{N}_1 \otimes \mathcal{N}_2$ . This follows directly from superactivation of the private capacity. If we have  $P^{(1)}(\mathcal{N}_{A \rightarrow BE}) = 0$ , it follows that the channel to Eve is less noisy than that to Bob. On the other hand, from  $P^{(1)}(\mathcal{N}_{A \rightarrow BE}^{\otimes 2}) > 0$ , it follows that this order does not carry over to the two copy case. Note that the additivity of the private information questions has been recently considered in [75], showing that when the sender is quantum but the receivers are classical additivity still holds. However, when either one of the receivers is quantum, additivity is violated, even when the sender is classical. Since the counterexamples in [75] exhibit the desired superactivation feature, the results directly carry over to the tensorization problem for the less noisy order.

It should be noted that all the examples above are based on wiretap channels that are not necessarily isometric. It would be interesting to find similar examples in the isometric case, i.e. where the channel to Eve is the complement of that to Bob. From the same argument, it can easily be seen that anti-less noisy channels have  $P^{(1)}(\mathcal{N}) = 0$ , however anti-regularized less noisy is needed to ensure that  $P(\mathcal{N}) = 0$ . Another consequence is that the less noisy ordering and the completely less noisy ordering (and therefore the degradable ordering) cannot be equivalent. An interesting open problem is whether the regularized or complete less noisy ordering could still be equivalent to the degradable ordering.

From the above it also follows that when checking for the regularized less noisy order one has to check  $n > 1$ . On the other hand, one can show that it is sufficient to check the condition in the asymptotic limit.

**Lemma 3.3.** *The following two conditions are equivalent:*

1.  $\mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{M}$ .
2. For all  $c - q$  states  $\rho_{UA^n}$ ,

$$\lim_{n \rightarrow \infty} I(U : B^n)_{(\text{id}_U \otimes \mathcal{M}^{\otimes n})(\rho_{UA^n})} \geq \lim_{n \rightarrow \infty} I(U : B^n)_{(\text{id}_U \otimes \mathcal{N}^{\otimes n})(\rho_{UA^n})}.$$

*Proof.* Since regularized less noisy requires the mutual information condition to hold for all  $n$  the direction (1)  $\Rightarrow$  (2) is obvious. The other direction follows by noting that if the condition holds for some  $n$  it also holds for  $n - 1$ : This is simply because if the condition holds for all states  $\rho_{UA^n}$  it in particular also holds for those of the form  $\rho_{UA^n} = \rho_{UA^{n-1}} \otimes \rho_{A_n}$ , for which we have

$$I(U : B^n) = I(U : B^{n-1}). \quad (3.3)$$

This directly implies the result for  $n - 1$  and therefore the claim follows by starting at  $n = \infty$  and stepwise reducing  $n$ .  $\square$

We could also ask for a weaker requirement than tensorization. We call a partial order  $\succeq$  *tensor stable* if for all channel the following holds: If  $\mathcal{N}_1 \succeq \mathcal{N}_2$  then  $\mathcal{N}_1 \otimes \mathcal{M} \succeq \mathcal{N}_2 \otimes \mathcal{M}$ . Every partial order that tensorizes is automatically tensor stable, however also those that do not tensorize can potentially still possess this property. We will show that for the less noisy ordering, while the property holds for some channels, in general the order is not tensor stable. Again an example based on erasure channels can serve as a primer.

**Lemma 3.4.** *Fix any two erasure channels  $\mathcal{E}_1$  and  $\mathcal{E}_2$  with erasure probability  $\epsilon_1$  and  $\epsilon_2$ , respectively. For any channel  $\mathcal{M}$  we have the following:*

$$\mathcal{E}_1 \succeq_{\text{l.n.}} \mathcal{E}_2 \Rightarrow \mathcal{E}_1 \otimes \mathcal{M} \succeq_{\text{l.n.}} \mathcal{E}_2 \otimes \mathcal{M}. \quad (3.4)$$

*Proof.* First note that  $\mathcal{E}_1 \succeq_{\text{l.n.}} \mathcal{E}_2$  is clearly equivalent to  $\epsilon_1 \geq \epsilon_2$ . We now need to show that

$$I(U : E_1 B) \geq I(U : E_2 B). \quad (3.5)$$

We are using

$$I(U : B_1 B_2) = (1 - \epsilon)I(U : B_1 A_2) + \epsilon I(U : B_1), \quad (3.6)$$

which follows essentially from [82, Eqn. (20.90) – (20.94)], to show that the above is equivalent to

$$0 \geq (\epsilon_1 - \epsilon_2) (I(U : A_1 B) - I(U : B)), \quad (3.7)$$

which holds true since both terms in the product are positive: The first by assumption and the second by data processing. This concludes the proof.  $\square$

A direct implication is that for two erasure channels,

$$\mathcal{E}_1 \succeq_{\text{l.n.}} \mathcal{E}_2 \Rightarrow \mathcal{E}_1 \succeq_{\text{l.n.}}^c \mathcal{E}_2. \quad (3.8)$$

The same would generalize to all channels if less noisy were tensor stable in general. In that case, the less noisy and the completely less noisy orderings would be equivalent, which we know they are not. Therefore, the less noisy ordering can not be tensor stable in general.

### 3.2 Bounds on contraction coefficients

In this section, we will discuss several properties and bounds of SDPI constants and contraction coefficients. Let's start with a simple bound for concatenated channels.

**Lemma 3.5.** *We have*

$$\eta_{\text{Re}}(\mathcal{M} \circ \mathcal{N}, \sigma) \leq \eta_{\text{Re}}(\mathcal{M}, \mathcal{N}(\sigma)) \eta_{\text{Re}}(\mathcal{N}, \sigma), \quad (3.9)$$

$$\eta_{\text{Re}}(\mathcal{M} \circ \mathcal{N}) \leq \eta_{\text{Re}}(\mathcal{M}) \eta_{\text{Re}}(\mathcal{N}). \quad (3.10)$$

*Proof.* Observe,

$$\eta_{\text{Re}}(\mathcal{M} \circ \mathcal{N}, \sigma) = \frac{D(\mathcal{M} \circ \mathcal{N}(\rho) \| \mathcal{M} \circ \mathcal{N}(\sigma))}{D(\rho \| \sigma)} = \frac{D(\mathcal{M} \circ \mathcal{N}(\rho) \| \mathcal{M} \circ \mathcal{N}(\sigma))}{D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma))} \frac{D(\mathcal{N}(\rho) \| \mathcal{N}(\sigma))}{D(\rho \| \sigma)}, \quad (3.11)$$

assuming that  $\rho$  is the optimizing state, from which the first claim follows directly. The second then follows by taking the supremum over  $\sigma$  on both sides.  $\square$

A common approach to bounding capacities is to consider flagged channels. This approach also translates to contraction coefficients. Let  $\mathcal{N} = \sum_i \lambda_i \mathcal{N}_i$  and define

$$\hat{\mathcal{N}} = \sum_i \lambda_i \mathcal{N}_i \otimes |i\rangle\langle i|. \quad (3.12)$$

**Lemma 3.6.** *For all state  $\sigma$ , we have*

$$\eta_{\text{Re}}(\mathcal{N}, \sigma) \leq \eta_{\text{Re}}(\hat{\mathcal{N}}, \sigma) \leq \sum_i \lambda_i \eta_{\text{Re}}(\mathcal{N}_i, \sigma). \quad (3.13)$$

*In fact, the two above bounds are generally true for any contraction coefficient based on a divergence satisfying the data processing inequality. The first inequality also extends to the case of non-trace preserving  $\mathcal{N}_i$ .*

*Proof.* The first inequality follows from data processing. The second from the properties of flagged channels and splitting the supremum.  $\square$

Note that, in general, the flags do not have to be orthogonal and one could define an extension

$$\tilde{\mathcal{N}} = \sum_i \lambda_i \mathcal{N}_i \otimes \tau_i, \quad (3.14)$$

leading to a potentially tighter bound

$$\eta_{\text{Re}}(\mathcal{N}, \sigma) \leq \eta_{\text{Re}}(\tilde{\mathcal{N}}, \sigma). \quad (3.15)$$

This approach has recently been employed to find very tight bounds on quantum capacities [30, 77].

We will now discuss whether contraction coefficients tensorize and give some simple examples to the contrary. We start with the contraction coefficient  $\eta_{\text{Re}}(\mathcal{M})$ . Our counterexample is going to be based on the erasure channel. As mentioned before, we have

$$\eta_{\text{Re}}(\mathcal{E}) = 1 - \epsilon. \quad (3.16)$$

Our goal will be to show that  $\eta_{\text{Re}}(\mathcal{E}) \neq \eta_{\text{Re}}(\mathcal{E} \otimes \mathcal{E})$ . First, we will see that the contraction coefficient cannot get smaller under tensorization.

**Lemma 3.7.** *The following holds for any two quantum channels:*

$$\eta_{\text{Re}}(\mathcal{M} \otimes \mathcal{N}) \geq \max\{\eta_{\text{Re}}(\mathcal{M}), \eta_{\text{Re}}(\mathcal{N})\}.$$

*Proof.* Recall that

$$\eta_{\text{Re}}(\mathcal{M}) := \sup_{\rho_{UA}} \frac{I(U : B)_{(\text{id}_U \otimes \mathcal{M})(\rho_{UA})}}{I(U : A)_{\rho_{UA}}}. \quad (3.17)$$

Pick  $\rho_{UA_1}$  as the optimizer in  $\eta_{\text{Re}}(\mathcal{M})$ . With  $\sigma_{UA_1A_2} = \rho_{UA_1} \otimes \rho_{A_2}$  with arbitrary  $\rho_{A_2}$ , we immediately have

$$\eta_{\text{Re}}(\mathcal{M} \otimes \mathcal{N}) \geq \eta_{\text{Re}}(\mathcal{M}) \quad (3.18)$$

and

$$\eta_{\text{Re}}(\mathcal{M} \otimes \mathcal{N}) \geq \eta_{\text{Re}}(\mathcal{N}) \quad (3.19)$$

follows by symmetry.  $\square$

Upper bounds are more involved, but one can give a nice one when one channel is an erasure channel.

**Lemma 3.8.** *For all quantum channel  $\mathcal{M}$ , we have*

$$\eta_{\text{Re}}(\mathcal{M} \otimes \mathcal{E}) \leq (1 - \epsilon) + \epsilon \eta_{\text{Re}}(\mathcal{M}) \leq (1 + \epsilon) \max\{\eta_{\text{Re}}(\mathcal{E}), \eta_{\text{Re}}(\mathcal{M})\}. \quad (3.20)$$

*Proof.* The main ingredient is the following equality

$$I(U : B_1 B_2) = (1 - \epsilon) I(U : B_1 A_2) + \epsilon I(U : B_1), \quad (3.21)$$

which follows essentially from [82, Eqn. (20.90) – (20.94)]. We have,

$$\frac{I(U : B_1 B_2)}{I(U : A_1 A_2)} = (1 - \epsilon) \frac{I(U : B_1 A_2)}{I(U : A_1 A_2)} + \epsilon \frac{I(U : B_1)}{I(U : A_1 A_2)} \quad (3.22)$$

$$\leq (1 - \epsilon) + \epsilon \frac{I(U : B_1)}{I(U : A_1)} \quad (3.23)$$

$$\leq (1 - \epsilon) + \epsilon \eta_{\text{Re}}(\mathcal{M}), \quad (3.24)$$

where the first inequality follows from applying data-processing twice and the second by taking the supremum over all  $\rho_{UA_1}$ .

Now the second inequality follows by noting that

$$(1 - \epsilon) + \epsilon \eta_{\text{Re}}(\mathcal{M}) = \eta_{\text{Re}}(\mathcal{E}) + \epsilon \eta_{\text{Re}}(\mathcal{M}) \quad (3.25)$$

and considering the two cases  $\eta_{\text{Re}}(\mathcal{E}) \geq \eta_{\text{Re}}(\mathcal{M})$  and  $\eta_{\text{Re}}(\mathcal{M}) \geq \eta_{\text{Re}}(\mathcal{E})$ .  $\square$

This allows us to calculate the contraction coefficient  $\eta_{\text{Re}}(\mathcal{E}^{\otimes n})$ .

**Lemma 3.9.** *Given the erasure channel  $\mathcal{E}$  of parameter  $\epsilon$ , we have*

$$\eta_{\text{Re}}(\mathcal{E}^{\otimes n}) = 1 - \epsilon^n. \quad (3.26)$$

*Proof.* The  $\leq$  direction follows from Lemma 3.8. Set  $\mathcal{M} = \mathcal{E}^{\otimes n-1}$  and define a function  $f(n)$  by  $f(1) = 1 - \epsilon$  and  $f(n) = (1 - \epsilon) + \epsilon f(n-1)$  for  $n \geq 2$ . It can easily be shown by induction that  $f(n) = 1 - \epsilon^n$  and therefore we have  $\eta_{\text{Re}}(\mathcal{E}^{\otimes n}) = 1 - \epsilon^n$ .

For the opposite direction we have to show that this bound is indeed achievable. Consider the following state

$$\rho_{UA_1A_2} = \frac{1}{2}|0\rangle\langle 0|_U \otimes |0\rangle\langle 0|_{A_1^n} + \frac{1}{2}|1\rangle\langle 1|_U \otimes |1\rangle\langle 1|_{A_1^n}. \quad (3.27)$$

It can be checked that

$$I(U : A_1A_2) = \log 2 \quad (3.28)$$

$$H(U) = \log 2 \quad (3.29)$$

$$H(UB_1B_2) - H(B_1B_2) = \epsilon^n \log 2 \quad (3.30)$$

$$\Rightarrow I(U : B_1B_2) = (1 - \epsilon^2) \log 2 \quad (3.31)$$

where the only non-trivial one is Equation 3.30, which follows from observing that the only overlap in support between the two output states of the erasure channels corresponds to the state being completely deleted. From this the desired result follows immediately.  $\square$

This completes our argument that the contraction coefficient does not tensorize.

However,  $\eta_{\text{Re}}(\mathcal{M}, \sigma)$  might still tensorize. Indeed, this is the case for erasure channels. Consider two erasure channels  $\mathcal{E}_1$  and  $\mathcal{E}_2$  with erasure probability  $\epsilon_1$  and  $\epsilon_2$ , respectively. By Lemma 3.8 we immediately get

$$\eta_{\text{Re}}(\mathcal{E}_1 \otimes \mathcal{E}_2) \leq 1 - \epsilon_1 \epsilon_2. \quad (3.32)$$

Opposed to that, in the restricted case, tensorization holds as can be seen in the following lemma.

**Lemma 3.10.** *For the erasure channels defined above, we have*

$$\eta_{\text{Re}}(\mathcal{E}_1 \otimes \mathcal{E}_2, \sigma_1 \otimes \sigma_2) \leq \max\{(1 - \epsilon_1), (1 - \epsilon_2)\}. \quad (3.33)$$

*Proof.* As before, we can use Equation 3.21, however, this time we apply it twice and we get

$$I(U : B_1B_2) = (1 - \epsilon_1)(1 - \epsilon_2)I(U : A_1A_2) + (1 - \epsilon_2)\epsilon_1 I(U : A_1) + (1 - \epsilon_1)\epsilon_2 I(U : A_2). \quad (3.34)$$

The second main ingredient is the following argument:

$$I(U : A_1A_2) = I(U : A_2) + I(U : A_1|A_2) \quad (3.35)$$

$$= I(U : A_2) + I(UA_2 : A_1) - I(A_2 : A_1) \quad (3.36)$$

$$= I(U : A_2) + I(UA_2 : A_1) \quad (3.37)$$

$$\geq I(U : A_2) + I(U : A_1), \quad (3.38)$$

where the first two inequalities follow by the chain rule, the third because  $\rho_{A_1 A_2} = \sigma_1 \otimes \sigma_2$  and the inequality by data processing.

Now, we have to distinguish two cases. First, consider  $\epsilon_1 \geq \epsilon_2$ . We have,

$$(1 - \epsilon_1)(1 - \epsilon_2)I(U : A_1 A_2) + (1 - \epsilon_2)\epsilon_1 I(U : A_1) + (1 - \epsilon_1)\epsilon_2 I(U : A_2) \quad (3.39)$$

$$\leq (1 - \epsilon_1)(1 - \epsilon_2)I(U : A_1 A_2) + (1 - \epsilon_2)\epsilon_1 [I(U : A_1 A_2) - I(U : A_2)] + (1 - \epsilon_1)\epsilon_2 I(U : A_2) \quad (3.40)$$

$$= (1 - \epsilon_1)I(U : A_1 A_2) + (\epsilon_1 - \epsilon_2)I(U : A_2) \quad (3.41)$$

$$\leq (1 - \epsilon_2)I(U : A_1 A_2), \quad (3.42)$$

where the first inequality follows by using the previously derived argument, the equality is simple rearranging and the final inequality is data-processing together with the assumption that  $(\epsilon_1 - \epsilon_2) \geq 0$ . It follows directly that in this case

$$\eta_{\text{Re}}(\mathcal{E}_1 \otimes \mathcal{E}_2, \sigma_1 \otimes \sigma_2) \leq (1 - \epsilon_2). \quad (3.43)$$

The second part follows directly by assuming  $\epsilon_2 \geq \epsilon_1$  and exchanging the roles of  $A_1$  and  $A_2$  in the previous derivation. Putting both cases together concludes the proof.  $\square$

Whether SDPI constants based on the relative entropy for fixed tensor product state tensorize in general remains an important open problem.

### 3.3 Additional examples

First, we recall an example from [34] for qubit channels. The original theorem applies to several contraction coefficients, we state it here for the relative entropy.

**Lemma 3.11** ([34, Theorem 6.1]). *For any unital map  $\mathcal{M}_T : I + w \cdot \sigma \rightarrow I + (Tw) \cdot \sigma$  where  $T$  is a real matrix with  $\|T\|_\infty \leq 1$ ,*

$$\eta_{\text{Re}}(\mathcal{M}_T) = \|T\|_\infty^2. \quad (3.44)$$

Examples to which this result applies include the depolarizing channel,

$$\mathcal{D}_p(\cdot) = (1 - p)(\cdot) + p\frac{\mathbb{1}}{2}, \quad (3.45)$$

and the dephasing and bit-flip channel,

$$\mathcal{D}_p^Z = (1 - p)(\cdot) + pZ(\cdot)Z, \quad (3.46)$$

$$\mathcal{D}_p^X = (1 - p)(\cdot) + pX(\cdot)X, \quad (3.47)$$

We get,

$$\eta(\mathcal{D}_p) = (1 - p)^2, \quad (3.48)$$

$$\eta(\mathcal{D}_p^X) = \eta(\mathcal{D}_p^Z) = 1. \quad (3.49)$$

We can combine this with results from the previous section to test our bounds on products of channels. Let  $\mathcal{E}_{1/2}$  be an erasure channel with erasure probability  $\epsilon = \frac{1}{2}$ , using the first upper bound in Lemma 3.8 and lower bounding by fixing the input state as that from Equation 3.27, we get

$$\frac{\log 2 - \frac{1}{2}h(\frac{p}{2})}{\log 2} \leq \eta_{\text{Re}}(\mathcal{E}_{1/2} \otimes \mathcal{D}_p) \leq \frac{1}{2}(1 + (1 - p)^2). \quad (3.50)$$

These simple bounds already limit the possible values significantly to the green area in Figure 2, where we also plot the looser second bound from Lemma 3.8 and the lower bound from Lemma 3.7 for comparison.

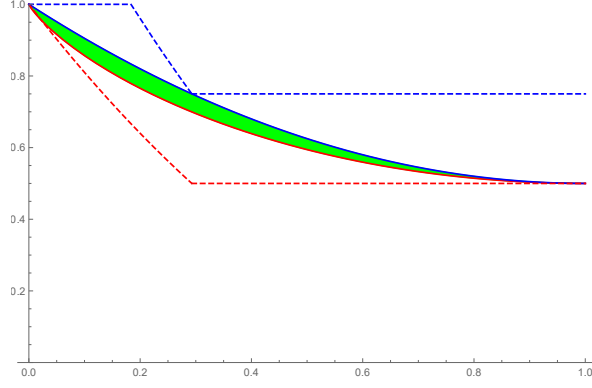


Figure 2: Bounds on  $\eta_{\text{Re}}(\mathcal{E}_{1/2} \otimes \mathcal{D}_p)$  as given in Equation (3.50). The green area marks the remaining possible range of the contraction coefficient.

Let us consider one more example. A channel that has recently proven useful in investigating the properties of quantum capacities is the dephasure channel [45],

$$\mathcal{D}_{\epsilon,p} = \mathcal{E}_{\epsilon} \circ \mathcal{D}_p^Z = (1 - \epsilon) ((1 - p)(\cdot) + pZ(\cdot)Z)) + \epsilon|e\rangle\langle e|. \quad (3.51)$$

We can easily get a lower bound on the contraction coefficient by picking the input state in Equation 3.27. An upper bound can either be found from using Lemma 3.5 or Lemma 3.6 (for the latter note that the dephasure channel can directly be written as a flag channel). Together we get,

$$\eta_{\text{Re}}(\mathcal{D}_{\epsilon,p}) = 1 - \epsilon. \quad (3.52)$$

Finally, a simple observation is that a replacer channel that always outputs a state  $\tau$ ,

$$\mathcal{R}_{\tau}(\cdot) = \tau \text{Tr}(\cdot). \quad (3.53)$$

is the noisiest channel and an isometric channel  $\mathcal{V}(\rho) = V\rho V^{\dagger}$  for some isometry  $V$  is the least noisy channel.

