Technische Universität München
Lehrstuhl für Mathematische Physik

Entropic inequalities for bosonic systems

Stefan Huber

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:
Prof. Dr. Daniel Matthes

Prüfende der Dissertation:

1. Prof. Dr. Robert König
2. Prof. Dr. Eric Carlen,
   Rutgers, The State University of New Jersey (nur schriftliche Beurteilung)
3. Prof. Dr. Andreas Winter,
   Universitat Autònoma de Barcelona

Die Dissertation wurde am 30.01.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 09.05.2019 angenommen.

Department of Mathematics
Technical Univeristy of Munich

Technical University of Munich
Chair of Mathematical Physics

Entropic inequalities for bosonic systems

Stefan Huber

Full imprint of the dissertation approved by the Department of Mathematics of the Technical University of Munich to obtain the academic degree of

Doctor of Natural Sciences (Dr. rer. nat.)

Chairman:
Prof. Dr. Daniel Matthes

Examiners of the dissertation:

1. Prof. Dr. Robert König
2. Prof. Dr. Eric Carlen,
   Rutgers, The State University of New Jersey
3. Prof. Dr. Andreas Winter,
   Universitat Autònoma de Barcelona

The dissertation was submitted to the Technical University of Munich on 30.01.2019 and was accepted by the Department of Mathematics on 09.05.2019.

Zusammenfassung

Diese Dissertation befasst sich mit entropischen Ungleichungen für bosonische Kanäle in der Quanteninformationstheorie. Derartige Ungleichungen finden Anwendung in grundlegenden Fragen der Quantenkommunikation und der Konvergenz von quantendynamischen Semigruppen. Wir beweisen unter anderem eine Ungleichung für die Ausgangsentropie eines Quantenkanals, der klassisches Rauschen modelliert, und verwenden diese Ungleichung, um Schranken an die klassische Kapazität einer allgemeinen Familie von nichtgaußschen Quantenkanälen zu finden.

Abstract

This dissertation deals with entropic inequalities for bosonic channels in quantum information theory. Such inequalities have applications in foundational questions of quantum communication and can be used to make statements about the convergence of quantum dynamical semigroups. We prove, among other results, an inequality for the output entropy of a quantum channel which models additive classical noise and apply this inequality to derive bounds on the classical capacity of a general family of non-Gaussian quantum channels.

Mathematics is an art of *human* understanding.

– *William Thurston*

Acknowledgements

First and foremost, I want to thank my advisor, Prof. Robert König, for the encouraging support throughout these years while advising this thesis.

Next, I want to thank M5, and especially Prof. Michael Wolf and Prof. Robert König for creating such a great work environment; Wilma and Silvia for being the best possible help navigating through all sorts of issues; the people with whom I have shared an office during this time: first Milán, and then Martina, simply for being there; all the PhD students that came and went, for countless discussions during lunch breaks and other things: Andreas, Anna, Axy, Daniel, Javier, Kech, Luca, Margret, Martin, Martina, Matthias, Max, Max, Yimin; the postdocs Anna, Daniela, Ion, Milán, Moto, Sabina; all the students and visitors who passed by, and many more.

In the past years, I have been honored to do research with a number of bright people, during which I have learned invaluably much. My gratitude in this context especially goes to my coauthors Giacomo De Palma, Robert König, Joe Renes, Volkher Scholz, Marco Tomamichel, and Anna Vershynina. Extra thanks go to Raúl García-Patrón for hosting me for a summer visit at ULB and teaching me about boson sampling.

Going back further to the past, I am obliged to Renato Renner for explaining information-theoretic aspects of the Maxwell Demon during his *Theorie der Wärme* lecture, which was the original spark of my interest in information theory. Moreover, I have to thank Volkher Scholz a second time for sparking my interest in the mathematical aspects of quantum information theory.

In addition, I want to thank all the new friends from the quantum information community I have met during the past years, for all the time we spent together. Extra thanks go to Norbert for burning the midnight oil at a number of conferences with us, and to my name-cousins Stefan and Marcus, who showed me that there is a place for someone like me in science.

Furthermore, I want to thank

> all the friends I met during my studies at ETH, for all the collaborative effort spent trying to understand mathematics and physics, and
>
> my dearly beloved friends from outside science. I cannot possibly describe how valuable you are to me. You know who you are.

Finally, I thank my family, especially my parents Barbara and Norbert for their patience and sympathy, and my cousin Emanuel, who spent countless hours explaining basic mathematics to me while I was a first-year student.