**Lemma 3.12.** *For any channel  $\mathcal{N}$ , replacer channel  $\mathcal{R}_{\tau}$  and isometric channel  $\mathcal{V}$ , we have*

$$\eta_{\text{Re}}(\mathcal{R}_{\tau}) = 0, \quad \eta_{\text{Re}}(\mathcal{V}) = 1 \quad (3.54)$$

$$\eta_{\text{Re}}(\mathcal{N} \otimes \mathcal{R}_{\tau}) = \eta_{\text{Re}}(\mathcal{N}) \quad (3.55)$$

$$\mathcal{V} \succeq_{\text{deg}} \mathcal{N} \succeq_{\text{deg}} \mathcal{R}_{\tau}, \quad (3.56)$$

where the final line implies a similar statement for all other partial orders defined here via Proposition 3.2.

*Proof.* Note that

$$D(\tau\|\tau) = 0, \quad (3.57)$$

$$D(\rho \otimes \tau\|\sigma \otimes \tau) = D(\rho\|\sigma), \quad (3.58)$$

$$D(V\rho V^{\dagger}\|V\sigma V^{\dagger}) = D(\rho\|\sigma), \quad (3.59)$$

by additivity, faithfulness and data processing of the relative entropy. The proof of the first two lines follows easily from there. The third line is obvious by choosing in the first step  $\mathcal{N} \circ \mathcal{V}^{-1}$  and in the second step  $\mathcal{R}_{\tau}$  as the degrading maps.  $\square$

## 4 Approximate partial orders and capacities

In this section, we explore approximations of the partial orders introduced in Section 2. These are powerful tools which can be used for instance in approximating capacities of a quantum channel in terms of the ones of another channel when the latter are simpler to compute.

## 4.1 Approximate partial orders

Previously, Sutter et. al. [73] defined  $\epsilon$ -approximate degradation as follows. We recall that the diamond norm distance between two quantum channels  $\mathcal{M}, \mathcal{N} : A \rightarrow B$  is defined as

$$\|\mathcal{M} - \mathcal{N}\|_{\diamond} := \sup_R \sup_{\rho_{AR}} \|(\mathcal{N} \otimes \text{id}_R - \mathcal{M} \otimes \text{id}_R)(\rho_{AR})\|_1.$$

**Definition 4.1.** A channel  $\mathcal{N}$  is said to be an  $\epsilon$ -degraded version of  $\mathcal{M}$  if there exists a channel  $\Theta$  such that  $\|\mathcal{N} - \Theta \circ \mathcal{M}\|_{\diamond} \leq \epsilon$ .

One can also define approximate versions of other partial orders, as we will do in the following.

**Definition 4.2.** A channel  $\mathcal{N}$  is said to be  $\epsilon$ -completely less noisy than  $\mathcal{M}$  (denoted  $\mathcal{N} \succeq_{\text{l.n.}}^{c, \epsilon} \mathcal{M}$ ) if

$$I(U : ER)_{\mathcal{M}} \leq I(U : BR)_{\mathcal{N}} + \epsilon \quad \forall \rho_{UAR}, \quad (4.1)$$

$\epsilon$ -regularized less noisy than  $\mathcal{M}$  (denoted  $\mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}, \epsilon} \mathcal{M}$ ) if

$$I(U : E^n)_{\mathcal{M}} \leq I(U : B^n)_{\mathcal{N}} + n\epsilon \quad \forall \rho_{UA_1^n} \quad \forall n, \quad (4.2)$$

$\epsilon$ -less noisy than  $\mathcal{M}$  (denoted  $\mathcal{N} \succeq_{\text{l.n.}}^{\epsilon} \mathcal{M}$ ) if

$$I(U : E)_{\mathcal{M}} \leq I(U : B)_{\mathcal{N}} + \epsilon \quad \forall \rho_{UA}, \quad (4.3)$$

$\epsilon$ -fully quantum less noisy than  $\mathcal{M}$  (denoted  $\mathcal{N} \succeq_{\text{l.n.}}^{\text{fq}, \epsilon} \mathcal{M}$ ) if

$$I(A : E)_{\mathcal{M}} \leq I(A : B)_{\mathcal{N}} + \epsilon \quad \forall \rho_{AA'}. \quad (4.4)$$

One defines similarly the approximate versions of the different more capable orders introduced in Section 2. Also  $\epsilon$ -anti orders are defined in the same way but exchanging  $\mathcal{N}$  and  $\mathcal{M}$ .

Following the same arguments as in the exact case, we have immediately that

$$\mathcal{N} \succeq_{\text{l.n.}}^{c, \epsilon} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}, \epsilon} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\epsilon} \mathcal{M}. \quad (4.5)$$

We would of course hope that  $\epsilon$ -approximate degradability implies the above orders. This is indeed the case as we will see in a moment. To relate the two partial orders we will make use of the continuity bounds introduced in [2] and [86]. In what follows,  $h(x) := -x \ln x - (1-x) \ln(1-x)$ ,  $x \in [0, 1]$ , denotes the binary entropy.

**Lemma 4.3** (Continuity bound, Theorem 1 in [2]). For quantum states  $\rho_A$  and  $\sigma_A$ , if  $\frac{1}{2}\|\rho - \sigma\|_1 \leq \epsilon \leq 1$ , then

$$|H(A)_{\rho} - H(A)_{\sigma}| \leq \epsilon \log(|A| - 1) + h(\epsilon). \quad (4.6)$$

**Lemma 4.4** (Alicki-Fannes-Winter bound, Lemma 2 in [86]). For quantum states  $\rho_{AB}$  and  $\sigma_{AB}$ , if  $\frac{1}{2}\|\rho - \sigma\|_1 \leq \epsilon \leq 1$ , then

$$|H(A|B)_{\rho} - H(A|B)_{\sigma}| \leq 2\epsilon \log |A| + (1 + \epsilon) h\left(\frac{\epsilon}{1 + \epsilon}\right), \quad (4.7)$$

where we recall that the conditional entropy of a state  $\rho_{AB}$  is defined as  $H(A|B)_{\rho} := H(AB)_{\rho} - H(B)_{\rho}$ .

This allows us to directly show the following relation.

**Lemma 4.5.** If  $\mathcal{M} : A \rightarrow B$  is an  $\epsilon$ -degraded version of  $\mathcal{N} : A \rightarrow B'$  then  $\mathcal{N}$  is  $\tilde{\epsilon}$ -completely less noisy than  $\mathcal{M}$ , with

$$\tilde{\epsilon} = 2\epsilon \log |B| + (2 + \epsilon) h\left(\frac{\epsilon}{2 + \epsilon}\right).$$

Therefore,

$$\|\mathcal{M} - \Theta \circ \mathcal{N}\|_{\diamond} \leq \epsilon \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{c, \tilde{\epsilon}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}, \tilde{\epsilon}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\tilde{\epsilon}} \mathcal{M}.$$



*Proof.* We assume that  $\|\mathcal{M} - \Theta \circ \mathcal{N}\|_\diamond \leq \varepsilon$  and start with an input state  $\rho_{UAR}$ . Define  $\sigma_{UBR} = (\Theta \circ \mathcal{N}) \otimes \text{id}(\rho_{UAR})$  and  $\tau_{UBR} = \mathcal{M} \otimes \text{id}(\rho_{UAR})$ . First, observe the following:

$$\begin{aligned}
I(U : BR)_{\mathcal{M}(\rho)} - I(U : B'R)_{\mathcal{N}(\rho)} &\leq I(U : BR)_\tau - I(U : BR)_\sigma \\
&= -H(BR)_\sigma + H(UBR)_\sigma - H(UBR)_\tau + H(BR)_\tau \\
&= -H(B|R)_\sigma + H(B|R)_\tau - H(B|UR)_\tau + H(B|UR)_\sigma \\
&\leq 2\epsilon \log |B| + (2 + \epsilon) h\left(\frac{\epsilon}{2 + \epsilon}\right) \\
&=: \tilde{\epsilon},
\end{aligned}$$

where the first inequality follows by data-processing to remove the degrading map. The first equality follows by the definition of the mutual information, the second by adding a zero twice and the fact that  $\tau_{UR} = \sigma_{UR}$  and the last inequality by applying Lemma 4.4 twice. The second claim follows from (4.5).  $\square$

In the same manner, one can go via the fully quantum less noisy order.

**Lemma 4.6.** *If  $\mathcal{M} : A \rightarrow B$  is an  $\epsilon$ -degraded version of  $\mathcal{N} : A \rightarrow B'$  then  $\mathcal{N}$  is  $\hat{\epsilon}$ -fully quantum less noisy than  $\mathcal{M}$ , with*

$$\hat{\epsilon} = \frac{1}{2}\epsilon \log(|B| - 1) + \epsilon \log |B| + \left(1 + \frac{\epsilon}{2}\right) h\left(\frac{\epsilon}{2 + \epsilon}\right) + h\left(\frac{\epsilon}{2}\right).$$

Therefore,

$$\|\mathcal{M} - \Theta \circ \mathcal{N}\|_\diamond \leq \varepsilon \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\text{fq}, \hat{\epsilon}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\hat{\epsilon}} \mathcal{M}.$$

*Proof.* We assume that  $\|\mathcal{M} - \Theta \circ \mathcal{N}\|_\diamond \leq \varepsilon$  and start with an input state  $\rho_{AA'}$ . Define  $\sigma_{AB} = (\Theta \circ \mathcal{N}) \otimes \text{id}(\rho_{AA'})$  and  $\tau_{AB} = \mathcal{M} \otimes \text{id}(\rho_{AA'})$ . First, observe the following:

$$\begin{aligned}
I(A : B)_{\mathcal{M}(\rho)} - I(A : B')_{\mathcal{N}(\rho)} &\leq I(A : B)_\tau - I(A : B)_\sigma \\
&= -H(B)_\sigma + H(AB)_\sigma - H(AB)_\tau + H(B)_\tau \\
&= -H(B)_\sigma + H(B)_\tau - H(B|A)_\tau + H(B|A)_\sigma \\
&\leq \frac{1}{2}\epsilon \log(|B| - 1) + \epsilon \log |B| + \left(1 + \frac{\epsilon}{2}\right) h\left(\frac{\epsilon}{2 + \epsilon}\right) + h\left(\frac{\epsilon}{2}\right) \\
&=: \hat{\epsilon},
\end{aligned}$$

where the first inequality follows by data-processing to remove the degrading map, the first equality is by definition of the mutual information, the second by adding a zero once and the fact that  $\tau_A = \sigma_A$ , the last inequality by applying Lemma 4.3 and Lemma 4.4. The second claim follows from (4.5).  $\square$

This again immediately implies that for two channels  $\mathcal{N}$  and  $\mathcal{M}$ ,

$$\mathcal{N} \succeq_{\text{deg}}^\epsilon \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\text{fq}, \hat{\epsilon}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\hat{\epsilon}} \mathcal{M}, \quad (4.8)$$

with  $\hat{\epsilon} = \frac{1}{2}\epsilon \log(|B| - 1) + \epsilon \log |B| + \left(1 + \frac{\epsilon}{2}\right) h\left(\frac{\epsilon}{2 + \epsilon}\right) + h\left(\frac{\epsilon}{2}\right)$ .

Clearly, the diamond norm is symmetric under exchanging the arguments. This motivates the following definition.

**Definition 4.7.** *Two channels  $\mathcal{N}$  and  $\mathcal{M}$  are said to be comparably  $\epsilon$ - completely less noisy if  $\mathcal{N} \succeq_{\text{l.n.}}^{\text{c}, \epsilon} \mathcal{M}$  and also  $\mathcal{M} \succeq_{\text{l.n.}}^{\text{c}, \epsilon} \mathcal{N}$ , and similarly for comparably  $\epsilon$ - regularized less noisy and comparably  $\epsilon$ - less noisy.*

**Corollary 4.8.** *Assuming  $\mathcal{N}, \mathcal{M} : A \rightarrow B$ , if  $\|\mathcal{M} - \mathcal{N}\|_\diamond \leq \epsilon$ , then  $\mathcal{M}$  and  $\mathcal{N}$  are comparably  $\hat{\epsilon}$ -completely less noisy, with  $\hat{\epsilon} = 2\epsilon \log |B| + (2 + \epsilon)h\left(\frac{\epsilon}{2 + \epsilon}\right)$ .*

*Proof.* This follows easily by Lemma 4.5 and the symmetry of the diamond norm.  $\square$

As a special case, of course also the non-complete version ( $n = 1$ ) is implied. Furthermore, it can be easily seen that  $\hat{\epsilon}$ -completely less noisy implies  $\hat{\epsilon}$ -completely more capable. Again, the same holds for the non-complete case. In a similar way, we can directly derive a few other Lemmas.

**Lemma 4.9.** *If  $\mathcal{M}$  is an  $\epsilon$ -anti-degraded version of  $\mathcal{N}$ , then  $\mathcal{M}$  is  $\hat{\epsilon}$ -anti-completely less noisy than  $\mathcal{N}$ , with  $\hat{\epsilon} = 2\epsilon \log |E| + (1 + \frac{\epsilon}{2})h(\frac{\epsilon}{2+\epsilon})$ .*

*Proof.* The proof follows by simply exchanging the roles of  $\mathcal{M}$  and  $\mathcal{N}$  in the proof of Lemma 4.5.  $\square$

**Lemma 4.10.** *If  $\mathcal{M}$  is an  $\epsilon$ -conjugate degraded version of  $\mathcal{N}$ , then  $\mathcal{M}$  is  $\hat{\epsilon}$ -completely less noisy than  $\mathcal{N}$ , with  $\hat{\epsilon} = 2\epsilon \log |E| + (1 + \frac{\epsilon}{2})h(\frac{\epsilon}{2+\epsilon})$ .*

*Proof.* We again use the same technique as in the proof of Lemma 4.5. However, we need one additional step. Since we only want to show that a less noisy type ordering is implied we can restrict the input state to be classical-quantum. It needs to be shown that going from  $\sigma_{E_1^n U} = \mathcal{N}^{\otimes n}(\rho_{A_1^n U})$  to  $\nu_{E_1^n U} = \mathcal{C} \circ \mathcal{N}^{\otimes n}(\rho_{A_1^n U})$  does not change the mutual information, where  $\mathcal{C}$  is the conjugation map. We have that,

$$I(E_1^n : U) = H(E_1^n) - H(E_1^n | U) \quad (4.9)$$

$$= H(E_1^n) - \sum_u p(u) H(E_1^n | U = u). \quad (4.10)$$

Now, since  $(\mathcal{C}^{\otimes n} \circ \mathcal{N}^{\otimes n})(\rho)^T = \mathcal{N}^{\otimes n}(\rho)$  it follows that the mutual information remains unchanged.  $\square$

## 4.2 Approximate capacities of a quantum channel

Next, we consider a quantum channel  $\mathcal{N}$  with associated complementary channel  $\mathcal{N}^{c3}$ . We want to investigate properties of the private capacity  $P$  and the quantum capacity  $Q$ . Let us start with the following.

**Theorem 4.11.** *Let  $\mathcal{N}$  be a quantum channel.*

- (i) *If  $\mathcal{N}$  is  $\epsilon$ -more capable, then  $Q^{(1)}(\mathcal{N}) \leq P^{(1)}(\mathcal{N}) \leq Q^{(1)}(\mathcal{N}) + \epsilon$ .*
- (ii) *If  $\mathcal{N}$  is  $\epsilon$ -regularized more capable, then  $Q(\mathcal{N}) \leq P(\mathcal{N}) \leq Q(\mathcal{N}) + \epsilon$ .*
- (iii) *If  $\mathcal{N}$  is  $\epsilon$ -regularized less noisy, then*

$$\begin{aligned} Q^{(1)}(\mathcal{N}) &\leq Q(\mathcal{N}) \leq Q^{(1)}(\mathcal{N}) + \epsilon \\ P^{(1)}(\mathcal{N}) &\leq P(\mathcal{N}) \leq P^{(1)}(\mathcal{N}) + 2\epsilon. \end{aligned}$$

*Proof.* The proof of the first part is similar to that of [78, Proposition 1]. The second part is simply the same proof for arbitrary  $n$  and taking the limit, see also [78, Theorem 1]. The third part is similar to the proof of [78, Theorem 2]. Finally, the last part follows directly from the previous bounds.  $\square$

To compare the above results to [73, Theorem 3.4], we can easily make the following corollary:

**Corollary 4.12.** *Let  $\mathcal{N}$  be either  $\epsilon$ -degradable or  $\epsilon$ -conjugate degradable. Then we have*

$$Q^{(1)}(\mathcal{N}) \leq P^{(1)}(\mathcal{N}) \leq Q^{(1)}(\mathcal{N}) + \hat{\epsilon} \quad (4.11)$$

$$Q(\mathcal{N}) \leq P(\mathcal{N}) \leq Q(\mathcal{N}) + \tilde{\epsilon} \quad (4.12)$$

$$Q^{(1)}(\mathcal{N}) \leq Q(\mathcal{N}) \leq Q^{(1)}(\mathcal{N}) + \tilde{\epsilon} \quad (4.13)$$

$$P^{(1)}(\mathcal{N}) \leq P(\mathcal{N}) \leq P^{(1)}(\mathcal{N}) + 2\tilde{\epsilon} \quad (4.14)$$

with

$$\hat{\epsilon} = \frac{1}{2}\epsilon \log(|B| - 1) + \epsilon \log |B| + (1 + \frac{\epsilon}{2})h\left(\frac{\epsilon}{2+\epsilon}\right) + h\left(\frac{\epsilon}{2}\right), \quad (4.15)$$

$$\tilde{\epsilon} = 2\epsilon \log |B| + (2 + \epsilon)h\left(\frac{\epsilon}{2+\epsilon}\right). \quad (4.16)$$

---

<sup>3</sup>For detailed definitions, we refer to [82].

*Proof.* Follows directly from Theorem 4.11 and Lemma 4.6.  $\square$

We can furthermore bound the quantum and private capacity of almost anti-less noisy or anti-more capable channels.

**Theorem 4.13.** *Let  $\mathcal{N}$  be a quantum channel.*

- (i)  *$\mathcal{N}$  be  $\epsilon$ -anti more capable. Then  $Q^{(1)}(\mathcal{N}) \leq \epsilon$ .*
- (ii) *If  $\mathcal{N}$  is  $\epsilon$ -regularized anti more capable, then  $Q(\mathcal{N}) \leq \epsilon$ .*
- (iii) *If  $\mathcal{N}$  be  $\epsilon$ -anti less noisy, then  $P^{(1)}(\mathcal{N}) \leq \epsilon$ .*
- (iv) *If  $\mathcal{N}$  is  $\epsilon$ -regularized anti less noisy, then  $P(\mathcal{N}) \leq \epsilon$ .*

*Proof.* The first two bounds follow immediately by noting that the coherent information can be written as a difference of two mutual informations (see e.g. [82, Theorem 13.6.1]):

$$I(A \rangle B) = I(X : B) - I(X : E). \quad (4.17)$$

The proof of the next two bounds follows immediately from the definitions of the private information and the less noisy partial order.  $\square$

Again, this immediately implies the previously known results about degradable channels (and adding analog ones for conjugate degradable).

**Corollary 4.14.** *Let  $\mathcal{N}$  be either  $\epsilon$ -anti degradable or  $\epsilon$ -conjugate anti degradable. Then we have*

$$Q^{(1)}(\mathcal{N}) \leq P^{(1)}(\mathcal{N}) \leq \hat{\epsilon} \quad (4.18)$$

$$Q(\mathcal{N}) \leq P(\mathcal{N}) \leq \tilde{\epsilon}, \quad (4.19)$$

with

$$\hat{\epsilon} = \frac{1}{2}\epsilon \log(|B| - 1) + \epsilon \log |B| + (1 + \frac{\epsilon}{2}) h\left(\frac{\epsilon}{2 + \epsilon}\right) + h\left(\frac{\epsilon}{2}\right), \quad (4.20)$$

$$\tilde{\epsilon} = 2\epsilon \log |B| + (2 + \epsilon) h\left(\frac{\epsilon}{2 + \epsilon}\right). \quad (4.21)$$

This should be compared to [73, Theorem 3.8].

In a similar fashion we can also reproduce the results in [44].

**Lemma 4.15.** *If  $\mathcal{N}$  is  $\epsilon$ -less noisy than  $\mathcal{M}$ , we have*

$$\chi(\mathcal{M}) \leq \chi(\mathcal{N}) + \epsilon. \quad (4.22)$$

*If  $\mathcal{N}$  is  $\epsilon$ -regularized less noisy than  $\mathcal{M}$ , we have*

$$C(\mathcal{M}) \leq C(\mathcal{N}) + \epsilon. \quad (4.23)$$

*If  $\mathcal{N}$  and  $\mathcal{M}$  are comparably  $\epsilon$ -regularized less noisy and  $\mathcal{N}$  is weakly additive, we have*

$$C(\mathcal{M}) \leq \chi(\mathcal{M}) + 2\epsilon. \quad (4.24)$$

*If  $\mathcal{N}$  is  $\epsilon$ -fully quantum more capable than  $\mathcal{M}$ , we have*

$$C_E(\mathcal{M}) \leq C_E(\mathcal{N}) + \epsilon. \quad (4.25)$$

*Proof.* The first two statements essentially follow by definition of the involved quantities. The third statement follows by chaining the together and applying the additivity:

$$C(\mathcal{M}) \leq C(\mathcal{N}) + \epsilon = \chi(\mathcal{N}) + \epsilon \leq \chi(\mathcal{M}) + 2\epsilon. \quad (4.26)$$

The final statement again follows trivially from the definitions.  $\square$

Combining the previous results we easily get the following corollary.