List of contributed articles

This thesis is based on the following articles:

Core articles as principal author

I) Stefan Huber, Robert König, and Anna Vershynina
Geometric inequalities from phase space translations.
J. Math. Phys. 58, 012206 (2017).
(cf. Article [1] in the bibliography)

II) Stefan Huber and Robert König
Coherent state coding approaches the capacity of non-Gaussian bosonic channels.
J. Phys. A: Math. Theor. 51, 184001 (2018).
(cf. Article [2] in the bibliography)

III) Giacomo De Palma and Stefan Huber
The conditional entropy power inequality for quantum additive noise channels.
J. Math. Phys. 59, 122201 (2018).
(cf. Article [3] in the bibliography)

Further articles

IV) Joseph M. Renes, Volkher B. Scholz, and Stefan Huber
Uncertainty relations: An operational approach to the error-disturbance tradeoff.
Quantum 1, 20 (2017).
(cf. Article [4] in the bibliography)

V) Stefan Huber, Robert König, and Marco Tomamichel
Jointly constrained semidefinite bilinear programming with an application to Dobrushin curves.
arXiv:1808.03182 [quant-ph] (2018).
Submitted to *IEEE Trans. Inf. Theory.*
(cf. Article [5] in the bibliography)

I, Stefan Huber, am the principal author of articles I, II and III.

Contents

1 Introduction

Quantum information theory has been a very active field of research in the past decades. At its core lies the question how information theory (i.e., the study of information-processing tasks and their limitations) changes when we assume that quantum mechanics, as opposed to classical mechanics, governs the information carriers. There are multiple approaches as to which information carriers shall be used for quantum computing. One popular approach is the use of qubits, i.e., two-level systems. Another one, and the main focus of this thesis, is the deployment of continuous-variable carriers, which live in an infinite-dimensional Hilbert space, such as the Hilbert space for a fixed number of quantum harmonic oscillators. The primary concern of this thesis is bosonic quantum systems and noise acting on them: We are concerned with output entropies of noisy quantum channels and the information-carrying capacities of these channels. An understanding of noise and how to deal with it is essential on the way to a large-scale quantum computer, and this work develops tools in this direction.

This thesis deals with multiple approaches to continuous-variable information theory: One major part of the work presented here is an effort to investigate well-established information-theoretic inequalities from classical information theory and translate them to the quantum setting. The main mathematical tools we use for this are from functional analysis. This provides tools for a variety of tasks, such as bounding the classical capacity of quantum channels or bounding the convergence rate of certain semigroups. In this way, this thesis is concerned with both the development of new tools for quantum information theory and the application of these tools.

We start with a short presentation of the contributed articles and their scope. This is followed by an introduction of some basic notions which are ubiquitous in quantum mechanics and quantum information in Chapter 2. Chapter 3 then introduces the task of communicating classical information via quantum channels. We derive the classical capacity of quantum channels, one of the key quantities in quantum information theory. An introduction to the main concepts of continuous-variable quantum information, which is our main concern, is then given in Chapter 4. After this, we give a more detailed account of the state of the art of one specific topic in continuous-variable quantum information: entropic inequalities for bosonic channels, presented in Chapter 5. This topic is of central importance to the work presented in the contributed articles. We then change the topic to some applications of the presented functional inequalities. In this last part of our review of the current state of the art, we discuss the application of entropic inequalities to capacities of bosonic channels.

After this overview, we briefly present the contributed articles. Every embedded article in the Appendix is preceded by a more detailed and more technical summary of the main results and a description of the individual contribution of the author of this thesis. In cases where the article has already been published elsewhere, we include the permission to use it in this thesis.

1.1 Summary and Discussion of Results

The contributed articles take different approaches to the field of continuous-variable quantum information: First, articles I and III develop functional inequalities for bosonic channels which generalize a variety of established results from classical information theory to the quantum case. These inequalities are interesting from a purely mathematical point of view, and are not necessarily motivated by a concrete physical problem yet. Article II then applies these entropic inequalities to obtain bounds on the information-carrying capacity of a range of noisy bosonic channels. Lastly, articles IV and V do not focus directly on bosonic channels, but nonetheless cover continuous variables (Article IV) and concepts from information theory which we generalize to the quantum case, and which are intimately connected to the general information-theoretic questions we are interested in in this thesis (Article V). These two articles are not the primary contribution to this thesis, but are included for completeness. We note that the author of this thesis does not claim to be the principal author of these two articles.

Core articles as principal author

- *Article I [1]: Geometric inequalities from phase space translations*
  In classical information theory, there are important inequalities which can be viewed as analogs of inequalities from geometric analysis. One example of this connection is that the entropy power inequality is formally equivalent to the Brunn-Minkowski inequality when the entropy power plays the role of volume and the sum of random variables, which is defined in terms of the convolution of their respective probability densities, plays the role of the Minkowski sum. There is a wide range of information-theoretic inequalities which make statements about entropic quantities involving sums of random variables. In addition to the entropy power inequality, notable examples are the Fisher information inequality, the isoperimetric inequality for entropies, the Fisher information isoperimetric inequality, and the concavity of the entropy power under action of the heat diffusion semigroup.

  In Article I we study a convolution operation between a probability density function on phase space and a quantum state which was originally introduced by Werner [6], and connect it to a quantum diffusion semigroup which plays a role which is analogous to the heat semigroup. We prove a number of new inequalities involving quantum entropy and quantum Fisher information. These are quantum analogs of the information-theoretic inequalities mentioned above. As a main result, we prove a new entropy power inequality for classical noise channels. As an application, we derive a Log-Sobolev inequality for the quantum Ornstein-Uhlenbeck semigroup and apply it to obtain bounds on the entropy production rate of this semigroup. As an interesting side result, we show that Gaussian thermal states minimize the entropy production rate for the one-mode attenuator semigroup among all states with bounded mean photon number. The mathematical tools used in proving our results include the establishment and application of a data processing inequality for the convolution between a probability density function and a quantum state, as well as bounds on the entropy production rate of semigroups. For the latter, we employ recent majorization-type results for bosonic quantum channels. The connection between the quantum diffusion semigroup and the geometric inequalities relies on the fact that the Fisher information is equal to the entropy production rate under the diffusion, a fact referred to as the de Bruijn identity. In both the quantum and classical settings, the de Bruijn identity plays an important role in the proof of information-theoretic inequalities.

  In this article, we also conjectured that the quantum Ornstein-Uhlenbeck semigroup converges in relative entropy to its fixed point at a certain exponential rate. This conjecture has subsequently been proven by Carlen and Maas [7] using methods of gradient flow.

More recently we have provided a proof in Article III using the entropy power inequality directly.

This work inspired follow-up work in two different directions: Application to capacities (Article II) and generalizations of these inequalities to a setting with side information (Article III).

I was significantly involved in finding the ideas and carrying out the work of all parts of this article, and I was in charge of writing the article, with the exception of Section V A and Lemma 8.

- *Article II [2]: Coherent state coding approaches the capacity of non-Gaussian bosonic channels*
  A central question in quantum communication is whether entangled quantum states can be used to provide an advantage over classically correlated quantum states for communication of classical information over a given quantum channel. The maximal achievable communication rate using unentangled or entangled states is called the one-shot classical capacity and the full classical capacity, respectively. The question whether these two capacities are equal is commonly referred to as the addivity problem – if the answer to the above question is no, then the full classical capacity is said to be additive. In the setting of bosonic channels, it has been shown that entanglement does not provide an advantage for communication over a certain class of Gaussian channels [8]. Despite this landmark achievement, only little is known about non-Gaussian channels. The general additivity question for bosonic channels remains open.

  Article II investigates the consequences of recently proven entropy power inequalities and conjectured Entropy Photon-Number Inequalities on the classical capacity of a general class of bosonic channels, which includes non-Gaussian channels. These channels are beamsplitters with a generic, potentially non-Gaussian environment state, and classical noise channels with probabilistic noise that need not be Gaussian. We prove upper and lower bounds on the classical capacity of these channels. These are the first available bounds on the classical capacity of non-Gaussian bosonic channels. We show that for these channels, additivity violations for the classical capacity, if at all existent, are rather minor. In fact, we upper bound the maximal additivity violation by a constant independent of the input energy. This requires giving upper bounds on the full capacity and lower bounds on the one-shot capacity. The lower bounds are achievable by using classical modulation of coherent states for the encoding. Furthermore, we show similar results assuming the validity of the conjectured Entropy Photon-Number Inequality. In the case of classical noise channels, we conjecture a new Entropy Photon-Number-type Inequality for this purpose. Our results show that the Entropy Photon-Number Inequality only provides a small improvement on the upper bound on the full classical capacity for these channels. In addition to various forms of entropy power inequalities / Entropy Photon-Number Inequalities, the main tool used in the proofs is the fact that Gaussian states maximize the quantum entropy for a given energy. Furthermore, we can make use of recent results on the output entropy of one-mode phase-covariant Gaussian channels. For some particular cases, these are slightly better than those derived from the entropy power inequality.

  In spirit, this work translates results on classical additive noise channels which were originally obtained by Shannon [9, 10] to the setting of non-Gaussian bosonic quantum channels. It is inspired by earlier work by König and Smith [11] which proved upper bounds on the classical capacity of thermal noise channels, but additionally gives lower bounds.