**Corollary 4.16.** *If  $\|\mathcal{M} - \mathcal{N}\|_\diamond \leq \epsilon$  and  $\mathcal{N}$  is weakly additive, we have*

$$C(\mathcal{M}) \leq \chi(\mathcal{M}) + \hat{\epsilon} + \tilde{\epsilon}, \quad (4.27)$$

$$\hat{\epsilon} = \frac{1}{2}\epsilon \log(|B| - 1) + \epsilon \log |B| + \left(1 + \frac{\epsilon}{2}\right) h\left(\frac{\epsilon}{2 + \epsilon}\right) + h\left(\frac{\epsilon}{2}\right), \quad (4.28)$$

$$\tilde{\epsilon} = 2\epsilon \log |B| + (2 + \epsilon) h\left(\frac{\epsilon}{2 + \epsilon}\right). \quad (4.29)$$

*Proof.* This follows along Equation (4.26) and noting that the final inequality only requires the unregularized  $\epsilon$ -less noisy property.  $\square$

This should be compared to [44, Corollary II.7]. Moreover, we note that given channels  $\mathcal{N}, \mathcal{M}$  it is possible to efficiently compute the minimal  $\epsilon$  such that  $\mathcal{N}$  is an  $\epsilon$  degraded version of  $\mathcal{M}$  [44, 79].

## 5 Characterizations of Strong data processing

### 5.1 SDPI via hypercontractivity

Let us prove some characterizations of strong data processing inequalities (SDPI), i.e. the relative entropy contraction coefficient. Obtaining characterizations of strong SDPIs has been the focus of recent activity in quantum information theory. For instance, in [7] the authors show that:

$$D(\mathcal{M}(\rho) \parallel \mathcal{M}(\sigma)) \leq \eta D(\rho \parallel \sigma) \quad \forall \rho \in \mathcal{S}(A)$$

and

$$\text{tr} \exp(\log \sigma + \mathcal{M}^\dagger(\log \omega)) \leq \left\| \exp\left(\log \omega + \frac{1}{\eta} \log \mathcal{M}(\sigma)\right) \right\|_\eta \quad \forall \omega \in \mathcal{S}(B)$$

are equivalent. Here we will focus on a hypercontractive approach. In the classical setting, results connecting hypercontractivity of a Markov kernel and SDPIs go back to Alswede and Gács [1] and have also been the focus of the more recent [68, Theorem 6]. In order to state our results, let us define the  $\sigma$ -weighted  $p$ -quasi-norms for  $\sigma > 0$  and  $X \in \mathcal{B}(\mathcal{H})$  as:

$$\|X\|_{p,\sigma}^p = \text{tr} \left( |\sigma^{\frac{1}{2p}} X \sigma^{\frac{1}{2p}}|^p \right).$$

For  $p \geq 1$  these quantities are norms, while for  $p < 1$  only quasi-norms. They are closely related to the sandwiched Rényi divergences  $D_p$  of [59, 85], which can be defined as

$$D_p(\rho \parallel \sigma) = \frac{p}{p-1} \log(\|\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\|_{p,\sigma})$$

for  $p \in (0, 1) \cup (1, \infty)$  and are known to satisfy a data processing inequality for  $p \in (\frac{1}{2}, 1) \cup (1, \infty)$  [3]. Moreover, we have that as  $p \rightarrow 1$ , they converge to the usual relative entropy and that these entropies are monotone increasing in the parameter  $p$  [59]. The connection between these norms and Rényi divergences will allow us to obtain several conditions that imply bounds on the relative entropy contraction coefficients and the less noisy ordering based on the Petz recovery map of the channel. For a quantum channel  $\mathcal{M}$  and a reference state  $\sigma$ , define  $\Gamma_\sigma^p(X) = \sigma^{\frac{p}{2}} X \sigma^{\frac{p}{2}}$  with the convention  $\Gamma_\sigma = \Gamma_\sigma^1$  and the Petz recovery map as

$$\hat{\mathcal{M}} = \Gamma_\sigma \circ \mathcal{M}^* \circ \Gamma_{\mathcal{M}(\sigma)}^{-1}.$$

We omit the reference state  $\sigma$  in the definition, as it should always be clear from the context. One can readily check that restricting to the support of  $\sigma$  it is a quantum channel that satisfies  $\hat{\mathcal{M}} \circ \mathcal{M}(\sigma) = \sigma$  and it is known

that if  $D(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) = D(\rho\|\sigma)$ , then  $\hat{\mathcal{M}} \circ \mathcal{M}(\rho) = \rho$  [65]. That is, the relative entropy does not contract under  $\mathcal{M}$  if and only if one can reverse the action of the channel on two input states simultaneously with the Petz recovery map.

Our techniques will mostly be based on studying how the  $p, \sigma$  norms contract for different values of  $p$  and reference states  $\sigma$  under the Petz recovery map. Such statements are intimately related to strong data processing inequalities. Indeed, as shown in [57], the data-processing inequality for the Rényi divergences is equivalent to the statement that  $\|\hat{\mathcal{M}}^*\|_{(p,\sigma) \rightarrow (p,\mathcal{M}(\sigma))} \leq 1$  for all choices of input state  $\sigma$  and channels  $\mathcal{M}$ . The basic idea of this section is that if the inequality

$$\|\hat{\mathcal{M}}^*\|_{(p,\sigma) \rightarrow (q,\mathcal{M}(\sigma))} \leq 1 \quad (5.1)$$

holds for  $q > p$ , then we also obtain a strong data processing inequality for  $\mathcal{M}$  and some Rényi entropy. Note that this statement is stronger than the one with  $p = q$ , as the norms are monotone increasing in  $p$ . Denote by  $p', q'$  the Hölder conjugates of  $p, q$ . One can show by duality of norms that Eq. (5.1) is in fact equivalent to

$$\|\Gamma_\sigma^{-\frac{p'-1}{p'}} \circ \hat{\mathcal{M}}(\rho)\|_{p'} \leq \|\Gamma_{\mathcal{M}(\sigma)}^{-\frac{q'-1}{q'}}(\rho)\|_{q'}. \quad (5.2)$$

for all states  $\rho$ . As we will see in Proposition (5.4), the expression above is equivalent to an SDPI for Rényi entropies.

We start with the following standard Lemma:

**Lemma 5.1.** *For any two quantum states  $\rho, \sigma$  s.t.  $\text{supp } \rho \subset \text{supp } \sigma$  we have:*

$$\frac{d}{dp} \|\Gamma_\sigma^{-1}(\rho)\|_{p,\sigma} \Big|_{p=1} = D(\rho\|\sigma).$$

*Proof.* Note that

$$\frac{p}{p-1} \log(\|\Gamma_\sigma^{-1}(\rho)\|_{p,\sigma}) = D_p(\rho\|\sigma),$$

where  $D_p$  is the sandwiched Rényi divergence. Define the functions,

$$\phi(p) = p \log(\|\Gamma_\sigma^{-1}(\rho)\|_{p,\sigma}), \quad \varphi(p) = \|\Gamma_\sigma^{-1}(\rho)\|_{p,\sigma}.$$

As  $\lim_{p \rightarrow 1} D_p(\rho\|\sigma) = D(\rho\|\sigma)$  and  $\phi(1) = 0$  we have:

$$\frac{d}{dp} \phi(p) \Big|_{p=1} = D(\rho\|\sigma).$$

As  $\phi(p) = p \log(\varphi(p))$ , it follows that

$$\frac{d}{dp} \phi(p) = \log(\varphi(p)) + p \frac{\varphi'(p)}{\varphi(p)},$$

which yields the claim after noting that  $\varphi(1) = 1$ . □

The following characterization of a strong DPI then follows:

**Proposition 5.2.** *For some  $\sigma > 0$  and a quantum channel  $\mathcal{M}$ , let  $\hat{\mathcal{M}}$  be the Petz recovery map with respect to  $\sigma$ . Then the following are equivalent:*

1. *We have for all  $\rho \in \mathcal{D}(A)$*

$$D(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \leq \eta D(\rho\|\sigma) \quad (5.3)$$

2. *there exists an  $\epsilon > 0$  s.t. for all  $0 \leq \tau \leq \epsilon$  we have*

$$\|\hat{\mathcal{M}}^*\|_{(1+\eta\tau,\sigma) \rightarrow (1+\tau,\mathcal{M}(\sigma))} \leq 1 \quad (5.4)$$

3. there exists an  $\epsilon > 0$  s.t. for all  $0 \leq \tau \leq \epsilon$  we have

$$D_{\frac{\eta\tau+1}{\eta\tau}}(\hat{\mathcal{M}}(\rho)\|\sigma) \leq \frac{1+\eta\tau}{1+\tau} D_{\frac{\tau+1}{\tau}}(\rho\|\mathcal{M}(\sigma)) \quad (5.5)$$

*Proof.* Let us start by showing that (5.4) implies (5.3). Note that by setting  $X = \Gamma^{-1}(\rho)$ , (5.4) gives:

$$\|X\|_{1+\eta\tau,\sigma} \geq \|\Gamma_{\mathcal{M}(\sigma)}^{-1}(\mathcal{M}(\rho))\|_{1+\tau,\mathcal{M}(\sigma)}.$$

Taking the logarithm and rewriting the Equation above in terms of sandwiched Rényi divergences gives

$$\frac{(1+\tau)}{\eta(1+\eta\tau)} D_{1+\tau}(\mathcal{M}(\rho)\|\mathcal{M}(\sigma)) \leq D_{1+\eta\tau}(\rho\|\sigma)$$

and Eq. (5.9) follows by taking the limit  $\tau \rightarrow 0$ . Let us show the other direction. It follows from Lemma 5.1 and a Taylor expansion that

$$\begin{aligned} \|X_0\|_{1+\eta\tau,\sigma} &= 1 + \eta\tau D(\rho_0\|\sigma) + \mathcal{O}(\tau^2), \\ \|\Gamma_{\mathcal{M}(\sigma)}^{-1}(\mathcal{M}(\rho_0))\|_{1+\tau,\mathcal{M}(\sigma)} &= 1 + \tau D(\mathcal{M}(\rho_0)\|\mathcal{M}(\sigma)) + \mathcal{O}(\tau^2), \end{aligned}$$

from which it follows that

$$\|X_0\|_{1+\eta\tau,\sigma} \geq \|\Gamma_{\mathcal{M}(\sigma)}^{-1}(\mathcal{M}(\rho_0))\|_{1+\tau,\mathcal{M}(\sigma)}$$

for all  $\tau$  small enough. The claim follows from the fact that it suffices to restrict to positive operators when computing  $p \rightarrow q$  norms of completely positive maps [64]. Finally, Eq. (5.5) follows from a duality argument. Indeed, it is easy to show that Eq. (5.4) is equivalent to Eq. (5.2). Taking the logarithm, we obtain the inequality in terms of the Petz recovery map.  $\square$

Thus, we see that an SDPI is equivalent to a nontrivial  $p \rightarrow q$  inequality for the Petz recovery map for  $p, q$  for all  $p$  slightly larger than 1. We also immediately obtain a contractive characterization of the less noisy order:

**Corollary 5.3.** *Let  $\mathcal{M}, \mathcal{N}$  be two quantum channels. Then the following are equivalent:*

1.  $\mathcal{N} \succeq_{\text{l.n.}} \mathcal{M}$
2. For all  $\sigma \in \mathcal{D}(\mathcal{H})$  there and  $X > 0$  there exists an  $\epsilon$  such that for all  $\tau \leq \epsilon$  we have:

$$\|\hat{\mathcal{M}}^*(X)\|_{1+\tau,\mathcal{M}(\sigma)} \leq \|\hat{\mathcal{N}}^*(X)\|_{1+\tau,\mathcal{N}(\sigma)} \quad (5.6)$$

*Proof.* By rewriting the inequality in Eq. (5.6) in terms of the Rényi divergences and taking the limit  $\tau \rightarrow 0$  we see that they imply that

$$D(\mathcal{N}(\rho)\|\mathcal{N}(\sigma)) \geq D(\mathcal{M}(\rho)\|\mathcal{M}(\sigma))$$

for all  $\rho, \sigma$ . By Proposition 2.3, this is equivalent to  $\mathcal{N} \succeq_{\text{l.n.}} \mathcal{M}$ . To show the other direction we may again resort to a Taylor expansion.  $\square$

It is straightforward to adapt the results above to the completely less noisy ordering by suitably adapting the involved channels and states.

However, it might be difficult to compute the  $1+\tau$  norms analytically, as it is more convenient to work with integer values of  $p$ . Thus, we will show that any nontrivial  $p \rightarrow q$  inequality gives rise to an SDPI, possibly with an error term.

**Proposition 5.4.** *Suppose that for some  $1 < p \leq q < \infty$ ,  $C \geq 1$  and two channels  $\mathcal{M}, \mathcal{N}$ , where we assume that  $\mathcal{N}$  is invertible, we have*

$$\|\hat{\mathcal{M}}^* \circ (\hat{\mathcal{N}}^*)^{-1}\|_{(p, \mathcal{N}(\sigma)) \rightarrow (q, \mathcal{M}(\sigma))} \leq C. \quad (5.7)$$

and

$$\|\hat{\mathcal{M}}^* \circ (\hat{\mathcal{N}}^*)^{-1}\|_{(1, \mathcal{N}(\sigma)) \rightarrow (1, \mathcal{M}(\sigma))} \leq 1. \quad (5.8)$$

Then

$$D(\mathcal{M}(\rho) \|\mathcal{M}(\sigma)) \leq \frac{(p-1)q}{p(q-1)} D(\mathcal{N}(\rho) \|\mathcal{N}(\sigma)) + \frac{q}{q-1} \log(C) \quad \forall \rho \in \mathcal{D}(\mathcal{H}). \quad (5.9)$$

*Proof.* Combining Eq. (5.7), Eq. (5.8) and the Stein-Weiss interpolation theorem (see e.g. [57] for a quantum information friendly exposition) yields that:

$$\|\hat{\mathcal{M}}^* \circ (\hat{\mathcal{N}}^*)^{-1}\|_{(p_\theta, \mathcal{N}(\sigma)) \rightarrow (q_\theta, \mathcal{M}(\sigma))} \leq C^{1-\theta} \quad (5.10)$$

with

$$\frac{1}{p_\theta} = \theta + \frac{1-\theta}{p}, \quad \frac{1}{q_\theta} = \theta + \frac{1-\theta}{q}.$$

for  $0 \leq \theta \leq 1$ , which can be rewritten as:

$$\|\hat{\mathcal{M}}^*(X)\|_{(q_\theta, \mathcal{M}(\sigma))} \leq C^{1-\theta} \|\hat{\mathcal{N}}^*(X)\|_{(p_\theta, \mathcal{N}(\sigma))} \quad (5.11)$$

for all  $X$ . Solving for  $p_\theta$  and  $q_\theta$  we see that

$$p_\theta = \frac{p}{\theta(p-1)+1}, \quad q_\theta = \frac{q}{\theta(q-1)+1}.$$

Taking the logarithm of Eq. (5.11) and rewriting in terms of Rényi entropies we see that

$$D_{q_\theta}(\mathcal{M}(\rho) \|\mathcal{M}(\sigma)) \leq \frac{(p_\theta-1)q_\theta}{p_\theta(q_\theta-1)} D_{p_\theta}(\mathcal{N}(\rho) \|\mathcal{N}(\sigma)) + \frac{q_\theta(1-\theta)}{q_\theta-1} \log(C). \quad (5.12)$$

A close inspection of the expressions involved in the formula above shows that:

$$\frac{(p_\theta-1)q_\theta}{p_\theta(q_\theta-1)} = \frac{(p-1)q}{p(q-1)}, \quad \frac{q_\theta(1-\theta)}{q_\theta-1} = \frac{q}{q-1}.$$

The claim then follows by taking the limit  $\theta \rightarrow 1$  in Eq. (5.12) and noting that  $p_\theta \rightarrow 1$  and  $q_\theta \rightarrow 1$ .  $\square$

For any quantum channel  $\mathcal{M}$  we have

$$\|\hat{\mathcal{M}}^*\|_{(1, \sigma) \rightarrow (1, \mathcal{M}(\sigma))} \leq 1$$

and, thus, Eq.(5.8) is always satisfied for  $\mathcal{N} = \text{id}$ . That is, if we wish to prove an SDPI for a quantum channel with and additive error term, a non-trivial  $p \rightarrow q$  inequality suffices.

The case of  $2 \rightarrow 2$  norms of the Proposition above is particularly interesting, as it leads to entropic inequalities that tensorize:

**Corollary 5.5.** *Suppose that  $\mathcal{N}$  is invertible as a linear map. Then for some  $\sigma > 0$ ,*

$$\|\Gamma_{\mathcal{M}(\sigma)}^{-\frac{1}{2}} \circ \mathcal{M} \circ \mathcal{N}^{-1} \circ \Gamma_{\mathcal{N}(\sigma)}^{\frac{1}{2}}\|_{2 \rightarrow 2} \leq C. \quad (5.13)$$

implies that for all states  $\rho$

$$D_2(\mathcal{M}^{\otimes n}(\rho) \|\mathcal{M}(\sigma)^{\otimes n}) \leq D_2(\mathcal{N}^{\otimes n}(\rho) \|\mathcal{N}(\sigma)^{\otimes n}) + 2n \log(C). \quad (5.14)$$

*Proof.* Define  $\hat{\mathcal{M}} = \Gamma_{\mathcal{M}(\sigma)}^{-1} \circ \mathcal{M} \circ \Gamma_\sigma$  and  $\hat{\Psi} = \Gamma_{\mathcal{N}(\sigma)}^{-1} \circ \mathcal{N} \circ \Gamma_\sigma$ . Note that Eq. (5.13) is equivalent to

$$\|\Gamma_{\mathcal{M}(\sigma)}^{\frac{1}{2}} \circ \hat{\mathcal{M}} \circ \hat{\mathcal{N}}^{-1} \circ \Gamma_{\mathcal{N}(\sigma)}^{-\frac{1}{2}}\|_{2 \rightarrow 2} \leq C.$$

We can further massage this to

$$\|\Gamma_{\mathcal{M}(\sigma)}^{\frac{1}{2}} \circ \hat{\mathcal{M}}(X)\|_2 \leq C \|\Gamma_{\mathcal{N}(\sigma)}^{\frac{1}{2}} \circ \hat{\mathcal{N}}(X)\|_2. \quad (5.15)$$

for all operators  $X$ . To see this, set  $X = \hat{\mathcal{N}}^{-1} \circ \Gamma_{\mathcal{N}(\sigma)}^{-\frac{1}{2}}(Y)$ . Moreover, we have

$$\|\Gamma_{\mathcal{N}(\sigma)}^{\frac{1}{2}} \circ \hat{\mathcal{N}}(X)\|_2 = \|\hat{\mathcal{N}}(X)\|_{2, \mathcal{N}(\sigma)}, \quad \|\Gamma_{\mathcal{M}(\sigma)}^{\frac{1}{2}} \circ \hat{\mathcal{M}}(X)\|_2 = \|\hat{\mathcal{M}}(X)\|_{2, \mathcal{M}(\sigma)}.$$

Thus, we conclude that for all  $X$  we have:

$$\|\hat{\mathcal{M}}(X)\|_{2, \mathcal{M}(\sigma)} \leq C \|\hat{\mathcal{N}}(X)\|_{2, \mathcal{N}(\sigma)}.$$

Picking  $X = \sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}$  and taking the logarithm yields the claim for  $n = 1$ . The claim for other  $n > 1$  follows from noting that the condition in Eq. (5.13) tensorizes. Indeed, note that for  $\mathcal{N}^{\otimes n}$  and  $\mathcal{M}^{\otimes n}$  we have:

$$\begin{aligned} \|\Gamma_{\mathcal{M}(\sigma)^{\otimes n}}^{-\frac{1}{2}} \circ \mathcal{M}^{\otimes n} \circ (\mathcal{N}^{-1})^{\otimes n} \circ \Gamma_{\mathcal{N}(\sigma)^{\otimes n}}^{\frac{1}{2}}\|_{2 \rightarrow 2} &= \left\| \left( \Gamma_{\mathcal{M}(\sigma)}^{-\frac{1}{2}} \circ \mathcal{M} \circ \mathcal{N}^{-1} \circ \Gamma_{\mathcal{N}(\sigma)}^{\frac{1}{2}} \right)^{\otimes n} \right\|_{2 \rightarrow 2} \\ &= \left\| \left( \Gamma_{\mathcal{M}(\sigma)}^{-\frac{1}{2}} \circ \mathcal{M} \circ \mathcal{N}^{-1} \circ \Gamma_{\mathcal{N}(\sigma)}^{\frac{1}{2}} \right) \right\|_{2 \rightarrow 2}^n. \end{aligned}$$

The last equality follows from the fact that the  $2 \rightarrow 2$  norm just corresponds to the operator norm of the map, which is multiplicative.  $\square$

Note that the condition in Eq. (5.13) is equivalent to the operator norm of  $\Gamma_{\mathcal{M}(\sigma)}^{-\frac{1}{2}} \circ \mathcal{M} \circ \mathcal{N}^{-1} \circ \Gamma_{\mathcal{N}(\sigma)}^{\frac{1}{2}}$  being bounded by  $C$ . In particular, this means that it can be verified efficiently given the channels and a target state  $\sigma$ . Thus, we have identified a condition that can be verified efficiently and tensorizes.

This statement can also be used to obtain contraction coefficients that tensorize:

**Corollary 5.6.** *For some  $p > 0$  let  $\mathcal{N}_p$  be the channel  $\mathcal{N}_p(\rho) = p\rho + (1-p)\sigma$  and suppose that*

$$\|\Gamma_{\mathcal{M}(\sigma)}^{-\frac{1}{2}} \circ \mathcal{M} \circ \mathcal{N}_p^{-1} \circ \Gamma_\sigma^{\frac{1}{2}}\|_{2 \rightarrow 2} \leq 1 \quad (5.16)$$

*Then for all states  $\rho$*

$$D_2(\mathcal{M}^{\otimes n}(\rho) \| \mathcal{M}^{\otimes n}(\sigma^{\otimes n})) \leq \alpha(p, \sigma) D_2(\rho \| \sigma^{\otimes n})$$

*with  $\alpha(p, \sigma) = \exp\left((1 - \|\sigma^{-1}\|^{-1}) \frac{\log(p)}{\log(\|\sigma^{-1}\|)}\right)$ .*

*Proof.* By Prop. 5.5 we have that Eq. (5.16) implies that

$$D_2(\mathcal{M}^{\otimes n}(\rho) \| \mathcal{M}^{\otimes n}(\sigma^{\otimes n})) \leq D_2(\mathcal{N}_p^{\otimes n}(\rho) \| \mathcal{N}_p^{\otimes n}(\sigma^{\otimes n})).$$

In [56] the authors show that

$$D_2(\mathcal{N}_p^{\otimes n}(\rho) \| \mathcal{N}_p^{\otimes n}(\sigma^{\otimes n})) \leq \alpha(p, \sigma) D_2(\rho \| \sigma^{\otimes n}),$$

which completes the proof.  $\square$

The statement above can be used to bound capacities and entropic quantities of a quantum channel efficiently. For instance, if the channel  $\mathcal{M}$  is doubly stochastic, then we immediately see that if  $\|\mathcal{M} \circ \mathcal{N}_p^{-1}\|_{2 \rightarrow 2} \leq 1$ , then

$$D_2(\mathcal{M}^{\otimes n}(\rho) \| \frac{I}{n}) \leq \exp\left((1 - n^{-1}) \frac{\log(p)}{\log(n)}\right) D_2(\rho \| \frac{I}{n}).$$



By noting that  $D_2(\mathcal{M}^{\otimes n}(\rho) \parallel \frac{I}{n}) \geq n - S(\mathcal{M}^{\otimes n}(\rho))$ , we immediately obtain that the minimum output entropy of  $\mathcal{N}^{\otimes n}(\rho)$  is lower bounded by

$$S(\mathcal{M}^{\otimes n}(\rho)) \geq n \left( 1 - \exp \left( (1 - n^{-1}) \frac{\log(p)}{\log(n)} \right) \right).$$

Inspecting the expression  $\|\mathcal{M} \circ \mathcal{N}_p^{-1}\|_{2 \rightarrow 2}$  more closely, we see that:

$$\|\mathcal{M} \circ \mathcal{N}_p^{-1}\|_{2 \rightarrow 2} = \frac{1}{p} \|\mathcal{M} - (1 - p)\mathcal{N}_0\|_{2 \rightarrow 2}.$$

Thus, we see that  $\|\mathcal{M} \circ \mathcal{N}_p^{-1}\|_{2 \rightarrow 2} \leq 1$  if and only if the spectrum of  $\mathcal{M} - (1 - p)\mathcal{N}_0$  is contained in a ball of radius  $p$  around the origin in the complex plane. The largest value of  $p$  for which this holds is called the spectral gap of  $\mathcal{M}$ .

## 6 Generalized contraction coefficients and partial orders

In this section we will discuss extensions of the concepts discussed to other divergences, in particular  $f$ -divergences and  $\chi^2$ -divergences. Note that both partial orders and contraction coefficients can be naturally defined in terms of any divergence by simply replacing the relative entropy by the desired divergence.

Interestingly, classically contraction coefficients (and partial orders respectively) based on many different divergences have been shown actually to be the same, see e.g. [52]. In the quantum case this remains a mostly open problem with some of what is known being summarized in the following section.

### 6.1 f-divergences

Here, we extend the discussions of the previous sections to the setting of arbitrary  $f$ -divergences. The two most commonly used ones are the standard  $f$ -divergence: given  $\rho, \sigma \in \mathcal{D}(\mathcal{H})$  with  $\sigma$  full-rank:

$$D_f(\rho \parallel \sigma) := \text{Tr } \sigma^{1/2} f(\Delta_{\rho, \sigma}) (\sigma^{1/2}), \quad (6.1)$$

where  $\Delta_{\rho, \sigma}(X) := \rho X \sigma^{-1}$  is the so-called relative modular operator between states  $\sigma$  and  $\rho$  and can be interpreted as a non-commutative generalization of the Radon-Nikodym derivative between two probability mass functions. Similarly, the maximal  $f$ -divergence is defined as

$$\widehat{D}_f(\rho \parallel \sigma) := \text{Tr } \sigma f(\sigma^{-1/2} \rho \sigma^{-1/2}). \quad (6.2)$$

Note that, often, only those functions that obey  $f(1) = 0$  are considered valid for  $f$ -divergences as it ensures that  $D_f(\rho \parallel \rho) = 0$ . However, this excludes e.g. Rényi divergences.

Of course, we can define contraction coefficients  $\eta_{D_f}(\mathcal{N})$  and  $\eta_{\widehat{D}_f}(\mathcal{N})$  as for other divergences. Next, we want to consider mutual information like quantities based on  $f$ -divergences. Consider the following definitions,

$$I_f(A : B) := D_f(\rho_{AB} \parallel \rho_A \otimes \rho_B), \quad (6.3)$$

$$\widehat{I}_f(A : B) := \widehat{D}_f(\rho_{AB} \parallel \rho_A \otimes \rho_B). \quad (6.4)$$

In what follows, we denote  $\widetilde{I}_f$ , resp.  $\widetilde{D}_f$  for either  $I_f$  or  $\widehat{I}_f$ , resp.  $D_f$  or  $\widehat{D}_f$ . Note, that just like for the relative entropy, the following holds. Consider a classical quantum state  $\rho_{UB}$  and its marginal  $\rho_B$ , we have

$$\widetilde{I}_f(U : B) = \sum_u p(u) \widetilde{D}_f(\rho_B^u \parallel \rho_B). \quad (6.5)$$

As a direct consequence, for a quantum channel  $\mathcal{N}$ , we have

$$\widetilde{I}_f(U : B) = \sum_u p(u) \widetilde{D}_f(\rho_B^u \parallel \rho_B) \leq \eta_{\widetilde{D}_f}(\mathcal{N}) \sum_u p(u) \widetilde{D}_f(\rho_A^u \parallel \rho_A) = \eta_{\widetilde{D}_f}(\mathcal{N}) \widetilde{I}_f(U : A), \quad (6.6)$$

resulting in

$$\eta_{\tilde{D}_f}(\mathcal{N}) \geq \sup_{\rho_{UA}} \frac{\tilde{I}_f(U; B)}{\tilde{I}_f(U; A)}. \quad (6.7)$$

Before proving the equivalence of Proposition 6.2, we recall this very simple technical lemma:

**Lemma 6.1.** *Let  $A$  be a self-adjoint operator over a finite dimensional Hilbert space, and let  $f$  be a function on  $\mathbb{R}_+$  that is differentiable at  $1$ . Then*

$$Df[1](A) = f'(1)A.$$

*Proof.* Simple write the eigenvalue decomposition of  $A := \sum_i a_i |i\rangle\langle i|$ , so that

$$Df[1](A) = \lim_{\varepsilon \rightarrow 0} \frac{f(A + \varepsilon \mathbb{1}) - f(A)}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \sum_i \frac{f(a_i + \varepsilon) - f(a_i)}{\varepsilon} |i\rangle\langle i| = f'(1) A.$$

□

The next Proposition is a direct generalization of Theorem V.2 in [69]

**Proposition 6.2.** *Let  $f$  be differentiable, and let  $\mathcal{M} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$  and  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_{B'})$ ,  $\sigma_A \in \mathcal{D}(\mathcal{H}_A)$  and  $\eta \geq 0$ . The following are equivalent:*

(i) *For all c-q state  $\rho_{UA}$  with marginal  $\sigma_A$ , where  $U$  is an arbitrary classical system,*

$$\eta \tilde{I}_f(U; B)_{(\text{id}_U \otimes \mathcal{M})(\rho_{UA})} \geq \tilde{I}_f(U; B')_{(\text{id}_U \otimes \mathcal{N})(\rho_{UA})}.$$

(ii) *For any state  $\rho_A \in \mathcal{D}(\mathcal{H}_A)$  with  $\text{supp}(\rho_A) \subseteq \text{supp}(\sigma_A)$ ,*

$$\eta \tilde{D}_f(\mathcal{M}(\rho_A) \| \mathcal{M}(\sigma_A)) \geq \tilde{D}_f(\mathcal{N}(\rho_A) \| \mathcal{N}(\sigma_A)). \quad (6.8)$$

Therefore,  $\Phi \succeq_{1.n.} \Psi$  if and only if Equation (6.8) holds for  $\eta = 1$ .

*Proof.* We first prove (ii)  $\Rightarrow$  (i): for any  $U \sim p_U$  and conditional states  $\{\rho_{A^u}\} \in \mathcal{D}(\mathcal{H}_A)$ , we have

$$\begin{aligned} \tilde{I}_f(U; B) &= \sum_u P_U(u) \tilde{D}_f(\mathcal{M}(\rho_A^u) \| \mathcal{M}(\sigma)) \\ \tilde{I}_f(U; B') &= \sum_u P_U(u) \tilde{D}_f(\mathcal{N}(\rho_A^u) \| \mathcal{N}(\sigma)), \end{aligned}$$

where  $\sigma = \sum_u P_U(u) \rho_A^u$  is defined to be the marginal of  $\rho_{AU}$ . Then the result follows directly from (ii). Next, we prove that (i)  $\Rightarrow$  (ii): for any  $\rho_A \in \mathcal{D}(\mathcal{H}_A)$ , and  $0 \leq \lambda \leq 1$ , we let  $U \equiv U_\lambda$  be the binary random variable of distribution  $P_{U_\lambda(0)} = \lambda$ ,  $P_{U_\lambda(1)} = 1 - \lambda$ , and conditional states  $\rho_A^0 = \rho_A$ ,  $\rho_A^1 = \sigma_A$ . Then, let

$$\begin{aligned} \varphi(\lambda) &:= \eta \tilde{I}_f(U_\lambda; B) - \tilde{I}_f(U_\lambda; B') \\ &= \eta \lambda \tilde{D}_f(\mathcal{M}(\rho_A) \| \mathcal{M}(\bar{\rho}_A)) + \eta (1 - \lambda) \tilde{D}_f(\mathcal{M}(\sigma) \| \mathcal{M}(\bar{\rho}_A)) - \lambda \tilde{D}_f(\mathcal{N}(\rho_A) \| \mathcal{N}(\bar{\rho}_A)) - (1 - \lambda) \tilde{D}_f(\mathcal{N}(\sigma) \| \mathcal{N}(\bar{\rho}_A)), \end{aligned}$$

where  $\bar{\rho}_A = \lambda \rho_A + (1 - \lambda) \sigma$ . Since  $\bar{\rho}_A = \sigma$  when  $\lambda = 0$ , we have that  $\varphi(0) = 0$ . Moreover, (i) implies that  $\varphi(\lambda) \geq 0$  for any  $\lambda \geq 0$ . Therefore  $\varphi'(0) \geq 0$ . The result follows after computing the latter derivative.

$$\varphi'(0) = \eta \tilde{D}_f(\mathcal{M}(\rho_A) \| \mathcal{M}(\sigma_A)) - \tilde{D}_f(\mathcal{N}(\rho_A) \| \mathcal{N}(\sigma_A)) + \left. \frac{d}{d\lambda} \right|_{\lambda=0} \eta \tilde{D}_f(\mathcal{M}(\sigma) \| \mathcal{M}(\bar{\rho}_A)) - \tilde{D}_f(\mathcal{N}(\sigma) \| \mathcal{N}(\bar{\rho}_A)),$$

We will be done as soon as we can show that the derivative above is equal to 0. By linearity of the channels  $\mathcal{M}$  and  $\mathcal{N}$ , it is enough to show that

$$\left. \frac{d}{d\lambda} \right|_{\lambda=0} \tilde{D}_f(\sigma \| \bar{\rho}) = 0.$$

We first focus the case of the maximal  $f$ -divergences: we have

$$\left. \frac{d}{d\lambda} \right|_{\lambda=0} \widehat{D}_f(\sigma \| \bar{\rho}) = \text{Tr} \left[ \sigma Df[\mathbb{1}] \left( \left. \frac{d}{d\lambda} \right|_{\lambda=0} \bar{\rho}^{-\frac{1}{2}} \sigma \bar{\rho}^{-\frac{1}{2}} \right) \right] \quad (6.9)$$

$$= f'(1) \left. \frac{d}{d\lambda} \right|_{\lambda=0} \text{Tr} [\sigma \bar{\rho}^{-\frac{1}{2}} \sigma \bar{\rho}^{-\frac{1}{2}}] \quad (6.10)$$

$$= 2f'(1) \left. \frac{d}{d\lambda} \right|_{\lambda=0} \text{Tr} [\sigma^{\frac{3}{2}} \bar{\rho}^{-\frac{1}{2}}] \quad (6.11)$$

where the second identity comes from Lemma 6.1. Next, we use the integral representation of the square root (see [8])

$$t^{\frac{1}{2}} = \frac{1}{\pi} \int_0^\infty \frac{st}{s+t} s^{-\frac{3}{2}} ds,$$

so that

$$\begin{aligned} \left. \frac{d}{d\lambda} \right|_{\lambda=0} \bar{\rho}^{-\frac{1}{2}} &= -\sigma^{-\frac{1}{2}} \left( \left. \frac{d}{d\lambda} \right|_{\lambda=0} \bar{\rho}^{\frac{1}{2}} \right) \sigma^{-\frac{1}{2}} \\ &= -\frac{1}{\pi} \sigma^{-\frac{1}{2}} \int_0^\infty \left. \frac{d}{d\lambda} \right|_{\lambda=0} \frac{s\bar{\rho}}{s+\bar{\rho}} s^{-\frac{3}{2}} ds \sigma^{-\frac{1}{2}} \\ &= -\frac{1}{\pi} \sigma^{-\frac{1}{2}} \int_0^\infty s(\rho - \sigma)(s + \sigma)^{-1} s^{-\frac{3}{2}} ds \sigma^{-\frac{1}{2}} + \frac{1}{\pi} \sigma^{-\frac{1}{2}} \int_0^\infty s\sigma(s + \sigma)^{-1}(\rho - \sigma)(s + \sigma)^{-1} s^{-\frac{3}{2}} ds \sigma^{-\frac{1}{2}} \\ &= -\sigma^{-\frac{1}{2}}(\rho - \sigma)\sigma^{-1}\sigma^{\frac{1}{2}}\sigma^{-\frac{1}{2}} + \frac{1}{\pi} \sigma^{-\frac{1}{2}} \int_0^\infty s\sigma(s + \sigma)^{-1}(\rho - \sigma)(s + \sigma)^{-1} s^{-\frac{3}{2}} ds \sigma^{-\frac{1}{2}}. \end{aligned}$$

Therefore,

$$\begin{aligned} \left. \frac{d}{d\lambda} \right|_{\lambda=0} \widehat{D}_f(\sigma \| \bar{\rho}) &= -2f'(1) \text{Tr}[\rho - \sigma] + \frac{2f'(1)}{\pi} \int_0^\infty \text{Tr} [\sigma^2(s + \sigma)^{-1}(\rho - \sigma)(s + \sigma)^{-1} \sigma^{-\frac{1}{2}}] s^{-\frac{1}{2}} ds \\ &= \frac{2f'(1)}{\pi} \int_0^\infty \text{Tr} [\sigma^{\frac{3}{2}}(s + \sigma)^{-2}(\rho - \sigma)] s^{-\frac{1}{2}} ds \\ &= f'(1) \text{Tr}[(\rho - \sigma)] \\ &= 0, \end{aligned}$$

which is what we needed to prove. The case of the standard  $f$ -divergence can be treated very similarly:

$$\begin{aligned} \left. \frac{d}{d\lambda} \right|_{\lambda=0} D_f(\sigma \| \bar{\rho}) &= 2 \left. \frac{d}{d\lambda} \right|_{\lambda=0} \text{Tr} \sigma^{\frac{1}{2}} f(L_\sigma R_{\sigma^{-1}})(\bar{\rho}^{\frac{1}{2}}) + \text{Tr} \sigma^{\frac{1}{2}} \left. \frac{d}{d\lambda} \right|_{\lambda=0} f(L_\sigma R_{\bar{\rho}^{-1}}) \sigma^{\frac{1}{2}} \\ &= 2 \left. \frac{d}{d\lambda} \right|_{\lambda=0} \text{Tr} \sigma^{\frac{1}{2}} \bar{\rho}^{\frac{1}{2}} + \text{Tr} \sigma^{\frac{1}{2}} Df[\text{id}] \left( \left. \frac{d}{d\lambda} \right|_{\lambda=0} L_\sigma R_{\bar{\rho}^{-1}} \right) \sigma^{\frac{1}{2}} \\ &\stackrel{(1)}{=} -f'(1) \text{Tr} \sigma^{\frac{1}{2}} \sigma^{\frac{3}{2}} \sigma^{-1}(\rho - \sigma) \sigma^{-1} \\ &= 0. \end{aligned}$$

where in (1) we used once again the integral representation of the square root in order to cancel the first trace above. □

**Remark 6.3.** *The previous result can be easily extended to the case of optimized quantum  $f$  divergences as defined in [83].*

**Remark 6.4.** *Note that we only require  $f$  to be differentiable. Thus, Proposition 6.2 holds for a larger class of functionals than  $f$ -divergences.*

As a result we can express the contraction coefficient based on  $f$ -divergences in terms of the corresponding  $f$ -mutual information

$$\eta_{\tilde{D}_f}(\mathcal{N}) = \sup_{\rho_{U:A}} \frac{\tilde{I}_f(U : B)}{\tilde{I}_f(U : A)}, \quad (6.12)$$

as well as define partial orders based on  $f$ -mutual informations that can be equivalently formulated in terms of  $f$ -divergences.

Let us end this section by discussing our favorite example: the erasure channel. One can easily find the following.

**Lemma 6.5.** *For any function  $f$  with  $f(1) = 0$ , we have*

$$D_f(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) = (1 - \epsilon) D_f(\rho \| \sigma), \quad (6.13)$$

$$\hat{D}_f(\mathcal{E}(\rho) \| \mathcal{E}(\sigma)) = (1 - \epsilon) \hat{D}_f(\rho \| \sigma), \quad (6.14)$$

which immediately implies that for all functions  $f$  and  $g$  with the above property, we have

$$\eta_{D_f}(\mathcal{E}) = \eta_{\hat{D}_g}(\mathcal{E}) = \eta_{\text{Re}}(\mathcal{E}) = 1 - \epsilon. \quad (6.15)$$

*Proof.* By direct calculation. □

This implies that we can easily extend our previous Proposition 2.5.

**Proposition 6.6.** *Let  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$ ,  $\eta_e \in [0, 1]$  and  $f$  any function with  $f(1) = 0$ . Then the following are equivalent:*

$$(i) \ \mathcal{M}_{\eta_e}^{\text{er}} \succeq_{D_f, \text{l.n.}} \mathcal{N}.$$

$$(ii) \ \eta_{D_f}(\mathcal{N}) \leq \eta_{D_f}(\mathcal{M}_{\eta_e}^{\text{re}}) = (1 - \eta_e).$$

*Proof.* Analog to that of Proposition 2.5. □

And the same holds for the maximal  $f$ -divergence.