This work was motivated by discussions with Robert König on possible applications of our previously published article [1]. I proved all the results of the paper, and I wrote all sections with the exception of the Introduction and the first half of Section 2.

- *Article III [3]: The conditional entropy power inequality for quantum additive noise channels*
  In classical information theory, many applications of the entropy power inequality use a formulation of the inequality which makes a statement about *conditional* entropies. This conditional entropy power inequality is a simple corollary to the entropy power inequality. This is due to the fact that the conditional entropy is simply an expectation value of entropy of conditional distributions, where the expectation is taken over the random variable we condition on. For quantum entropy, this is no longer the case if the system on which we condition is not classical. Therefore, a conditional entropy power inequality does not follow immediately from the entropy power inequality.

  For the beamsplitter, a conditional entropy power inequality was first formulated and proven for Gaussian states in [12]. A full proof for general states was given in [13]. Therefore it is natural to ask whether a conditional version of the entropy power inequality for classical noise channels from Article I holds. In Article III we generalize the quantum entropy power inequality for classical noise channels (one of the central results of Article I) to the setting with side information: We consider a bipartite quantum system one part of which is affected by noise. The proof of this inequality makes use of an integral form of the Fisher information. As a consequence, it does not exhibit certain regularity issues present in previous proofs of the quantum entropy power inequality without side information. As such, it can be seen as a generalization of Article I to the conditional setting, which also implies the main results of Article I without regularity issues. We show the remarkable fact that the conditional version of the quantum entropy power inequality is optimal in the following sense: For every fixed pair of values of the conditional entropies at the input, there exists a sequence of Gaussian input states such that the conditional entropy power inequality is saturated in the limit. In contrast to this, the version of the entropy power inequality for classical noise channels without side information is not tight. Furthermore, we prove a variety of information-theoretic inequalities, the classical analogs of which were established a long time ago. This includes the conditional Stam inequality, the conditional Fisher information inequality, and the isoperimetric inequality for conditional quantum entropies. As an application, we prove an upper bound on the entanglement-assisted classical capacity of a non-Gaussian bosonic channel, namely a classical noise channel where the probability density function of the noise is not Gaussian. We also show how the quantum entropy power inequality implies fast convergence of the quantum Ornstein-Uhlenbeck semigroup in relative entropy, a conjecture first stated in Article I and proven by different methods in [7]. In fact, we prove a more general statement regarding the convergence in relative entropy of a bipartite system one part of which undergoes a quantum Ornstein-Uhlenbeck evolution.

  This work was motivated by discussions with Giacomo De Palma during a visit he made to Munich. A sketch of the proof of the main result was worked out during discussions, after which I completed all the proofs and wrote the article, with the exception of Lemma 3 and Theorem 9, the proofs of which came from Giacomo De Palma. I was significantly involved in the scientific work of all parts of the article, with the aforementioned exceptions.

- *Article IV [4]: Uncertainty relations: An operational approach to the error-disturbance tradeoff*
  The Heisenberg uncertainty relation is perhaps one of the most famous aspects of quantum mechanics. The original formulation by Heisenberg in 1927 was somewhat vague, and only much later there have been proofs of formal statements. In recent years, an active field of research has opened up discussing uncertainty relations in different settings.

  In Article IV we focus on two aspects of uncertainty. In spirit, these already appear in Heisenberg's article: Joint measurability and the error-disturbance tradeoff. The former deals with the question to which precision two observables can be simultaneously measured, and the latter states that the more precise a measurement of one observable is, the larger is the disturbance to another non-commuting observable. In this context, it is not clear how the notions of "error" and "disturbance" are to be defined. There are a number of different approaches.

  In this article we take an operational approach: we seek uncertainty relations which make statements about measurement devices, and not about the physical quantities themselves: We define error and disturbance in terms of the *distinguishing probability*, i.e., the probability that the actual behavior of a measurement apparatus can be distinguished from the ideal behavior in any single experiment. This approach has the benefit that the notion of distinguishability does not depend on concepts of quantum mechanics. It therefore avoids some conceptual difficulties. Our notions of error and disturbance are related to the completely bounded norm, which is a well-known norm in operator theory. We use this approach to derive Heisenberg-type uncertainty relations for both joint measurability and error-disturbance tradeoff for arbitrary finite-dimensional observables, as well as for position and momentum. A key tool in our proofs is the continuity of the Stinespring dilation, a remarkable mathematical result by Kretschmann, Schlingemann, and Werner [14]. The latter relates the distance between quantum channels with respect to the completely bounded norm to the distance of their respective Stinespring dilations with respect to the operator norm. We apply our error-disturbance relation to an information processing setting: We prove that quantum channels which can faithfully transmit information regarding one observable do not leak any information about conjugate observables to the environment. Moreover, we discuss a connection to wave-particle duality relations. These quantify a tradeoff between the observation of interference patterns and the gain of information about the path of the particle in a Mach-Zehnder interferometer.