## 6.2 Spectral approaches to quantum less-noisy pre-orders

In Theorem 1 of [50], it was shown in the classical setting that the less noisy pre-order  $\mathcal{M} \succeq_{\text{l.n.}} \Psi$  is equivalent to the following condition: for any probability mass functions  $p, q$

$$\chi^2(\mathcal{M}(p), \mathcal{M}(q)) \geq \chi^2(\mathcal{N}(p), \mathcal{N}(q)),$$

where the  $\chi^2$  divergence is defined as follows:

$$\chi^2(p, q) := \sum_i \frac{(p(i) - q(i))^2}{q(i)}. \quad (6.16)$$

This characterization is powerful as it allows for a spectral analysis of the less noisy pre-order [50]. The proof of this fact is two-fold: to prove that less noisiness implies (6.16), it uses the characterization of less noisy seen in Section 2.2, namely that  $\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}$  is equivalent to

$$D(\mathcal{M}(p) \| \mathcal{M}(q)) \geq D(\mathcal{N}(p) \| \mathcal{N}(q)), \quad (6.17)$$

for any two probability mass functions  $p$  and  $q$ , together with the well-known fact that the  $\chi^2$  divergence locally approximates the relative entropy:

$$\lim_{\lambda \rightarrow 0^+} \frac{2}{\lambda^2} D(\lambda p + (1 - \lambda)q \| q) = \chi^2(p, q). \quad (6.18)$$

On the other hand, the relative entropy can be rewritten in terms of the following integral [50] (see also [54, 60, 61] for slightly different related expressions):

$$D(p\|q) := \int_0^\infty \frac{\chi^2(p, \frac{tp+q}{1+t})}{1+t} dt. \quad (6.19)$$

The characterization of less noisiness in terms Equation (6.16) follows from the joint use of Equations (6.18) and (6.19). This result implies as a special case the following result [15] (see also refinements, e.g. Theorem 3.6 of [69], and [69, 1] for the link to the contraction coefficient for the maximal correlation):

$$\eta_{\text{Re}}(\mathcal{M}) = \eta_{\chi^2}(\mathcal{M}) := \sup_{\substack{p, q \\ 0 < \chi^2(p, q) < \infty}} \frac{\chi^2(\mathcal{M}(p), \mathcal{M}(q))}{\chi^2(p, q)}. \quad (6.20)$$

The  $\chi^2$  quantity is known classically to locally approximate any  $f$  divergence [11, 14]. This is no longer the case in the non-commutative world, where a complete characterization of  $\chi^2$  quantities, also known as Fisher information metrics, was found [47, 74]. First of all, we define the following set of functions:

$$\mathcal{K} := \{k; -k \text{ is operator monotone}, k(w^{-1}) = wk(w), k(1) = 1\}.$$

Next, for  $\rho, \sigma \in \mathcal{D}(\mathcal{H})$ , and  $k \in \mathcal{K}$ , define the quantum  $\chi^2$ -divergence

$$\chi_k^2(\rho, \sigma) := \langle \rho - \sigma, \Omega_\sigma^k(\rho - \sigma) \rangle,$$

when  $\text{supp}(\rho) \subseteq \text{supp}(\sigma)$ , and infinity otherwise. In the previous case, the inversion  $\Omega_\sigma^k$  is defined as

$$\Omega_\sigma^k := R_\sigma^{-1} k(\Delta_{\sigma, \sigma}).$$

Here, we also restrict ourselves to a subclass of  $f$ -divergences: consider the class  $\mathcal{G}$  of continuous operator convex functions  $g$  from  $\mathbb{R}_+$  to  $\mathbb{R}$  that satisfy  $g(1) = 0$ . These functions can all be expressed in terms of the following integral representation:

$$g(w) = a(w-1) + b(w-1)^2 + c \frac{(w-1)^2}{w} + \int_0^\infty \frac{(w-1)^2}{w+s} d\nu(s), \quad (6.21)$$

where  $a, b, c > 0$  and the integral of the positive measure  $d\nu(s)$  on  $(0, \infty)$  is bounded. Then, for any  $k \in \mathcal{K}$ , there exists  $g \in \mathcal{G}$  such that [47, 74]:

$$\chi_k^2(\rho, \sigma) = - \frac{\partial^2}{\partial \alpha \partial \beta} D_g(\sigma + \alpha(\rho - \sigma) \| \sigma + \beta(\rho - \sigma)) \Big|_{\alpha=\beta=0}. \quad (6.22)$$

The function  $k$  is related to  $g$  by

$$k(w) := \frac{g(w) + wg(w^{-1})}{(w-1)^2}. \quad (6.23)$$

Hence, like its classical restriction, the quantum  $\chi^2$  divergence can be understood as a local approximation of the quantum relative entropy. However, Equation (6.19) is not known to hold in general in the quantum case. Rather, the following integral representation of the quantum relative entropy is known and has been recently used to get tightenings of the data processing inequality in [13] as a special case of the integral representation (6.21):

$$D(\rho\|\sigma) = \int_0^\infty \text{Tr} [(t + \Delta_{\sigma, \rho})^{-1}(\rho)] - \frac{1}{1+t} dt, \quad (6.24)$$

It can be easily shown that the integrands in Equation (6.24) and Equation (6.19) coincide when  $\rho$  and  $\sigma$  commute and can be identified with  $p$  and  $q$ . However, the integrand (6.24) cannot be identified in general with an adequately normalized  $\chi^2$  divergence. In particular, Equation (6.20) is not known to hold in general

in the quantum setting, despite some recent progress in that direction [46, 34]. We also refer to the more recent paper [12] where the strong data processing for the  $\chi^2$  divergence was shown to tensorize in the case where the state  $\sigma$  is taken to be a tensor product.

For the reasons mentioned above, it seems unlikely that the equivalence between less noisy and the ordering of  $\chi^2$  is true in the quantum setting. However, the following result is a direct consequence of the above discussion:

**Proposition 6.7.** *For any two quantum channels  $\mathcal{M} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_B)$ ,  $\mathcal{N} \in \text{CPTP}(\mathcal{H}_A, \mathcal{K}_{B'})$  and  $\eta \geq 0$ , (i)  $\Rightarrow$  (ii)  $\Rightarrow$  (iii), where:*

(i) *For any two quantum states  $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$  and any  $t \geq 0$ :*

$$\begin{aligned} \text{Tr } \mathcal{N}(\sigma)^{\frac{1}{2}} (\Delta_{\mathcal{N}(\rho), \mathcal{N}(\sigma)} - \text{id})^2 (\mathcal{N}(\sigma)^{\frac{1}{2}}) &\leq \eta \text{Tr } \mathcal{M}(\sigma)^{\frac{1}{2}} (\Delta_{\mathcal{M}(\rho), \mathcal{M}(\sigma)} - \text{id})^2 (\mathcal{M}(\sigma)^{\frac{1}{2}}); \quad \text{and} \\ \text{Tr } \mathcal{N}(\sigma)^{\frac{1}{2}} \frac{(\Delta_{\mathcal{N}(\rho), \mathcal{N}(\sigma)} - \text{id})^2}{\Delta_{\mathcal{N}(\rho), \mathcal{N}(\sigma)} + t} (\mathcal{N}(\sigma)^{\frac{1}{2}}) &\leq \eta \text{Tr } \mathcal{M}(\sigma)^{\frac{1}{2}} \frac{(\Delta_{\mathcal{M}(\rho), \mathcal{M}(\sigma)} - \text{id})^2}{\Delta_{\mathcal{M}(\rho), \mathcal{M}(\sigma)} + t} (\mathcal{M}(\sigma)^{\frac{1}{2}}). \end{aligned}$$

(ii) *For all  $g \in \mathcal{G}$  and any two states  $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$ :*

$$D_g(\mathcal{N}(\rho) \| \mathcal{N}(\sigma)) \leq \eta D_g(\mathcal{M}(\rho) \| \mathcal{M}(\sigma)).$$

(iii) *For any  $k \in \mathcal{K}$  and any two quantum states  $\rho, \sigma \in \mathcal{D}(\mathcal{H}_A)$ ,*

$$\chi_k^2(\mathcal{N}(\rho), \mathcal{N}(\sigma)) \leq \eta \chi_k^2(\mathcal{M}(\rho), \mathcal{M}(\sigma)).$$

Moreover, if (ii) holds only for a given  $g \in \mathcal{G}$ , then (iii) holds for  $k \in \mathcal{K}$  satisfying (6.23). Finally, obvious extensions of the above chain of implications hold true in the case of regularized and complete less noisy pre-orders, as well as for the related contraction coefficients.

*Proof.* This is a direct consequence of the expression (6.21) as well as the relation (6.22).  $\square$

Finally, we will once more mention the erasure channel as an illustrating example.

**Lemma 6.8.** *Let  $\mathcal{E} : A \rightarrow B$  be the quantum erasure channel of parameter  $\epsilon$ ,  $\rho_A$  and  $\sigma_A$  be some quantum states. Then for any  $k \in \mathcal{K}$  we have that*

$$\chi_k^2(\mathcal{E}(\rho), \mathcal{E}(\sigma)) = (1 - \epsilon) \chi_k^2(\rho, \sigma). \quad (6.25)$$

Therefore,

$$\eta_{\text{Re}}(\mathcal{E}) = \eta_{\chi^2}(\mathcal{E}) = 1 - \epsilon.$$

This means that at least for the erasure channel the contraction coefficients for relative entropy, f-divergences and  $\chi_k^2$  divergence are the same.

## 7 Functional inequalities

Another standard way of obtaining strong data processing is through functional inequalities [68], which we will describe in a bit more detail below. This approach is particularly effective whenever we have a semigroup  $(\mathcal{P}_t)_{t \geq 0}$  converging to a unique full rank state  $\sigma$ . Such semigroups are called primitive. In [50] the authors relate such functional inequalities to the less noisy ordering, especially when comparing to symmetric channels. Here we will recover their results in the quantum setting, especially Proposition 12, while also revisiting the connections between functional and entropic inequalities.

We will start by defining logarithmic Sobolev inequalities and showing their consequences. Let  $\mathcal{N}$  be a primitive channel whose unique invariant state is  $\sigma$ , and denote by  $\mathcal{N}^*$  the corresponding completely, positive

unital dual map. Corresponding to the map  $\mathcal{N}^*$ , we construct a quantum Markov semigroup (QMS) of completely positive unital maps  $\mathcal{P}_t = e^{-t(\text{id} - \mathcal{N}^*)}$  for all  $t \geq 0$ . In other words, the resulting semigroup  $(\mathcal{P}_t)_{t \geq 0}$  has  $\mathcal{S} := \mathcal{N}^* - \text{id}$  as its generator. The semigroup  $(\mathcal{P}_t)_{t \geq 0}$  is by construction also primitive, and  $\sigma$  is its unique invariant state. We can then define a scalar product

$$\langle X, Y \rangle_\sigma := \text{Tr}(\sigma^{\frac{1}{2}} X^\dagger \sigma^{\frac{1}{2}} Y)$$

which induces the  $\|\cdot\|_{2,\sigma}$  norm. Next we define the Dirichlet form  $\mathcal{E}_\mathcal{S}$  corresponding to  $\mathcal{S}$  as follows: for all  $X, Y \in \mathcal{B}_{sa}(\mathcal{H})$ ,

$$\mathcal{E}_\mathcal{S}(X, Y) = -\frac{1}{2} \langle X, \mathcal{S} + \hat{\mathcal{S}}(Y) \rangle_\sigma \equiv \langle X, (\mathbf{1} - \frac{\mathcal{N}^* + \hat{\mathcal{N}}^*}{2})(Y) \rangle_\sigma,$$

where Next, we define the 2-entropy as follows: for any  $X \in \mathcal{B}_{sa}(\mathcal{H})$ ,

$$\text{Ent}_{2,\sigma}(X) := D(|\sigma^{\frac{1}{4}} X \sigma^{\frac{1}{4}}|^2 \| \sigma) - \text{Tr}(|\sigma^{\frac{1}{2}} X \sigma^{\frac{1}{2}}|^2) \ln \text{Tr}(|\sigma^{\frac{1}{2}} X \sigma^{\frac{1}{2}}|^2),$$

where here the relative entropy  $D(\cdot \| \cdot)$  has been extended to non-normalized positive operators  $D(A \| B) = \text{Tr}(A(\ln A - \ln B))$  whenever  $\text{supp}(A) \subset \text{supp}(B)$ . The QMS  $(\mathcal{P}_t)_{t \geq 0}$  satisfies the *logarithmic Sobolev inequality* (LSI) if there exists  $\alpha > 0$  such that for every  $X \in \mathcal{B}_{sa}(\mathcal{H})$

$$\alpha \text{Ent}_{2,\sigma}(X) \leq \mathcal{E}_\mathcal{S}(X, X). \quad (7.1)$$

The largest  $\alpha$  such that Equation (7.1) holds

$$\alpha(\mathcal{N}^*) := \inf_{\substack{X \in \mathcal{B}_{sa}(\mathcal{H}) \\ \text{Ent}_{2,\sigma}(X) \neq 0}} \frac{\mathcal{E}_\mathcal{S}(X, X)}{\text{Ent}_{2,\sigma}(X)}$$

is called the *logarithmic Sobolev constant* of  $(\mathcal{P}_t)_{t \geq 0}$ . Likewise, the logarithmic Sobolev inequality for the *discrete-time* Markov chain with constant  $\alpha > 0$  states that for every  $X \in \mathcal{B}_{sa}(\mathcal{H})$ ,

$$\alpha \text{Ent}_{2,\sigma}(X) \leq \mathcal{E}_{\text{id} - \mathcal{N}^* \circ \hat{\mathcal{N}}^*}(X, X),$$

where  $\mathcal{E}_{\text{id} - \mathcal{N}^* \circ \hat{\mathcal{N}}^*}(X, Y) := \langle X, (\text{id} - \mathcal{N}^* \hat{\mathcal{N}}^*)(Y) \rangle_\sigma$  is sometimes referred to as the discrete Dirichlet form. Logarithmic Sobolev inequalities are particularly useful for reversible semigroups, i.e. semigroups such that their Petz recovery map with respect to the invariant state is the semigroup itself. In this case, it is known that a LSI implies the following hypercontractive inequality:

$$\|\mathcal{P}_t^*\|_{2 \rightarrow q(t)} \leq 1, \quad q(t) = 1 + e^{2\alpha t}. \quad (7.2)$$

By the results of Proposition 5.4, we know that this can be used to obtain strong data processing inequalities involving Rényi divergences. Indeed, we immediately obtain:

$$D_2(\mathcal{P}_t(\rho) \| \sigma) \leq \frac{1 + e^{-2\alpha t}}{2} D_{q(t)}(\rho \| \sigma).$$

Nevertheless, the inequality above is a bit unsatisfactory, as it does not give complete contraction even as  $t \rightarrow \infty$ . But this can be easily remedied by resorting to reversibility. As the semigroup is reversible, Eq. (7.2) also implies

$$D_2(\mathcal{P}_t(\rho) \| \sigma) \leq \frac{2e^{-2\alpha t}}{1 + e^{-2\alpha t}} D_{1+e^{-\alpha t}}(\rho \| \sigma). \quad (7.3)$$

by duality. By Proposition 5.4, we conclude that:

$$D(\mathcal{P}_t(\rho) \| \sigma) \leq \frac{2e^{-2\alpha t}}{1 + e^{-2\alpha t}} D(\rho \| \sigma).$$

Exploiting the fact that  $\frac{2e^{-2\alpha t}}{1+e^{-2\alpha t}} \leq e^{-\alpha t}$  we then get an exponential decay of the relative entropy from the logarithmic Sobolev inequality. Although this result is by no means new [29, 63, 41], we believe that the approach presented here showcases transparently how a LSI implies several different entropic inequalities. Computing the LSI constant is, in general, an arduous task. For the depolarizing semigroup in dimension  $d$ , it is known that:

$$\alpha = \frac{2(1 - 2d^{-1})}{\log(d - 1)}. \quad (7.4)$$

It is possible to obtain estimates of LSI through 7.4 for any doubly stochastic channel through so-called comparison techniques. The same is true for generalized depolarizing channels.

Indeed, an easy way of estimating the LSI constant of a semigroup assuming that we know the LSI constant of another QMS is via the comparison of their Dirichlet forms: assume that  $\mathcal{S}_1$  and  $\mathcal{S}_2$  have the same unique invariant state  $\sigma$ , then:

$$\exists \lambda > 0, \forall X \in \mathcal{B}_{sa}(\mathcal{H}), \mathcal{E}_{\mathcal{S}_1}(X, X) \leq \lambda \mathcal{E}_{\mathcal{S}_2}(X, X) \Rightarrow \alpha(\mathcal{S}_2) \geq \frac{\alpha(\mathcal{S}_1)}{\lambda}.$$

In the next theorem, we show that the information-theoretic notion of less noisy domination is a sufficient condition for the comparison of Dirichlet forms. First, we recall the following Lemma from [50]:

**Lemma 7.1.** *Given a positive semi-definite real matrix  $A$  and a normal real matrix  $B$ ,*

$$A^2 = AA^* \geq_{\text{PSD}} BB^T \Rightarrow A = \frac{A + A^T}{2} \geq_{\text{PSD}} \frac{B + B^T}{2}.$$

From this we get:

**Theorem 7.2.** *Let  $\mathcal{N}, \mathcal{M}$  be two primitive quantum channels, with a common unique full-rank invariant state  $\sigma$ . Then the following are true:*

(i) *If  $\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}$ , then:*

$$\forall X \in \mathcal{B}(\mathcal{H})_+, \mathcal{E}_{\text{id}-\mathcal{N}^*\hat{\mathcal{N}}^*}(X, X) \geq \mathcal{E}_{\text{id}-\mathcal{M}^*\hat{\mathcal{M}}^*}(X, X),$$

*or equivalently*

$$D_2(\mathcal{N}(\rho) \parallel \sigma) \leq D_2(\mathcal{M}(\rho) \parallel \sigma)$$

*for all states  $\rho$ .*

(ii) *Assume that  $\mathcal{M}$  is positive semi-definite on  $(\mathcal{B}_{sa}(\mathcal{H}), \langle \cdot, \cdot \rangle_\sigma)$  and that  $\mathcal{N}^*\hat{\mathcal{N}}^* = \hat{\mathcal{N}}^*\mathcal{N}^*$ . If  $\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}$ , then for all  $X \in \mathcal{B}_{sa}(\mathcal{H})$*

$$\mathcal{E}_{\mathcal{S}_{\mathcal{N}}}(X, X) \geq \mathcal{E}_{\mathcal{S}_{\mathcal{M}}}(X, X)$$

(iii) *Assume that  $\mathcal{N} = \mathcal{N}_{p,\sigma}$ , with  $\mathcal{N}_{p,\sigma}$  the generalized depolarizing channel with depolarizing probability  $p$  to a state  $\sigma > 0$ , and  $\mathcal{N}_p \succeq_{\text{l.n.}} \mathcal{M}$  for some channel  $\mathcal{M}$  with  $\sigma$  as its stationary state. Then:*

$$\forall X \in \mathcal{B}(\mathcal{H})_+, p\mathcal{E}_{\text{id}-\mathcal{N}_{1,\sigma}^*}(X, X) \leq \mathcal{E}_{\text{id}-\mathcal{M}^*}(X, X),$$

*Proof.* (i) We use the implication (i)  $\Rightarrow$  (ii) of Proposition 6.7 in order to get that  $\mathcal{M} \succeq_{\text{l.n.}} \mathcal{N}$  implies that for any two states  $\rho_1, \rho_2$

$$\chi^2(\mathcal{M}(\rho_1), \mathcal{M}(\rho_2)) \geq \chi^2(\mathcal{N}(\rho_1), \mathcal{N}(\rho_2)).$$

This implies, taking  $\rho_1 := \rho \equiv \sigma^{\frac{1}{2}} X \sigma^{\frac{1}{2}}$  and  $\rho_2 = \sigma$  that

$$\text{Tr}(\mathcal{M}(\rho) \sigma^{-\frac{1}{2}} \mathcal{M}(\rho) \sigma^{-\frac{1}{2}}) \geq \text{Tr}(\mathcal{N}(\rho) \sigma^{-\frac{1}{2}} \mathcal{N}(\rho) \sigma^{-\frac{1}{2}})$$



Note that the adjoint of  $\mathcal{M}^*$  with respect to the weighted scalar product is  $\hat{\mathcal{M}}^*$ . Thus, we can rewrite this last condition as follows:

$$\langle \mathcal{M}^*(X), \mathcal{M}^*(X) \rangle_\sigma \geq \langle \mathcal{N}^*(X), \mathcal{N}^*(X) \rangle_\sigma \Leftrightarrow \mathcal{E}_{\text{id}-\mathcal{M}^*\hat{\mathcal{M}}^*}(X, X) \leq \mathcal{E}_{\text{id}-\mathcal{N}^*\hat{\mathcal{N}}^*}(X, X)$$

(ii) The previously proved result implies in particular that the maps  $\mathcal{M}^*$  and  $\mathcal{N}^*$  seen as matrices acting on the real vector space of self-adjoint operators provided with the  $\sigma$ -KMS inner product  $\langle \cdot, \cdot \rangle_\sigma$ ,

$$\mathcal{N}^*\hat{\mathcal{N}}^* \leq_{\text{PSD}} \mathcal{M}^*\hat{\mathcal{M}}^*$$

From Lemma 7.1, we find that

$$\frac{\mathcal{M}^* + \hat{\mathcal{M}}^*}{2} \geq_{\text{PSD}} \frac{\mathcal{N}^* + \hat{\mathcal{N}}^*}{2} \Rightarrow \forall X \in \mathcal{B}_{sa}(\mathcal{H}), \mathcal{E}_{\mathcal{S}_{\mathcal{N}}}(X, X) \geq \mathcal{E}_{\mathcal{S}_{\mathcal{M}}}(X, X)$$

(iii) Note that  $\mathcal{N}_{p,\sigma}^*\hat{\mathcal{N}}_{p,\sigma}^* = \hat{\mathcal{N}}_{p,\sigma}^*\mathcal{N}_{p,\sigma}^*$  for the depolarizing channel. Thus, the assumptions of (ii) are fulfilled. The claim then follows by a simple manipulation.  $\square$

Note that  $p\mathcal{E}_{\text{id}-\mathcal{N}_{1,\sigma}^*}(X, X)$  is nothing but the variance of the observable  $X$  with respect to the state  $\sigma$ . Thus, we conclude that from a less-noisy comparison between channels with the same invariant state we immediately get a comparison of Dirichlet forms and a bound on logarithmic Sobolev constants if one of them is known. Unfortunately, such techniques do not yield sharp inequalities for the SDPI constant for channels. This is because it is known that for generalized depolarizing channels, we have that the optimal SDPI constant [58] is given by:

$$D(\mathcal{N}_{p,\sigma}(\rho)\|\sigma) \leq (1-p)^{1+\alpha(\sigma)} D(\rho\|\sigma)$$

where

$$\alpha(\sigma) = \min_{x \in [0,1]} (1 + q_{s_{\min}(\sigma)}(x))$$

with  $q_{s_{\min}(\sigma)}$  the smallest eigenvalue of  $\sigma$  and

$$q_y(x) = \begin{cases} \frac{D_2(y\|x)}{D_2(x\|y)}, & x \neq y \\ 1, & x = y \end{cases},$$

where  $D_2$  is the binary relative entropy. Moreover, for qubit depolarizing channels we have  $\alpha(I/2) = 1$  and we know that  $\mathcal{N}_{p,\sigma}^{\otimes n}$  satisfies a SDPI with constant  $(1-p)$  [4]. This is to be contrasted with the LS estimate, which decays with the local dimension.

## 8 Special classes of channels

### 8.1 Weyl-covariant channels

In this section, we provide new comparison bounds for Weyl-covariant channels [81, 23, 37, 38] (see also [80]). We consider the finite group  $G := \mathbb{Z}_n \times \mathbb{Z}_n$  with the following projective representation on  $\mathcal{H} := \mathbb{C}^n$ :

$$(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n \mapsto W_{a,b} := U^a V^b, \quad \text{where} \quad U := \sum_{c \in \mathbb{Z}_n} |c+1\rangle\langle c| \quad \text{and} \quad V := \sum_{c \in \mathbb{Z}_n} e^{\frac{2\pi i c}{n}} |c\rangle\langle c|.$$

Next, define the *discrete Fourier transform*  $F$  to be the following unitary matrix on  $\mathbb{C}^n$

$$F := \frac{1}{\sqrt{n}} \sum_{a,b \in \mathbb{Z}_n} e^{\frac{2\pi i a b}{n}} E_{a,b}.$$

A map  $\mathcal{M}$  on  $\mathcal{B}(\mathcal{H})$  is a *Weyl-covariant map* if

$$\mathcal{M}(W_{a,b} X W_{a,b}^*) = W_{a,b} \mathcal{M}(X) W_{a,b}^*$$

for every  $X \in \mathcal{B}(\mathcal{H})$  and  $(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n$ . In particular, any additive noise channel of the form

$$\mathcal{M}_f(\rho) := \sum_{(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n} f(a,b) W_{a,b} \rho W_{a,b}^\dagger, \quad (8.1)$$

where  $f : \mathbb{Z}_n \times \mathbb{Z}_n \rightarrow \mathbb{R}_+$  is a probability mass function, is Weyl-covariant, by the irreducibility of the representation  $W$ . Moreover, one can show that any Weyl-covariant channel has this form (see Corollary 4.15 in [80]). The next Lemma is an extension of Lemma 1 in [50]:

**Lemma 8.1.** *Any two Weyl-covariant noise channels  $\mathcal{M}_f$  and  $\mathcal{M}_h$  satisfy  $\mathcal{M}_h \succeq_{\text{deg}} \mathcal{M}_f$  if and only if there exists a probability distribution function  $k$  on  $\mathbb{Z}_n \times \mathbb{Z}_n$  such that  $\mathcal{M}_f = \mathcal{M}_h \circ \mathcal{M}_k = \mathcal{M}_k \circ \mathcal{M}_h$ .*

*Proof.* We simply need to prove the forward implication. By definition, there exists a channel  $\mathcal{N}$  such that  $\mathcal{M}_f = \mathcal{N} \circ \mathcal{M}_h$ . Now, for any  $(a, b) \in \mathbb{Z}_n \times \mathbb{Z}_n$ , since  $W_{a,b} W_{c,d} = \alpha W_{c,d} W_{a,b}$  where  $\alpha := e^{\frac{2\pi i(ad-bc)}{n}}$ ,

$$\mathcal{N} \circ \mathcal{M}_h(W_{a,b}) = \sum_{(a',b') \in \mathbb{Z}_n \times \mathbb{Z}_n} h(a',b') \mathcal{N}(W_{a',b'} W_{a,b} W_{a',b'}^\dagger) \quad (8.2)$$

$$= \sum_{(a',b') \in \mathbb{Z}_n \times \mathbb{Z}_n} h(a',b') e^{\frac{2\pi i(a'b-b'a)}{n}} \mathcal{N}(W_{a,b}) \quad (8.3)$$

$$\equiv A_h(a,b) \mathcal{N}(W_{a,b}). \quad (8.4)$$

On the other hand,  $\mathcal{N} \circ \mathcal{M}_h(W_{a,b}) = \mathcal{M}_f(W_{a,b}) = A_f(a,b) W_{a,b}$ . Assume first that  $A_h(a,b) = 0$  for some  $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$ . Then  $A_f(a,b) = 0$  from the degradability condition, and we can choose without loss of generality  $\mathcal{N}(W_{a,b}) = 0$ . Now, assume that  $A_h(a,b) \neq 0$ . Then we have  $\mathcal{N}(W_{a,b}) = A_f(a,b)/A_h(a,b) W_{a,b}$ . Therefore, we found that there exists a function  $C : \mathbb{Z}_n \times \mathbb{Z}_n \mapsto \mathbb{C}$  such that for all  $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$ ,  $\mathcal{N}(W_{a,b}) = C(a,b) W_{a,b}$ . We conclude by an appeal to Theorem 4.14 in [80].  $\square$

In Section 2.1, we introduced the notion of a less noisy domination region. Restricting ourselves to comparisons with additive noise channels, we can also define the so-called *additive less noisy domination region* of a channel  $\mathcal{M}$  acting on system  $\mathcal{H}$  as

$$\mathcal{L}_{\mathcal{M}}^{\text{add}} := \{\mathcal{M}_f : \mathcal{M} \succeq_{\text{l.n.}} \mathcal{M}_f\}$$

and its *additive degradation region*

$$\mathcal{D}_{\mathcal{M}}^{\text{add}} := \{\mathcal{M}_f : \mathcal{M} \succeq_{\text{deg}} \mathcal{M}_f\},$$

where both are subsets of the set of Weyl-covariant channels of the form of (8.1).

**Proposition 8.2.** *Let  $\mathcal{M}_k$  be a Weyl-covariant quantum channel corresponding to the probability distribution function  $k$ . Then,  $\mathcal{L}_{\mathcal{M}_f}^{\text{add}}$  and  $\mathcal{L}_{\mathcal{M}_f}^{\text{add}}$  are non-empty, closed, convex, and invariant under the conjugate action of  $\{W_{a,b}\}_{(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n}$  for all  $(a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n$ ,*

$$\mathcal{M}_h \in \mathcal{L}_{\mathcal{M}_f}^{\text{add}} \Rightarrow W_{a,b} \circ \mathcal{M}_h \in \mathcal{L}_{\mathcal{M}_f}^{\text{add}} \quad \text{and} \quad \mathcal{M}_h \in \mathcal{D}_{\mathcal{M}_f}^{\text{add}} \Rightarrow W_{a,b} \circ \mathcal{M}_h \in \mathcal{D}_{\mathcal{M}_f}^{\text{add}},$$

where  $W_{a,b}(X) := W_{a,b} X W_{a,b}^\dagger$ . Moreover,  $\mathcal{D}_{\mathcal{M}_f}^{\text{add}} = \text{conv}(\{W_{a,b} \circ \mathcal{M}_f : (a,b) \in \mathbb{Z}_n \times \mathbb{Z}_n\})$ .

*Proof.* Non-emptiness, closure and convexity follow in the same way as in Proposition 2.2 in addition with the fact that the set of additive noise channels is closed and convex. Invariance under the conjugate action of the Weyl operators is also apparent. The identity in (2) follows from Lemma 8.1.  $\square$

In the case when  $f$  is the distribution

$$\omega_\delta := \left(1 - \delta, \frac{\delta}{n^2 - 1}, \dots, \frac{\delta}{n^2 - 1}\right),$$

the map  $\mathcal{M}_\delta := \mathcal{M}_f$  reduces to

$$\begin{aligned} \mathcal{M}_\delta(\rho) &= (1 - \delta)\rho + \frac{\delta}{n^2 - 1} \sum_{(a,b) \neq (0,0)} W_{a,b} \rho W_{a,b}^\dagger = (1 - \delta)\rho + \frac{\delta}{n^2 - 1} (n \operatorname{Tr}(\rho) \mathbb{1} - \rho) \\ &= \rho p + (1 - p) \operatorname{Tr}(\rho) \frac{\mathbb{1}}{n}, \end{aligned}$$

with  $p = 1 - \frac{\delta n^2}{(n^2 - 1)}$ . This is the depolarizing channel  $\mathcal{N}_p^{\text{depol}}$  of noise parameter  $p$ . We recall that this map is completely positive if and only if  $-\frac{1}{n^2 - 1} \leq p \leq 1$ , i.e. if and only if  $0 \leq \delta \leq 1$  [40, 43]. Moreover, the family  $(\mathcal{N}_p^{\text{depol}})_{p \in \mathbb{R}}$  forms a semigroup:  $\mathcal{N}_{p_1 p_2}^{\text{depol}} = \mathcal{N}_{p_1}^{\text{depol}} \circ \mathcal{N}_{p_2}^{\text{depol}} = \mathcal{N}_{p_2}^{\text{depol}} \circ \mathcal{N}_{p_1}^{\text{depol}}$ . From this property, one can easily derive that for  $\delta \leq \frac{n^2 - 1}{n^2}$ ,

$$\delta \leq \gamma \leq 1 - \frac{\delta}{n^2 - 1} \quad \Leftrightarrow \quad \mathcal{M}_\delta \succeq_{\text{deg}} \mathcal{M}_\gamma. \quad (8.5)$$

In the next result, we prove that a depolarizing channel can never be less noisy than an erasure channel, hence extending Proposition 7 of [50]:

**Proposition 8.3.** *For  $n \geq 2$ , given an erasure channel  $\mathcal{E} : \mathcal{B}(\mathbb{C}^n) \rightarrow \mathcal{B}(\mathbb{C}^{n+1})$  with erasure probability  $\varepsilon \in (0, 1)$ , there does not exist  $\delta \in (0, 1)$  such that the corresponding depolarizing channel  $\mathcal{M}_\delta : \mathcal{B}(\mathbb{C}^n) \rightarrow \mathcal{B}(\mathbb{C}^n)$  satisfies  $\mathcal{M}_\delta \succeq_{\text{l.n.}} \mathcal{E}$ .*

*Proof.* For an erasure channel  $\mathcal{E}$  with  $\varepsilon \in (0, 1)$ , we always have  $D(\mathcal{E}(\mathbb{1}/n) \| \mathcal{E}(|1\rangle\langle 1|)) = +\infty$ . On the other hand, for any  $\delta \in (0, 1)$ , we have  $D(\mathcal{M}_\delta(\rho) \| \mathcal{M}_\delta(\sigma)) < +\infty$  for all  $\rho, \sigma \in \mathcal{D}(\mathbb{C}^n)$ . Hence  $\mathcal{M}_\delta \not\succeq_{\text{l.n.}} \mathcal{E}$  for any  $\delta \in (0, 1)$  by the characterization in Proposition 2.3.  $\square$

As we saw in Proposition 3.2, degradability is at least as strong as any notion of less noisy ordering. This raises the question of whether one can find pair of channels  $\mathcal{M}, \mathcal{N}$  such that  $\mathcal{M} \succeq_{\text{l.n.}}^c \mathcal{N}$  without  $\mathcal{M} \succeq_{\text{deg}} \mathcal{N}$ . In the next result, we make one step towards this direction by proving a separation between the (non-complete) less noisy ordering and degradability between two quantum channels. This result extends Proposition 11 of [50]:

**Proposition 8.4.** *Given a Weyl-covariant channel  $\mathcal{M}_f$  over  $\mathbb{Z}_n \times \mathbb{Z}_n$ ,  $n \geq 2$ ,  $\delta \in [0, \frac{n^2 - 1}{n^2}]$  and*

$$\gamma_0 := \frac{1 - \delta}{1 - \delta + \frac{\delta}{(n^2 - 1)^2}}.$$

*If  $\mathcal{M}_f$  is a convex mixture of the channels  $\mathcal{W}_{a,b} \circ \mathcal{M}_\delta$  and  $\mathcal{W}_{a',b'} \circ \mathcal{M}_{\gamma_0}$ ,  $(a, b), (a', b') \in \mathbb{Z}_n \times \mathbb{Z}_n$ , then*

$$\mathcal{M}_\delta \succeq_{\text{l.n.}} \mathcal{M}_f.$$

**Remark 8.5.** *Choosing  $\mathcal{M}_f$  to be the depolarizing channel  $\mathcal{M}_\delta$ , we already saw in (8.5) that  $\mathcal{M}_\delta \succeq_{\text{deg}} \mathcal{M}_\gamma$ , and hence  $\mathcal{M}_\delta \succeq_{\text{l.n.}} \mathcal{M}_\gamma$ , whenever  $\delta \leq \gamma \leq 1 - \frac{\delta}{n^2 - 1}$ . Therefore, since  $\gamma_0 \geq 1 - \frac{\delta}{n^2 - 1}$ , Proposition 8.4 provides us with a class of channels that belong to  $\mathcal{L}_{\mathcal{M}_\delta}^{\text{add}} \setminus \mathcal{D}_{\mathcal{M}_\delta}^{\text{add}}$ . However the problem of finding two channels  $\mathcal{M}, \mathcal{N}$  such that  $\mathcal{M} \succeq_{\text{l.n.}}^c \mathcal{N}$  but  $\mathcal{M} \not\succeq_{\text{deg}} \mathcal{N}$  still remains open, even when the channels  $\mathcal{M}$  and  $\mathcal{N}$  are taken to be classical.*

*Proof.* Thanks to Proposition 8.2, it is enough to prove that  $\mathcal{M}_\delta \succeq_{\text{l.n.}} \mathcal{M}_{\gamma_0}$ . Then, by the characterization in Proposition 2.3, we simply need to show that for any two states  $\rho, \sigma \in \mathcal{D}(\mathbb{C}^n)$ ,

$$D(\mathcal{M}_\delta(\rho) \| \mathcal{M}_\delta(\sigma)) \geq D(\mathcal{M}_{\gamma_0}(\rho) \| \mathcal{M}_{\gamma_0}(\sigma)).$$

Denoting by  $\rho := \sum_j \lambda_j |e_j\rangle\langle e_j|$  and  $\sigma := \sum_k \mu_k |f_k\rangle\langle f_k|$  the eigenvalue decompositions of the states, we can easily show that the above relative entropies are linear functions of the square overlaps between the eigenvectors of  $\rho$  and those of  $\sigma$ . For instance, denoting by  $p = 1 - \frac{\delta n^2}{n^2-1}$ , we have

$$\mathcal{M}_\delta(\rho) = \sum_j \lambda_j(p) |e_j\rangle\langle e_j|, \quad \mathcal{M}_\delta(\sigma) = \sum_k \mu_k(p) |f_k\rangle\langle f_k|, \quad (8.6)$$

where  $\lambda_j(p) := p\lambda_j + (1-p)\frac{1}{n}$  and  $\mu_k(p) := p\mu_k + (1-p)\frac{1}{n}$ . Hence,

$$D(\mathcal{M}_\delta(\rho) \parallel \mathcal{M}_\delta(\sigma)) = \sum_j \lambda_j(p) \ln \lambda_j(p) - \sum_{j,k} \lambda_j(p) \ln \mu_k(p) |\langle e_j | f_k \rangle|^2.$$

Similarly, we denote by  $q = 1 - \frac{\gamma_0 n^2}{n^2-1}$  and define  $\lambda_j(q)$  and  $\mu_j(q)$  as the corresponding convex mixtures of the eigenvalues of  $\rho$ , resp.  $\sigma$  and the ones of the completely mixed state. Denote by  $A$  the  $n \times n$  matrix whose entropies are given by  $a_{jk} := |\langle e_j | f_k \rangle|^2$ . Since both  $\{|e_j\rangle\}$  and  $\{|f_k\rangle\}$  form orthonormal bases,  $A$  is doubly stochastic. Then, by Birkhoff's theorem,  $A$  can be written as a convex combination of permutation matrices. Therefore, we can conclude that if an inequality of the form

$$\sum_j \lambda_j(p) \ln \lambda_j(p) - \sum_{j,k} \lambda_j(p) \ln \mu_k(p) a_{jk} \geq \sum_j \lambda_j(q) \ln \lambda_j(q) - \sum_{j,k} \lambda_j(q) \ln \mu_k(q) a_{jk}$$

holds for all permutation matrices  $A$ , then it holds for all doubly stochastic matrices  $A$ , and therefore (8.6) holds for all states  $\rho, \sigma$ . However, restricting to the case of permutation matrices means that we can, without loss of generality, assume that the states  $\rho$  and  $\sigma$  commute. In other words, we reduced the proof to that of classical depolarizing channels over classical probability mass functions, for which we can directly invoke Proposition 11 in [50].  $\square$

**Remark 8.6.** *The use of Birkhoff's theorem for reducing a proof to the commutative regime is not new. It was recently used in the context of functional inequalities in order to find the exact (modified) logarithmic Sobolev constant of the depolarizing semigroup [58, 4].*

## 8.2 Gaussian channels

In this section, we consider the setting of Gaussian quantum channels [39]: an  $n$ -mode quantum system is modeled by the Hilbert space  $L^2(\mathbb{R}^n) \simeq L^2(\mathbb{R})^{\otimes n}$  of complex-valued, square-integrable functions on  $\mathbb{R}^n$ . On this space, we define the so-called *creation*  $a_j^\dagger$  and *annihilation*  $a_j$  operators,  $j \in [n]$ , by means of their commutation relations

$$[a_i, a_j^\dagger] = \delta_{ij} \mathbb{1}, \quad [a_i, a_j] = [a_i^\dagger, a_j^\dagger] = 0, \quad i, j = 1, \dots, n,$$

where for each  $j \in [n]$ ,  $a_j$  and  $a_j^\dagger$  act nontrivially on the  $j$ -th copy of the system. The number operator is defined as

$$N^{(n)} := \sum_{j=1}^n a_j^\dagger a_j$$

and corresponds to the energy of the system. In the case  $n = 1$ , it is known to have the following spectral decomposition onto the Fock basis  $\{|k\rangle\}_{k \in \mathbb{N}}$ :

$$N^{(1)} = \sum_{k=0}^{\infty} k |k\rangle\langle k|.$$

Next, we denote by  $|z\rangle$  the coherent state of parameter  $z \in \mathbb{C}$ . These are the eigenvectors of the one-mode annihilation operator,  $a|z\rangle = z|z\rangle$ . In what follows, for a vector  $\mathbf{z} \in \mathbb{C}^n \equiv \mathbb{R}^{2n}$ , the tensor product of coherent states  $|z_j\rangle$  will be denoted by  $|\mathbf{z}\rangle$ . The Weyl displacement operators  $D(\mathbf{z})$ ,  $\mathbf{z} \in \mathbb{C}^n$ , are unitary operators

that rotate the  $n$ -mode vacuum state  $|\mathbf{0}\rangle$  to the coherent state  $|\mathbf{z}\rangle = D(\mathbf{z})|\mathbf{0}\rangle$ . In terms of the creation and annihilation operators, they are defined as

$$D(\mathbf{z}) = \exp \left[ \mathbf{a}^\dagger \cdot \mathbf{z} - \mathbf{z}^* \cdot \mathbf{a} \right]. \quad (8.7)$$

Next, an  $n$ -mode quantum Gaussian state is proportional to the exponential of a quadratic polynomial in the creation and annihilation operators. Among these states, thermal Gaussian states play an important role. They are defined by  $n$ -fold products of the following geometric probability distribution for the energy:

$$\sigma(E) := \frac{1}{E+1} \sum_{k=0}^{\infty} \left( \frac{E}{E+1} \right)^k |k\rangle\langle k|, \quad E > 0.$$

The average energy of  $\sigma(E)$  is equal to  $\text{Tr}(\sigma(E)N^{(1)}) = E$ , whereas its entropy is equal to

$$g(E) := H(\sigma(E)) = (E+1) \ln(E+1) - E \ln(E). \quad (8.8)$$

Quantum Gaussian channels are those quantum channels  $\mathcal{N} : \mathcal{D}(L^2(\mathbb{R}^n)) \rightarrow \mathcal{D}(L^2(\mathbb{R}^m))$  that preserve the set of quantum Gaussian states. The most important classes of such channels are the beam-splitter, the squeezing, quantum Gaussian attenuators and quantum Gaussian amplifiers. The first two are unitary quantum analogs of the operation of linearly mixing random variables: a *two-mode beam-splitter*  $U_\lambda^{\text{bs}} : \mathcal{D}(L^2(\mathbb{R}^2)) \rightarrow \mathcal{D}(L^2(\mathbb{R}^2))$  of transmissivity  $0 < \lambda < 1$  is defined by the following action on the ladder operators:

$$\begin{aligned} U_\lambda^{\text{bs}\dagger} a_1 U_\lambda^{\text{bs}} &= \sqrt{\lambda} a_1 + \sqrt{1-\lambda} a_2 \\ U_\lambda^{\text{bs}\dagger} a_2 U_\lambda^{\text{bs}} &= -\sqrt{1-\lambda} a_1 + \sqrt{\lambda} a_2. \end{aligned}$$

This is a simple example of a passive Gaussian channel in the sense that it preserve the total energy:  $[U_\lambda^{\text{bs}}, N^{(2)}] = 0$ . On the other hand, the *two-mode squeezing*  $U_\kappa^{\text{sq}}$  of parameter  $\kappa \geq 1$  increases the energy of the input. It is defined similarly to the beam-splitter as follows:

$$\begin{aligned} U_\kappa^{\text{sq}\dagger} a_1 U_\kappa^{\text{sq}} &= \sqrt{\kappa} a_1 + \sqrt{\kappa-1} a_2 \\ U_\kappa^{\text{sq}\dagger} a_2 U_\kappa^{\text{sq}} &= \sqrt{\kappa-1} a_1 + \sqrt{\kappa} a_2. \end{aligned}$$

In both cases, it is standard to interpret the first factor of the tensor product as the system and the second factor as the environment. Next, for  $\lambda \geq 0$ , we denote by  $\mathcal{B}_\lambda : \mathcal{D}(L^2(\mathbb{R}^2)) \rightarrow \mathcal{D}(L^2(\mathbb{R}))$  the reduced state on the system of either the beam-splitter ( $\lambda \leq 1$ ) or the squeezer ( $\lambda \geq 1$ ). The *quantum Gaussian attenuator*  $\mathcal{E}_{\lambda,E} : \mathcal{D}(L^2(\mathbb{R})) \rightarrow \mathcal{D}(L^2(\mathbb{R}))$ , of parameters  $0 \leq \lambda \leq 1$  and  $E \geq 0$ , corresponds to the channel  $\rho \mapsto \mathcal{B}_\lambda(\rho \otimes \sigma(E))$ . Similarly, the *quantum Gaussian amplifier*  $\mathcal{A}_{\kappa,E} : \mathcal{D}(L^2(\mathbb{R})) \rightarrow \mathcal{D}(L^2(\mathbb{R}))$  of parameters  $\kappa \geq 1$  and  $E \geq 0$ , corresponds to the channel  $\rho \mapsto \mathcal{B}_\kappa(\rho \otimes \sigma(E))$ .