  This project started while I was working on my Master's thesis project at ETH Zurich together with Joseph Renes and Volkher Scholz. This was a project about uncertainty relations in the same setting as the one discussed in this article. After my graduation, we continued to work on this topic, proving stronger statements about the position-momentum uncertainty relations, extending the results, and significantly changing the proof method employed. These extended results were then combined with earlier results found by Joseph Renes and Volkher Scholz [15] in the finite-dimensional case and published together in this article.

- *Article V [5]: Jointly constrained semidefinite bilinear programming with an application to Dobrushin curves*
  We consider a problem we call *jointly constrained semidefinite bilinear programming.* This

asks to minimize a bilinear function over a set of self-adjoint operators specified by joint semidefinite programming (SDP) constraints. It is given by

$$\min_{(X,Y)\in\mathcal{S}} \operatorname{tr}((X\otimes Y)Q) + \operatorname{tr}(AX) + \operatorname{tr}(BY)\ , \tag{1.1}$$

where $\mathcal{S}$ is a set of pairs of self-adjoint operators (X, Y) defined by a family of SDP constraints and Q, A, B are given self-adjoint operators. Such programs appear in a number of contexts in quantum information theory: As an example, the entanglement fidelity can be cast as such a program. The latter plays an important role in quantum communication and in entanglement distribution. The jointly constrained semidefinite bilinear program also appears in the context of quantum games and Bell inequalities. In addition, we show that the computation of Dobrushin curves, which give bounds on classical coding with energy constraints, can also be cast as a program of this form. In the quantum information theory literature, the so-called seesaw algorithm has been applied in various contexts. It tackles the jointly constrained semidefinite bilinear program by alternately fixing a value of X and Y and solving the resulting affine-linear problem for the other variable, has been applied in various contexts. The downside of the seesaw algorithm is that it is heuristic – in general, it will not produce an optimum of the problem.

The goal of our work is to give a new algorithm for jointly constrained semidefinite bilinear programming from the quantum information point of view. In this article, we give a branch-and-bound algorithm for the jointly constrained semidefinite bilinear program, which produces a sequence of feasible points which converge to the global optimum. The algorithm is a generalization of the branch-and-bound algorithm given by Al-Khayyal and Falk [16] for a jointly constrained bilinear program. Moreover, the algorithm gives upper and lower bounds on the value of the program at each step as well as values of X and Y at which the upper bound on the value (1.1) is attained. As an application, we use our algorithm to numerically compute Dobrushin curves for quantum channels. As mentioned, these give upper bounds on optimal codes for classical information in a scenario where the noise acts repeatedly.

It should be noted that this project works in an exclusively finite-dimensional setting. However, the concept of Dobrushin curves is intimately related to the type of quantum communication questions we are interested in in this thesis. The concept of Dobrushin curves has not been studied for bosonic channels yet. The main idea and a sketch of the algorithm was worked out by Robert König and Marco Tomamichel. I was responsible for the applications and for writing the article and the documentation of the code, with the exception of the Introduction and Section 4.4.2. The code itself was written by Marco Tomamichel and Robert König.

2 Basic structure of Quantum Mechanics

We give a short introduction to the basic mathematical concepts and the formalism underlying quantum mechanics and quantum information theory. All the results presented in this chapter are covered in several fairly standard textbooks about quantum mechanics and quantum information theory, such as [17–20]. The approach presented here generally follows the excellent books by Holevo [18,19], though the presentation of quantum states and the way we map them to density operators is inspired by [21] and [17].

Let us first fix some notation. In the following, $\mathcal{H}$ will always denote a separable Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ which is antilinear in the first argument and linear in the second argument, and norm $\| \cdot \|$ induced by this scalar product. We are going to make extensive use of the bra-ket notation, which denotes vectors $\phi \in \mathcal{H}$ via a ket[1] $|\phi\rangle$, and their corresponding dual vectors via the bra $\langle\phi| \in \mathcal{H}^*$. The latter stands for the continuous linear form $\langle\phi| : \mathcal{H} \to \mathbb{C}$, $\psi \mapsto \langle\phi|\psi\rangle := \langle\phi, \psi\rangle$. For $\psi, \phi \in \mathcal{H}$ we use the notation $|\psi\rangle\langle\phi|$ for the operator $\mathcal{H} \to \mathcal{H}$ which maps $\chi \mapsto |\psi\rangle \langle\phi|\chi\rangle = \langle\phi, \chi\rangle\psi$.