Another important Gaussian channel considered in the literature is the *Gaussian additive noise channel*. It is generated by a convex combination of displacement operators with a Gaussian probability measure:

$$\mathcal{N}_E(\rho) := \int_{\mathbb{C}} D(\sqrt{E}\mathbf{z}) \rho D(\sqrt{E}\mathbf{z})^\dagger e^{-|\mathbf{z}|^2} \frac{dz}{\pi}. \quad (8.9)$$

In fact, the Gaussian additive noise channel can be seen as a large particle number limit of the attenuator channel (see [31]).

All the channels mentioned above are typical examples of a *gauge-covariant* Gaussian channel [39]: A channel  $\mathcal{M} : \mathcal{D}(L^2(\mathbb{R}^n)) \rightarrow \mathcal{D}(L^2(\mathbb{R}^m))$  is called gauge-covariant if for all input state  $\rho \in \mathcal{D}(L^2(\mathbb{R}^n))$ ,

$$\mathcal{M}(U_\phi^{(n)} \rho (U_\phi^{(n)})^\dagger) = U_\phi^{(m)} \mathcal{M}(\rho) (U_\phi^{(m)})^\dagger$$

where  $U_\phi^{(n)} := e^{-i\phi N^{(n)}}$  is the gauge transformation corresponding to the  $n$ -mode harmonic oscillator. In fact, every phase-covariant quantum Gaussian channel can be expressed as a quantum-limited amplifier composed with a quantum-limited attenuator.

The following conjecture, stated in [32], was recently proved in the restricted one-mode case [27, 26] as well as for the range of parameters for which the channels become entanglement-breaking [24]; that is

- (i) For any quantum Gaussian attenuator  $\mathcal{E}_{\lambda,E}$  with  $E \geq \frac{\lambda}{1-\lambda}$ ;
- (ii) For any quantum Gaussian amplifier  $\mathcal{A}_{\kappa,E}$  with  $E \geq \frac{1}{\kappa-1}$ ;
- (iii) For any quantum Gaussian additive noise channel  $\mathcal{N}_E$  with  $E \geq 1$ .

**Conjecture 8.7** (Constrained minimum output entropy conjecture for phase-covariant quantum channels). *For any  $n \in \mathbb{N}$  and any  $\rho \in \mathcal{D}(L^2(\mathbb{R}^n))$  with finite entropy, let  $\sigma$  be the one-mode thermal Gaussian state with entropy  $\frac{H(\rho)}{n}$ . Then, for any  $E \geq 0$ ,  $0 \leq \lambda \leq 1$ ,  $\kappa \geq 1$ :*

$$H(\mathcal{E}_{\lambda,E}^{\otimes n}(\rho)) \geq H(\mathcal{E}_{\lambda,E}^{\otimes n}(\sigma^{\otimes n})) = n g\left(\lambda g^{-1}\left(\frac{H(\rho)}{n}\right) + (1-\lambda)E\right), \quad (8.10)$$

$$H(\mathcal{A}_{\kappa,E}^{\otimes n}(\rho)) \geq H(\mathcal{A}_{\kappa,E}^{\otimes n}(\sigma^{\otimes n})) = n g\left(\kappa g^{-1}\left(\frac{H(\rho)}{n}\right) + (\kappa-1)(E+1)\right), \quad (8.11)$$

$$H(\mathcal{N}_E^{\otimes n}(\rho)) \geq H(\mathcal{N}_E^{\otimes n}(\sigma^{\otimes n})) = n g\left(g^{-1}\left(\frac{H(\rho)}{n}\right) + E\right). \quad (8.12)$$

In the next result, we provide an exact expression for contraction coefficients of any of these Gaussian channels and their tensor products under the condition that Conjecture 8.7 holds. More precisely, let  $\sigma_j \in \mathcal{D}(L^2(\mathbb{R}))$ ,  $j \in [n]$ , be one-mode Gaussian states,  $\mathcal{G}$  be a one-mode Gaussian quantum channel, and define the contraction coefficients for the relative entropy as

$$\eta_{\text{Re}}\left(\bigotimes_j \sigma_j, \mathcal{G}^{\otimes n}\right) := \sup_{\substack{\rho \in \mathcal{D}(L^2(\mathbb{R}^n)) \\ 0 < D(\rho \| \bigotimes_j \sigma_j) < \infty}} \frac{D(\mathcal{G}^{\otimes n}(\rho) \| \bigotimes_j \mathcal{G}(\sigma_j))}{D(\rho \| \bigotimes_j \sigma_j)}. \quad (8.13)$$

Here, we define the relative entropy as Lindblad in [49, 48]: given any two positive, trace-class operators  $A, B$ , if  $\{|a_i\rangle\}$ , resp.  $\{|b_j\rangle\}$ , is a complete orthonormal set of eigenvectors of  $A$  with corresponding eigenvalues  $\{a_i\}$ , resp.  $B$  with eigenvalues  $\{b_j\}$ ,

$$\begin{aligned} D(A \| B) &:= \sum_i \langle a_i | (A \ln A - A \ln B + B - A) | a_i \rangle \\ &= \sum_j \langle b_j | (A \ln A - A \ln B + B - A) | b_j \rangle \\ &= \sum_{i,j} |\langle a_i | b_j \rangle|^2 (a_i \ln a_i - a_i \ln b_j + b_j - a_i). \end{aligned}$$

where the sum is taken to be  $+\infty$  if the series diverges. Furthermore, we define the *energy-constrained* relative entropy contraction coefficient as follows: for  $p > 0$

$$\eta_{\text{Re}}^p(\mathcal{G}^{\otimes n}, \bigotimes_j \sigma_j) := \sup_{\substack{\rho \in \mathcal{D}(L^2(\mathbb{R}^n)), \text{Tr}(\rho N^{(n)}) \leq p \\ 0 < D(\rho \| \bigotimes_j \sigma_j) < \infty}} \frac{D(\mathcal{G}^{\otimes n}(\rho) \| \bigotimes_j \mathcal{G}(\sigma_j))}{D(\rho \| \bigotimes_j \sigma_j)}.$$

**Theorem 8.8.** *Assume that Conjecture 8.7 holds true. Then for any family  $\{E_j\}_{j \in [n]}$  of energies, any  $0 \leq \lambda \leq 1$ ,  $1 \geq \kappa$ ,  $E \geq 0$ , and  $p \geq \max_j E_j$ :*

$$\begin{aligned} \eta_{\text{Re}}(\mathcal{E}_{\lambda,E}^{\otimes n}, \bigotimes_j \sigma(E_j)) &= \eta_{\text{Re}}^p(\mathcal{E}_{\lambda,E}^{\otimes n}, \bigotimes_j \sigma(E_j)) = \lambda \max_{j \in [n]} \frac{\ln\left(\frac{\lambda E_j + (1-\lambda)E+1}{\lambda E_j + (1-\lambda)E}\right)}{\ln\left(\frac{E_j+1}{E_j}\right)}, \\ \eta_{\text{Re}}(\mathcal{A}_{\kappa,E}^{\otimes n}, \bigotimes_j \sigma(E_j)) &= \eta_{\text{Re}}^p(\mathcal{A}_{\kappa,E}^{\otimes n}, \bigotimes_j \sigma(E_j)) = \kappa \max_{j \in [n]} \frac{\ln\left(\frac{\kappa E_j + (\kappa-1)(E+1)+1}{\kappa E_j + (\kappa-1)(E+1)}\right)}{\ln\left(\frac{E_j+1}{E_j}\right)}, \\ \eta_{\text{Re}}(\mathcal{N}_E^{\otimes n}, \bigotimes_j \sigma(E_j)) &= \eta_{\text{Re}}^p(\mathcal{N}_E^{\otimes n}, \bigotimes_j \sigma(E_j)) = \max_{j \in [n]} \frac{\ln\left(\frac{E_j+E+1}{E_j+E}\right)}{\ln\left(\frac{E_j+1}{E_j}\right)}. \end{aligned} \quad (8.14)$$

Before proving Theorem 8.8, we state and prove a couple of technical lemmas. The first one allows us to restrict the optimization (8.13) to finite-rank input states (see Lemma 21 in [51] for a classical analogue):

**Lemma 8.9** (Finite-rank state characterization of  $\eta_{\text{Re}}$ ). *Let  $\mathcal{T}$  be the set of finite-rank  $n$ -mode quantum states  $\rho \in \mathcal{D}(L^2(\mathbb{R}^n))$  supported on  $\mathcal{H}_0 := \text{lin}\{|k_1\rangle \otimes \dots \otimes |k_n\rangle, k_j \in \mathbb{N}\}$ . Then for any 1-mode bosonic channel  $\mathcal{N}$ , and any thermal Gaussian states  $\sigma_1, \dots, \sigma_n$ :*

$$\eta_{\text{Re}}(\mathcal{N}^{\otimes n}, \bigotimes_j \sigma_j) = \sup_{\substack{\rho \in \mathcal{T} \\ D(\rho \| \bigotimes_j \sigma_j) < \infty}} \frac{D(\mathcal{N}^{\otimes n}(\rho) \| \bigotimes_j \mathcal{N}(\sigma_j))}{D(\rho \| \bigotimes_j \sigma_j)}. \quad (8.15)$$

**Remark 8.10.** *Remark that the optimization does not need to assume that  $D(\rho \| \bigotimes_j \sigma_j) > 0$ : this is automatically true since  $\bigotimes_j \sigma_j$  is faithful, and hence cannot be equal to any finite-rank state.*

*Proof.* Consider a quantum state  $\rho \in \mathcal{D}(L^2(\mathbb{R}^n))$  such that  $0 < D(\rho \| \bigotimes_j \sigma_j) < \infty$ , and define a sequence  $\Pi_k$  of finite-dimensional orthogonal projections on  $L^2(\mathbb{R}^n)$  which commute with  $\bigotimes_j \sigma_j$  and such that  $\Pi_k \rightarrow \mathbb{1}$  strongly as  $k \rightarrow \infty$  (i.e.  $\lim_{k \rightarrow \infty} \|(\Pi_k - \mathbb{1})\psi\| = 0$  for all  $\psi \in L^2(\mathbb{R}^n)$ ). Then, for  $\rho^{(k)} := \frac{\Pi_k \rho \Pi_k}{\text{Tr}(\Pi_k \rho)}$ ,

$$D(\rho^{(k)} \| \bigotimes_j \sigma_j) = \frac{1}{\text{Tr}(\Pi_k \rho)} D(\Pi_k \rho \Pi_k \| \Pi_k \bigotimes_j \sigma_j \Pi_k) + 1 - \frac{\text{Tr}(\Pi_k \sigma)}{\text{Tr}(\Pi_k \rho)} - \ln \text{Tr}(\Pi_k \rho),$$

where the identity follows from the fact that  $[\Pi_k, \bigotimes_j \sigma_j] = 0$  by assumption. Now by strong convergence of the sequence  $\{\Pi_k\}_{k \in \mathbb{N}}$ ,  $\text{Tr}(\Pi_k \sigma), \text{Tr}(\Pi_k \rho) \rightarrow 1$  as  $k \rightarrow \infty$ . Moreover, by Lemma 4 in [49], we have that  $D(\Pi_k \rho \Pi_k \| \Pi_k \bigotimes_j \sigma_j \Pi_k) \rightarrow D(\rho \| \sigma)$  as  $k \rightarrow \infty$ . Hence,

$$D(\rho^{(k)} \| \bigotimes_j \sigma_j) \xrightarrow{k \rightarrow \infty} D(\rho \| \bigotimes_j \sigma_j). \quad (8.16)$$

Moreover, by lower semi-continuity of the quantum relative entropy, we also have that

$$\liminf_{k \rightarrow \infty} D(\mathcal{N}^{\otimes n}(\rho^{(k)}) \| \bigotimes_j \mathcal{N}(\sigma_j)) \geq D(\mathcal{N}^{\otimes n}(\rho) \| \bigotimes_j \mathcal{N}(\sigma_j)). \quad (8.17)$$

Combining Equation (8.16) and Equation (8.17), we have that

$$\liminf_{k \rightarrow \infty} \frac{D(\mathcal{N}^{\otimes n}(\rho^{(k)}) \| \bigotimes_j \mathcal{N}(\sigma_j))}{D(\rho^{(k)} \| \bigotimes_j \sigma_j)} \geq \frac{D(\mathcal{N}^{\otimes n}(\rho) \| \bigotimes_j \mathcal{N}(\sigma_j))}{D(\rho \| \bigotimes_j \sigma_j)}. \quad (8.18)$$

The rest of the proof relies on a diagonalization argument that was already used in Lemma 21 in [51]: suppose that  $\{\rho_m\}_{m \in \mathbb{N}}$  is a sequence of quantum states that satisfies  $0 < D(\rho_m \| \bigotimes_j \sigma_j) < \infty$  for all  $m \in \mathbb{N}$  and achieves the supremum in (8.13):

$$\lim_{m \rightarrow \infty} \frac{D(\mathcal{N}^{\otimes n}(\rho_m) \| \bigotimes_j \mathcal{N}(\sigma_j))}{D(\rho_m \| \bigotimes_j \sigma_j)} = \eta_{\text{Re}}(\mathcal{N}^{\otimes n}, \bigotimes_j \sigma_j).$$

Then, by (8.18), we can construct a sequence  $\{\rho_m^{(k(m))}\} \in \mathcal{T}_{m \in \mathbb{N}}$  where each  $k(m)$  is chosen such that for every  $m \in \mathbb{N}$ :

$$\frac{D(\mathcal{N}^{\otimes n}(\rho_m^{(k(m))}) \| \bigotimes_j \mathcal{N}(\sigma_j))}{D(\rho_m^{(k(m))} \| \bigotimes_j \sigma_j)} \geq \frac{D(\mathcal{N}^{\otimes n}(\rho_m) \| \bigotimes_j \mathcal{N}(\sigma_j))}{D(\rho_m \| \bigotimes_j \sigma_j)} - \frac{1}{2^m}.$$

Letting  $m \rightarrow \infty$ , we end up with

$$\liminf_{m \rightarrow \infty} \frac{D(\mathcal{N}^{\otimes n}(\rho_m^{(k(m))}) \| \bigotimes_j \mathcal{N}(\sigma_j))}{D(\rho_m^{(k(m))} \| \bigotimes_j \sigma_j)} \geq \eta_{\text{Re}}(\mathcal{N}^{\otimes n}, \bigotimes_j \sigma_j).$$

Since the supremum in (8.13) is over a smaller set than in (8.15), the above inequality is actually an equality, and the result follows.  $\square$

The next technical lemma will serve as a replacement of Bernoulli's inequality in [51]:

**Lemma 8.11.** *For any three numbers  $\nu$ ,  $0 \leq \beta \leq \alpha$ ,*

$$g(\nu + \alpha) - g(\nu + \beta) \leq \frac{\ln\left(\frac{\nu + \alpha + 1}{\nu + \alpha}\right)}{\ln\left(\frac{\alpha + 1}{\alpha}\right)} (g(\alpha) - g(\beta)), \quad (8.19)$$

where  $g(x) := (x + 1) \ln(x + 1) - x \ln(x)$ . Similarly, for any  $\nu$ ,  $0 \leq \beta \leq \alpha$  and  $0 \leq \eta$ , we have

$$g(\eta\alpha + |1 - \eta|\nu) - g(\eta\beta + |1 - \eta|\nu) \leq \eta \frac{\ln\left(\frac{\eta\alpha + |1 - \eta|\nu + 1}{\eta\alpha + |1 - \eta|\nu}\right)}{\ln\left(\frac{\alpha + 1}{\alpha}\right)} (g(\alpha) - g(\beta)), \quad (8.20)$$

*Proof.* The proof of (8.19) amounts to proving that the function  $\nu \mapsto \frac{g(\nu + \alpha) - g(\nu + \beta)}{\ln(\nu + \alpha + 1) - \ln(\nu + \alpha)} \equiv \frac{g(\nu + \alpha) - g(\nu + \beta)}{g'(\nu + \alpha)}$  is monotone decreasing. Now, since  $g$  is differentiable, we have

$$g(\nu + \alpha) - g(\nu + \beta) = \int_{\nu + \beta}^{\nu + \alpha} g'(s) ds = \int_{\beta}^{\alpha} g'(s + \nu) ds = \int_{\beta}^{\alpha} \ln\left(\frac{s + \nu + 1}{s + \nu}\right) ds.$$

Therefore, after dividing by  $g'(\nu + \alpha)$ , we reduced the proof to that of showing that the function  $\nu \mapsto \frac{\ln(s + \nu + 1) - \ln(s + \nu)}{\ln(\nu + \alpha + 1) - \ln(\nu + \alpha)}$  is monotone decreasing for all  $\beta \leq s \leq \alpha$ . After differentiation, we further simplify the proof to that of showing that the function

$$x \mapsto -\frac{1}{(x + 1)x} \frac{1}{\ln(x + 1) - \ln(x)} \quad (8.21)$$

is increasing. This last assertion can be showed to hold after differentiating the function defined in (8.21). To obtain (8.20), we simply operate the following replacements in (8.19):  $\alpha \rightarrow \eta\alpha$ ,  $\nu \rightarrow |1 - \eta|\nu$  and  $\beta \rightarrow \eta\beta$ . Then we are reduced to having to prove that

$$\frac{g(\eta\alpha) - g(\eta\beta)}{\ln(\eta\alpha + 1) - \ln(\eta\beta)} \leq \eta \frac{g(\alpha) - g(\beta)}{\ln(\alpha + 1) - \ln(\alpha)}.$$

Using once again the fundamental theorem of calculus and an obvious rescaling, we reduce the problem to that of proving that for all  $u \leq \alpha$  the function

$$\eta \mapsto \frac{\ln\left(\frac{\eta u + 1}{\eta u}\right)}{\ln\left(\frac{\eta \alpha + 1}{\eta \alpha}\right)}$$

is increasing. By differentiating this function, it is enough to show that the function

$$x \mapsto (x + 1) \ln\left(\frac{x + 1}{x}\right)$$

on  $\mathbb{R}_+$  is decreasing. One last differentiation reduces the problem to the basic inequality  $\ln(1 + u) \leq u$  on  $\mathbb{R}_+$ .  $\square$

We are now ready to prove our main result.