We will denote by $\mathcal{B}(\mathcal{H})$ the set of bounded linear operators $\mathcal{H} \to \mathcal{H}$, and by $\mathcal{B}_1(\mathcal{H})$ the set of trace-class operators

$$\mathcal{B}_1(\mathcal{H}) := \{A \in \mathcal{B}(\mathcal{H}) \bigm| \|A\|_1 := \operatorname{tr}|A| := \sum_{k=1}^{\infty} \langle \sqrt{A^\dagger A} e_k, e_k \rangle < \infty \} \ ,$$

where $\{e_k\}_{k \in \mathbb{N}}$ is any countable orthonormal basis of $\mathcal{H}$ and $A^\dagger \in \mathcal{B}(\mathcal{H})$ is the *adjoint* of A, defined by

$$\langle \phi, A\psi \rangle = \langle A^\dagger \phi, \psi \rangle \qquad \text{for all } \phi, \psi \in \mathcal{H} \ .$$

The set of *self-adjoint* bounded operators is denoted by $\mathcal{B}_{\mathrm{sa}}(\mathcal{H}) := \{A \in \mathcal{B}(\mathcal{H}) \bigm| A^\dagger = A\}$. For an operator $A_{12} \in \mathcal{B}_1(\mathcal{H}_1 \otimes \mathcal{H}_2)$, we denote its partial trace over the first system as $\operatorname{tr}_1(A_{12}) \in \mathcal{B}_1(\mathcal{H}_2)$, which is the unique operator $B \in \mathcal{B}_1(\mathcal{H}_2)$ such that

$$\operatorname{tr}\left(A_{12}(\mathbb{1} \otimes Y)\right) = \operatorname{tr}(BY) \qquad \text{for all } Y \in \mathcal{B}(\mathcal{H}_2) \ .$$

Given a linear map $\mathcal{T} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ and $n \in \mathbb{N}$, we define the map $\mathcal{T}^{\otimes n} : \mathcal{B}(\mathcal{H}^{\otimes n}) \to \mathcal{B}(\mathcal{H}^{\otimes n})$ via

$$\mathcal{T}^{\otimes n}(A_1 \otimes \cdots \otimes A_n) = \mathcal{T}(A_1) \otimes \cdots \otimes \mathcal{T}(A_n) \qquad \text{for } A_k \in \mathcal{B}(\mathcal{H}),\ 1 \le k \le n \ ,$$

and linearly extended.

2.1 Quantum states and measurements

Quantum mechanics, like any physical theory, aims to predict the outcomes of statistical experiments. A statistical experiment is divided into two parts, *preparation* and *measurement*. On the one hand, if we specify the preparation of a quantum system (which we will later call

[1] It can be useful to view a ket $|\phi\rangle$ as a linear map $\mathbb{C} \to \mathcal{H}$, $\alpha \mapsto \alpha\phi$. We will not distinguish between the two notions.

its *state*), we fix the outcome probabilities of all possible measurements. On the other hand, specifying a measurement fixes the outcome distribution of the statistical experiment for all states. Our basic assumptions on the structure of a statistical theory are the following [19]:

(i) Let there be given a set $\mathcal{S}$, whose elements are called *states*, and a set $\mathcal{M}$, whose elements are called *observables*. For arbitrary $S \in \mathcal{S}$ and $X \in \mathcal{M}$ there is a probability distribution μ_S^X on the σ-algebra $B(O)$ of Borel subsets of a set of outcomes O, called the probability distribution of the observable X in the state S.

(ii) For arbitrary $S_1, S_2 \in \mathcal{S}$, and an arbitrary number p with $0 < p < 1$, there exists $S \in \mathcal{S}$ such that $\mu_S^X = p\mu_{S_1}^X + (1-p)\mu_{S_2}^X$ for all $X \in \mathcal{M}$. The state S is said to be a *mixture of the states* S_1 *and* S_2 in the proportion $p : (1-p)$.

(iii) For arbitrary $X_1 \in \mathcal{M}$ and an arbitrary Borel function $f: O \to O$ there exists $X_2 \in \mathcal{M}$ such that $X_2 = f \circ X_1$, i.e., $\mu_S^{X_2}(B) = \mu_S^{X_1}(f^{-1}(B))$ for all Borel sets $B \in B(O)$. We say that the observable X_2 is *functionally subordinate* to the observable X_1.