*Proof of Theorem 8.8. Step 1: Upper bounding  $\eta_{\text{Re}}$ :* by Lemma 8.9, we can restrict the optimization to that over finite-rank input states  $\rho \in \mathcal{D}(L^2(\mathbb{R}^n))$  supported on  $\mathcal{H}_0 = \text{lin}\{|k_1\rangle \otimes \cdots \otimes |k_n\rangle, k_j \in \mathbb{N}\}$ , so that  $\text{Tr}(\rho N^{(n)}) < \infty$ . Such states also have finite von Neumann entropy. Moreover, the energy at the output of any



Gaussian quantum channel is also finite [71]. Then given any sequence  $\{\sigma(E_j)\}$  of thermal Gaussian states, and denoting by  $q_j := \text{Tr}((\rho_j - \sigma(E_j))N^{(1)})$  the difference in energy between  $\rho_j = \text{Tr}_{\{j\}^c}(\rho)$  and  $\sigma(E_j)$ ,

$$\begin{aligned} D(\rho \| \bigotimes_j \sigma(E_j)) &= -H(\rho) - \text{Tr}[\rho \ln(\bigotimes_j \sigma(E_j))] \\ &= -H(\rho) + \sum_j \left( S(\sigma(E_j)) + q_j \ln\left(\frac{E_j + 1}{E_j}\right) \right) \\ &= -H(\rho) + \sum_j \left( g(E_j) + q_j \ln\left(\frac{E_j + 1}{E_j}\right) \right). \end{aligned} \quad (8.22)$$

Let us now consider the attenuator channel  $\mathcal{E}_{\lambda,E}$ . Since  $\mathcal{E}_{\lambda,E}^\dagger(N^{(1)}) = \lambda N^{(1)} + (1-\lambda)E\mathbb{1}$ , we have

$$\begin{aligned} D(\mathcal{E}_{\lambda,E}^{\otimes n}(\rho) \| \bigotimes_j \mathcal{E}_{\lambda,E}(\sigma(E_j))) &= -H(\mathcal{E}_{\lambda,E}^{\otimes n}(\rho)) + \sum_j \left( H(\mathcal{E}_{\lambda,E}(\sigma(E_j))) + \lambda q_j \ln\left(\frac{\lambda E_j + (1-\lambda)E + 1}{\lambda E_j + (1-\lambda)E}\right) \right) \\ &= -H(\mathcal{E}_{\lambda,E}^{\otimes n}(\rho)) + \sum_j \left( g(\lambda E_j + (1-\lambda)E) + \lambda q_j \ln\left(\frac{\lambda E_j + (1-\lambda)E + 1}{\lambda E_j + (1-\lambda)E}\right) \right). \end{aligned} \quad (8.23)$$

Hence, for  $\eta_{\text{Re}}(\bigotimes_j \sigma(E_j), \mathcal{E}_{\lambda,E}^{\otimes n})$  to be upper bounded by the right-hand side of (8.14), it suffices to prove that for any  $j \in [n]$ :

$$g(\lambda E_j + (1-\lambda)E) - \frac{H(\mathcal{E}_{\lambda,E}^{\otimes n}(\rho))}{n} \leq \frac{\ln\left(\frac{\lambda E_j + (1-\lambda)E + 1}{\lambda E_j + (1-\lambda)E}\right)}{\ln\left(\frac{E_j + 1}{E_j}\right)} \lambda \left( g(E_j) - \frac{H(\rho)}{n} \right).$$

By the constrained minimum entropy conjecture (8.10), we can further simplify the problem to that of proving, for any  $\beta := g^{-1}\left(\frac{H(\rho)}{n}\right)$ ,

$$g(\lambda E_j + (1-\lambda)E) - g(\lambda \beta + (1-\lambda)E) \leq \lambda \frac{\ln\left(\frac{\lambda E_j + (1-\lambda)E + 1}{\lambda E_j + (1-\lambda)E}\right)}{\ln\left(\frac{E_j + 1}{E_j}\right)} (g(E_j) - g(\beta)). \quad (8.24)$$

The result directly follows from (8.20) for  $\alpha := E_j$  and  $\nu := E$ . The case of the amplifier  $\mathcal{A}_{\kappa,E}$  follow the exact same proof, since  $\mathcal{A}_{\kappa,E}^\dagger(N^{(1)}) = \kappa N^{(1)} + (\kappa - 1)(E + 1)\mathbb{1}$ .

Now, we turn our attention to the additive noise channel  $\mathcal{N}_E$ : here, since  $\mathcal{N}_E^\dagger(N^{(1)}) = N^{(1)} + E\mathbb{1}$ , (8.23) has to be replaced by

$$\begin{aligned} D(\mathcal{N}_E^{\otimes n}(\rho) \| \bigotimes_j \mathcal{N}_E(\sigma(E_j))) &= -H(\mathcal{N}_E^{\otimes n}(\rho)) + \sum_j \left( H(\mathcal{N}_E(\sigma(E_j))) + q_j \ln\left(\frac{E_j + E + 1}{E_j + E}\right) \right) \\ &= -H(\mathcal{N}_E^{\otimes n}(\rho)) + \sum_j \left( g(E_j + E) + q_j \ln\left(\frac{E_j + E + 1}{E_j + E}\right) \right). \end{aligned} \quad (8.25)$$

Therefore, as previously, by conjecture (8.12) it is enough to prove that for any  $j \in [n]$ , denoting by  $\beta := g^{-1}\left(\frac{H(\rho)}{n}\right)$ ,

$$g(E_j + E) - g(\beta + E) \leq \frac{\ln\left(\frac{E_j + E + 1}{E_j + E}\right)}{\ln\left(\frac{E_j + 1}{E_j}\right)} (g(E_j) - g(\beta))$$

Then, invoking (8.19) for  $\alpha := E_j$  and  $\nu := E$  allows us to conclude.

**Step 2:**  $\eta_{\text{Re}} \geq \eta_{\text{Re}}^{(p)}$ : this is obvious by definition of the contraction coefficients.

**Step 3: Lower bounding  $\eta_{\text{Re}}^{(p)}$  for all  $p \geq \max_j E_j$ :** let us start with the additive noise channel  $\mathcal{N}_E$ : choosing  $\rho$  to be a tensor product state, we can reduce to the case  $n = 1$  without loss of generality. Moreover, we choose  $\rho$  to be the thermal Gaussian state  $\sigma(E')$ . Then, we have for any displaced thermal state  $\sigma(E')_z := D(z)\sigma(E')D(z)^\dagger$ ,  $z \in \mathbb{C}$ ,

$$\frac{D(\mathcal{N}_E(\sigma(E')_z) \| \mathcal{N}_E(\sigma(E_1)))}{D(\sigma(E')_z \| \sigma(E_1))} = \frac{-g(E' + E) + (1 + E' + |z|^2 + E) \ln(E_1 + 1 + E) - (E' + E + |z|^2) \ln(E_1 + E)}{-g(E') + (1 + E' + |z|^2) \ln(E_1 + 1) - (E' + |z|^2) \ln(E_1)}.$$

The lower bound follows after taking  $E' = E_1 - \delta$ ,  $|z|^2 = \delta$  and taking the limit  $\delta \rightarrow 0$ . Since  $\text{Tr}(\sigma(E')_z N^{(1)}) = E_1$  for all  $\delta > 0$ , the result follows. The cases of the amplifier and of the attenuator follow the exact same computations. The other two cases follows the exact same argument.  $\square$

**Remark 8.12.** In the case of the Gaussian attenuator channel  $\mathcal{E}_{\lambda, E}$ , we find that the contraction coefficient  $\eta_{\text{Re}}(\sigma(E)^{\otimes n}, \mathcal{E}_{\lambda, E}^{\otimes n}) = \lambda$ . This is also a consequence of the conditional entropy power inequality, as previously observed in [25].

**Remark 8.13.** In the classical setting, it is well-known that the contraction coefficient  $\eta(\mathcal{G}) = 1$  for the classical additive white Gaussian noise channel, even when restricting to inputs of finite energy. In order to get nontrivial contraction, the following non-linear strong data processing inequality was proposed in [66, 21]: for the channel  $\mathcal{G} : X \rightarrow Y$

$$F(t, \mathcal{G}, E) := \sup_{\substack{U \rightarrow X \rightarrow Y \\ I(U; X) \leq t \\ \mathbb{E}[X^2] \leq E}} I(U; Y),$$

where the supremum is over all joint distributions  $P_{UX}$  with constrained input energy over system  $X$ . In [21], it was shown that such curves are always strictly less than  $t$  for all  $t > 0$ . A similar analysis in the quantum setting will be the subject of future work.

## 9 Conclusions

In this work, we discussed the relationship between the concept of contraction coefficients and the less noisy partial order. The equivalence between their respective relative entropy and mutual information formulations allowed us a new point of view on their properties and bringing our knowledge in the quantum setting closer to what is currently available in the classical case.

Nevertheless, many open questions remain. Most crucially, it is not yet clear whether the relations between the different partial orders are actually strict. For example, we know that

$$\mathcal{N} \succeq_{\text{deg}} \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^c \mathcal{M} \Rightarrow \mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{M}$$

but it is not known whether the opposite directions might also hold, namely

$$\mathcal{N} \succeq_{\text{l.n.}}^{\text{reg}} \mathcal{M} \stackrel{?}{\Rightarrow} \mathcal{N} \succeq_{\text{l.n.}}^c \mathcal{M} \stackrel{?}{\Rightarrow} \mathcal{N} \succeq_{\text{deg}} \mathcal{M}$$

For SDPI constants and contraction coefficients, natural open problems include a better understanding of their tensorization properties and the possible equivalence of constants based on different divergences.

## Acknowledgments

CH and DSF acknowledge financial support from the VILLUM FONDEN via the QMATH Centre of Excellence (Grant no. 10059) and the QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union's Horizon 2020 Programme (QuantAlgo project) via the Innovation Fund Denmark. CR is partially supported by a Junior Researcher START Fellowship from the MCQST. CR acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germanys Excellence Strategy EXC-2111 390814868.

## References

- [1] R. Ahlswede and P. Gács. Spreading of sets in product spaces and hypercontraction of the markov operator. *The annals of probability*, pages 925–939, 1976.
- [2] K. M. Audenaert. A sharp continuity estimate for the von neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127, 2007.
- [3] S. Beigi. Sandwiched Rényi divergence satisfies data processing inequality. *Journal of Mathematical Physics*, 54(12):122202, jun 2013.
- [4] S. Beigi, N. Datta, and C. Rouzé. Quantum reverse hypercontractivity: Its tensorization and application to strong converses. *Communications in Mathematical Physics*, 376(2):753–794, 2020.
- [5] S. Beigi and A. Gohari. On dimension bounds for auxiliary quantum systems. *IEEE transactions on information theory*, 60(1):368–387, 2013.
- [6] P. Bergmans. Random coding theorem for broadcast channels with degraded components. *IEEE Transactions on Information Theory*, 19(2):197–207, 1973.
- [7] M. Berta, D. Sutter, and M. Walter. Quantum Brascamp-Lieb Dualities, 2019. arXiv:1909.02383v2.
- [8] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.
- [9] F. Buscemi. Degradable channels, less noisy channels, and quantum statistical morphisms: An equivalence relation. *Problems of Information Transmission*, 52(3):201–213, 2016.
- [10] F. Buscemi. Comparison of noisy channels and reverse data-processing theorems. In *2017 IEEE Information Theory Workshop (ITW)*, pages 489–493. IEEE, 2017.
- [11] L. Campbel. An extended cencov characterization of the information metric. In *Proc. AMS*, volume 98, pages 135–141, 1996.
- [12] Y. Cao and J. Lu. Tensorization of the strong data processing inequality for quantum chi-square divergences. *Quantum*, 3:199, 2019.
- [13] E. A. Carlen and A. Vershynina. Recovery map stability for the Data Processing Inequality. *arXiv preprint 1710.02409*, 2017.
- [14] N. N. Cencov. *Statistical decision rules and optimal inference*. Number 53. American Mathematical Soc., 2000.
- [15] M.-D. Choi, M. B. Ruskai, and E. Seneta. Equivalence of certain entropy contraction coefficients. *Linear algebra and its applications*, 208:29–36, 1994.
- [16] M. Christandl, C. Hirche, and A. Winter. to be published. 2020.
- [17] J. Cohen, J. H. Kempermann, and G. Zbaganu. *Comparisons of stochastic matrices with applications in information theory, statistics, economics and population*. Springer Science & Business Media, 1998.
- [18] J. E. Cohen, Y. Iwasa, G. Rautu, M. B. Ruskai, E. Seneta, and G. Zbaganu. Relative entropy under mappings by stochastic matrices. *Linear algebra and its applications*, 179:211–235, 1993.
- [19] T. Cover. Broadcast channels. *IEEE Transactions on Information Theory*, 18(1):2–14, 1972.
- [20] I. Csiszár and J. Korner. Broadcast channels with confidential messages. *IEEE transactions on information theory*, 24(3):339–348, 1978.
- [21] F. d. P. Calmon, Y. Polyanskiy, and Y. Wu. Strong data processing inequalities for input constrained additive noise channels. *IEEE Transactions on Information Theory*, 64(3):1879–1892, 2018.

- [22] 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)*, pages 1157–1161. IEEE, 2019.
- [23] E. B. Davies. On the repeated measurement of continuous observables in quantum mechanics. *Journal of Functional Analysis*, 6(2):318–346, 1970.
- [24] G. De Palma. New lower bounds to the output entropy of multi-mode quantum gaussian channels. *IEEE Transactions on Information Theory*, 65(9):5959–5968, 2019.
- [25] G. De Palma and S. Huber. The conditional entropy power inequality for quantum additive noise channels. *Journal of Mathematical Physics*, 59(12):122201, 2018.
- [26] G. De Palma, D. Trevisan, and V. Giovannetti. Gaussian states minimize the output entropy of the one-mode quantum attenuator. *IEEE Transactions on Information Theory*, 63(1):728–737, 2016.
- [27] G. De Palma, D. Trevisan, and V. Giovannetti. Gaussian states minimize the output entropy of one-mode quantum gaussian channels. *Physical review letters*, 118(16):160503, 2017.
- [28] P. Del Moral, M. Ledoux, and L. Miclo. On contraction properties of markov kernels. *Probability theory and related fields*, 126(3):395–420, 2003.
- [29] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 08 1996.
- [30] M. Fanizza, F. Kianvash, and V. Giovannetti. Quantum flags, and new bounds on the quantum capacity of the depolarizing channel. *arXiv preprint arXiv:1911.01977*, 2019.
- [31] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro. Minimum output entropy of bosonic channels: a conjecture. *Physical Review A*, 70(3):032315, 2004.
- [32] S. Guha, J. H. Shapiro, and B. I. Erkmen. Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture. *Physical Review A*, 76(3):032303, 2007.
- [33] F. Hiai and D. Petz. The proper formula for relative entropy and its asymptotics in quantum probability. *Comm. Math. Phys.*, 143(1):99–114, 1991.
- [34] F. Hiai and M. B. Ruskai. Contraction coefficients for noisy quantum channels. *Journal of Mathematical Physics*, 57(1):015211, 2016.
- [35] C. Hirche and D. Reeb. Bounds on information combining with quantum side information. *IEEE Transactions on Information Theory*, 64(7):4739–4757, 2018.
- [36] C. Hirche and A. Winter. An alphabet-size bound for the information bottleneck function. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2383–2388. IEEE, 2020.
- [37] A. S. Holevo. A note on covariant dynamical semigroups. *Reports on mathematical physics*, 32(2):211–216, 1993.
- [38] A. S. Holevo. Covariant quantum markovian evolutions. *Journal of Mathematical Physics*, 37(4):1812–1832, 1996.
- [39] A. S. Holevo. *Quantum systems, channels, information: a mathematical introduction*, volume 16. Walter de Gruyter, 2012.
- [40] M. Horodecki and P. Horodecki. Reduction criterion of separability and limits for a class of distillation protocols. *Physical Review A*, 59(6):4206, 1999.
- [41] M. J. Kastoryano and K. Temme. Quantum logarithmic Sobolev inequalities and rapid mixing. *Journal of Mathematical Physics*, 54(5), 2013.

- [42] J. Korner and K. Marton. Comparison of two noisy channels. *Topics in information theory*, pages 411–423, 1977.
- [43] L. Lami and M. Huber. Bipartite depolarizing maps. *Journal of Mathematical Physics*, 57(9):092201, 2016.
- [44] F. Leditzky, E. Kaur, N. Datta, and M. M. Wilde. Approaches for approximate additivity of the holevo information of quantum channels. *Physical Review A*, 97(1):012332, 2018.
- [45] F. Leditzky, D. Leung, and G. Smith. Dephasure channel and superadditivity of coherent information. *Physical review letters*, 121(16):160501, 2018.
- [46] A. Lesniewski and M. B. Ruskai. Monotone Riemannian Metrics and Relative Entropy on Non-Commutative Probability Spaces. page 33, 1998.
- [47] A. Lesniewski and M. B. Ruskai. Monotone Riemannian metrics and relative entropy on noncommutative probability spaces. *Journal of Mathematical Physics*, 40(11):5702–5724, 1999.
- [48] G. Lindblad. Entropy, information and quantum measurements. *Communications in Mathematical Physics*, 33(4):305–322, 1973.
- [49] G. Lindblad. Expectations and entropy inequalities for finite quantum systems. *Communications in Mathematical Physics*, 39(2):111–119, 1974.
- [50] A. Makur and Y. Polyanskiy. Comparison of channels: criteria for domination by a symmetric channel. *IEEE Transactions on Information Theory*, 64(8):5704–5725, 2018.
- [51] A. Makur and L. Zheng. Linear bounds between contraction coefficients for  $f$ -divergences. *arXiv preprint arXiv:1510.01844*, 2015.
- [52] A. Makur and L. Zheng. Comparison of contraction coefficients for  $f$ -divergences. *Problems of Information Transmission*, 56(2):103–156, 2020.
- [53] K. Marton. A coding theorem for the discrete memoryless broadcast channel. *IEEE Transactions on Information Theory*, 25(3):306–311, 1979.
- [54] J. Melbourne, M. Madiman, and M. V. Salapaka. Relationships between certain  $f$ -divergences. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1068–1073. IEEE, 2019.
- [55] L. Miclo. Remarques sur l’hypercontractivité et l’évolution de l’entropie pour des chaînes de markov finies. In *Séminaire de Probabilités XXXI*, pages 136–167. Springer, 1997.
- [56] A. Müller-Hermes and D. S. Franca. Sandwiched Rényi Convergence for Quantum Evolutions. *arXiv:1607.00041*, 2016.
- [57] A. Müller-Hermes and D. Reeb. Monotonicity of the quantum relative entropy under positive maps. In *Annales Henri Poincaré*, volume 18, pages 1777–1788. Springer, 2017.
- [58] A. Müller-Hermes, D. Stilck França, and M. M. Wolf. Relative entropy convergence for depolarizing channels. *Journal of Mathematical Physics*, 57(2):022202, 2016.
- [59] M. Muller-Lennert, F. Dupuis, O. Szehr, S. Fehr, and M. Tomamichel. On quantum Rényi entropies: A new generalization and some properties. *Journal of Mathematical Physics*, 54(12):122203, jun 2013.
- [60] T. Nishiyama. A new lower bound for kullback-leibler divergence based on hammersley-chapman-robbins bound. *arXiv preprint arXiv:1907.00288*, 2019.
- [61] T. Nishiyama and I. Sason. On relations between the relative entropy and  $\chi^2$ -divergence, generalizations and applications, 2020.

- [62] T. Ogawa and H. Nagaoka. Strong converse and Stein’s lemma in quantum hypothesis testing. *IEEE Trans. Inform. Theory*, 46(7):2428–2433, 2000.
- [63] R. Olkiewicz and B. Zegarlinski. Hypercontractivity in Noncommutative  $L_p$  Spaces. *Journal of Functional Analysis*, 161(1):246–285, 1999.
- [64] R. Olkiewicz and B. Zegarlinski. Hypercontractivity in noncommutative  $L_p$  Spaces. *Journal of Functional Analysis*, 161(1):246–285, Jan. 1999.
- [65] D. Petz. Sufficient subalgebras and the relative entropy of states of a von neumann algebra. *Communications in Mathematical Physics*, 105(1):123–131, Mar. 1986.
- [66] Y. Polyanskiy and Y. Wu. Dissipation of information in channels with input constraints. *IEEE Transactions on Information Theory*, 62(1):35–55, 2016.
- [67] Y. Polyanskiy and Y. Wu. Strong data-processing inequalities for channels and bayesian networks. In *Convexity and Concentration*, pages 211–249. Springer, 2017.
- [68] M. Raginsky. Logarithmic sobolev inequalities and strong data processing theorems for discrete channels. In *2013 IEEE International Symposium on Information Theory*. IEEE, jul 2013.
- [69] M. Raginsky. Strong data processing inequalities and  $\phi$ -sobolev inequalities for discrete channels. *IEEE Transactions on Information Theory*, 62(6):3355–3389, 2016.
- [70] N. Ramakrishnan, R. Iten, V. Scholz, and M. Berta. Quantum blahut-arimoto algorithms. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1909–1914. IEEE, 2020.
- [71] M. Shirokov. On extension of quantum channels and operations to the space of relatively bounded operators. *arXiv preprint arXiv:1903.06086*, 2019.
- [72] D. Sutter and J. M. Renes. Universal polar codes for more capable and less noisy channels and sources. In *2014 IEEE International Symposium on Information Theory*, pages 1461–1465. IEEE, 2014.
- [73] D. Sutter, V. B. Scholz, A. Winter, and R. Renner. Approximate degradable quantum channels. *IEEE Transactions on Information Theory*, 63(12):7832–7844, 2017.
- [74] K. Temme, M. J. Kastoryano, M. B. Ruskai, M. M. Wolf, and F. Verstraete. The  $\chi^2$ -divergence and mixing times of quantum Markov processes. *Journal of Mathematical Physics*, 51(12), 2010.
- [75] A. Tikku, M. Berta, and J. M. Renes. Non-additivity in classical-quantum wiretap channels. *arXiv preprint arXiv:2002.06580*, 2020.
- [76] M. Tomamichel. *Quantum Information Processing with Finite Resources: Mathematical Foundations*, volume 5. Springer, 2015.
- [77] X. Wang. Optimizing the fundamental limits for quantum and private communication. *arXiv preprint arXiv:1912.00931*, 2019.
- [78] S. Watanabe. Private and quantum capacities of more capable and less noisy quantum channels. *Physical Review A*, 85(1):012326, 2012.
- [79] J. Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5(11):217–238, 2009.
- [80] J. Watrous. *The theory of quantum information*. Cambridge University Press, 2018.
- [81] H. Weyl. *The theory of groups and quantum mechanics*. Courier Corporation, 1950.
- [82] M. M. Wilde. *Quantum information theory*. Cambridge University Press, 2013.
- [83] M. M. Wilde. Optimized quantum f-divergences and data processing. *Journal of Physics A: Mathematical and Theoretical*, 51(37):374002, 2018.

- [84] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur. Amortized channel divergence for asymptotic quantum channel discrimination. *Letters in Mathematical Physics*, 110(8):2277–2336, 2020.
- [85] M. M. Wilde, A. Winter, and D. Yang. Strong converse for the classical capacity of entanglement-breaking and hadamard channels via a sandwiched rényi relative entropy. *Communications in Mathematical Physics*, 331(2):593–622, July 2014.
- [86] A. Winter. Tight uniform continuity bounds for quantum entropies: conditional entropy, relative entropy distance and energy constraints. *Communications in Mathematical Physics*, 347(1):291–313, 2016.