A pair of non-empty sets $\{\mathcal{S}, \mathcal{M}\}$ which satisfies assumptions (i)-(iii) is called a *statistical model*. If, in addition, we have that

$$\mu_{S_1}^M = \mu_{S_2}^M \qquad \text{for all } M \in \mathcal{M} \ ,$$

implies that $S_1 = S_2$, and

$$\mu_S^{M_1} = \mu_S^{M_2} \qquad \text{for all } S \in \mathcal{S} \ ,$$

implies that $M_1 = M_2$, we call the statistical model *separable*. Separable models have the property that the mixture of states from assumption (ii) and the functional subordination of assumption (iii) are uniquely defined. This means that for separable models, the state set $\mathcal{S}$ has a convex structure, and the set of observables $\mathcal{M}$ has a partial order.

The function μ_S^M predicts the measurement statistics of a statistical experiment, i.e., $\mu_S^M(B)$ gives the probability that the measurement outcome of an experiment which measures the observable M in the state S lies in B. We note that the set of outcomes might be finite, countable, or even uncountably infinite. We will always assume $(O, B(O))$ to be a *standard Borel space*, i.e., O is a complete separable metric space and $B(O)$ is its Borel σ-algebra. Since standard Borel spaces of the same cardinality are isomorphic, $B(O)$ will always be equivalent to either a finite set, $\mathbb{N}$, or the Borel subsets of the real line, $B(\mathbb{R})$.

We now specify some more assumptions for the framework in which we want to formulate quantum theory [21].

(i) Observables are elements of a C^*-algebra.

(ii) The potential measurement outcomes lie in the spectrum of the elements of the C^*-algebra.

(iii) States are positive linear functionals which are normalized to 1.

These assumptions are sufficient to give a fairly concrete visualization of the structure. The Gelfand-Naimark theorem [21, 22] guarantees that we can always work with bounded linear operators on a Hilbert space.

Theorem 2.1.1 (Gelfand-Naimark). *For every C^*-algebra $\mathcal{A}$ there exists a Hilbert space $\mathcal{H}$ and an isometric $*-$homomorphism $\Xi : \mathcal{A} \to \mathcal{B}(\mathcal{H})$. If $\mathcal{A}$ is separable, then so is $\mathcal{H}$.*

The Gelfand-Naimark theorem allows us to consider a suitable Hilbert space $\mathcal{H}$ instead of an abstract C^*-algebra to realize the algebra of observables. Then the C^*-algebra of observables is identified with $\mathcal{B}(\mathcal{H})$, and observables are Hermitian elements in $\mathcal{B}(\mathcal{H})$.

Before we give the definition of a quantum state, let us define some notation. Let $\{A_k\}_{k\in\mathbb{N}}$ be a family of norm-bounded increasing operators with smallest upper bound $A \in \mathcal{B}(\mathcal{H})$ in the sense that $A \geq A_k$ for all k and if $B \geq A_k$ for all k then $B \geq A$. The relation $A \geq B$ for self-adjoint operators here means that $A - B$ is a positive operator. Then we write $A_k \uparrow A$. This notation comes from the fact that if $A_k \uparrow A$, then A_k converges to A weakly, ultraweakly, and strongly [17, Chapter 1.6].

Definition 2.1.2 (Quantum state)**.** *A quantum state is a linear functional* $\omega : \mathcal{B}(\mathcal{H}) \to \mathbb{C}$ *which is positive, normalized, and normal:*

(i) (positivity) $\omega(A) \geq 0$ *if* A *is a positive operator.*

(ii) (normalization) $\omega(\mathbb{1}_{\mathcal{H}}) = 1$.

(iii) (normality) If $A_k \uparrow A$, *then* $\lim_{k\to\infty} \omega(A_k) = \omega(A)$.

Here $\mathbb{1}_{\mathcal{H}}$ *is the identity map on* $\mathcal{H}$.

The normality assumption is not always used in literature. However, it leads to a useful description of quantum states in terms of so-called *density operators*:

Lemma 2.1.3 (States and density operators [17, Lemma 6.1])**.** *A positive linear functional* $\omega : \mathcal{B}_{\mathrm{sa}}(\mathcal{H}) \to \mathbb{R}$ *is normal if and only if there exists a positive* $\rho \in \mathcal{B}_1(\mathcal{H}) \cap \mathcal{B}_{\mathrm{sa}}(\mathcal{H})$ *such that*

$$\omega(A) = \mathrm{tr}(A\rho) \qquad \textit{for all } A \in \mathcal{B}_{\mathrm{sa}}(\mathcal{H}) \ .$$

If ω *is normalized, then* $\mathrm{tr}(\rho) = 1$.

Positive operators of trace one are called *density operators*. Applying Lemma 2.1.3 to the restriction $\omega|_{\mathcal{B}_{\mathrm{sa}}(\mathcal{H})}$ of a quantum state ω to the set of self-adjoint operators gives us a density operator ρ which describes the quantum state ω. The functional $\mathrm{tr}(\cdot\rho)$ can easily be extended back to $\mathcal{B}(\mathcal{H})$ by linearity. The duality between states and density operators given by Lemma 2.1.3 is central to quantum mechanics. The definition of states given in Definition 2.1.2 describes states in the so-called *Heisenberg picture*. The description in terms of density operators is referred to as the *Schrödinger picture*. If we had dropped the assumption of normality, there would exist states in the Heisenberg picture which do not correspond to density operators in the case of an infinite-dimensional Hilbert space $\mathcal{H}$ [17, Lemma 6.1]. In the following, we will use the terms quantum state and density operator interchangeably, and we will denote the set of density operators on $\mathcal{H}$ by $\mathcal{S}(\mathcal{H})$.

A special set of states are the so-called *pure states*. These correspond to vectors in $\mathcal{H}$ up to a phase. For any $\psi \in \mathcal{H}$ with $\|\psi\| = 1$, the projection $|\psi\rangle\langle\psi|$ onto $\mathbb{C}\psi$ is a density operator describing a quantum state ω_ψ via

$$\omega_\psi(A) = \mathrm{tr}(|\psi\rangle\langle\psi|\, A) = \langle\psi, A\psi\rangle \ . \tag{2.1}$$

States which can be written in the form (2.1) for $\psi \in \mathcal{H}$ are called *pure*. Note that all elements of the so-called *unit ray*

$$[\psi] := \{e^{i\alpha}\psi \;\big|\; \alpha \in [0, 2\pi]\}$$

define the same quantum state.

Every quantum state can be written as a convex combination of pure states. This can easily be seen by considering the spectral decomposition of a general density operator:

Theorem 2.1.4 (Convex combination of projections [21])**.** *Every state ω can be written as a convex linear combination of pure states, i.e., there exists a complete orthonormal system $\{\psi_j\}_{j=1}^{\infty} \subset \mathcal{H}$ and nonnegative numbers $p_1 \geq p_2 \geq \ldots$ such that $\sum_{j=1}^{\infty} p_j = 1$ and such that*

$$\omega = \sum_{j=1}^{\infty} p_j \; \omega_{[\psi_j]} \ .$$

Conversely, every convex linear combination of pure states $\omega_{[\phi_j]}, \{\phi_j\}_{j\in\mathbb{N}} \subset \mathcal{H}$, defines a quantum state.

Let us next define the mathematical objects which we will refer to as quantum-mechanical *measurements*.

Definition 2.1.5 (Measurement)**.** *A measurement is a positive operator-valued measure (POVM) $M : B(O) \to \mathcal{B}(\mathcal{H})$, i.e., a function $B(O) \to \mathcal{B}(\mathcal{H})$ with the following properties:*

1. *$M(B)$ is a positive operator in $\mathcal{H}$ for any $B \in B(O)$.*
2. *If $\{B_j\}_j$ is a finite or countable partition of O into pairwise disjoint measurable sets, then*

$$\sum_j M(B_j) = \mathbb{1}_{\mathcal{H}} \ ,$$

where the series converges strongly.

If $M(B)^2 = M(B)$ for all $B \in B(O)$, then M is a projective measurement*, also called a* projection-valued measure (PVM)*.*

In the case of finitely many outcomes $O = \{1, \ldots, n\}$, POVMs have a simpler description: They are simply collections of positive operators $\{M_j\}_{j=1}^{n} \subset \mathcal{B}(\mathcal{H})$ such that

$$\sum_{j=1}^{n} M_j = \mathbb{1}.$$

Furthermore, it is easy to see that every self-adjoint observable induces a projective measurement via the spectral theorem. The latter assigns to every self-adjoint operator X its spectral measure $E_X : B(\mathbb{R}) \to \mathcal{B}(\mathcal{H})$ such that

$$X = \int_{\mathbb{R}} x E_X(\mathrm{d}x) \ .$$

The spectral measure E_X is then a PVM. The probability distribution $\mu_\rho^X : B(\mathbb{R}) \to \mathbb{R}$ associated with the outcome statistics of an observable X in a state ρ is then given by

$$\mu_\rho^X(B) = \mathrm{tr}(\rho E_X(B)) \qquad \text{for all } B \in B(\mathbb{R}) \ .$$

Not every general POVM corresponds to a self-adjoint element of $\mathcal{B}(\mathcal{H})$. However, the two concepts are very closely related, because any POVM can be seen as a projective measurement on a possibly larger Hilbert space. Physically, this makes sense because the laboratory might only have access to a subsystem of a larger quantum system. This is, in essence, the content of Naimark's dilation theorem.

Theorem 2.1.6 (Naimark dilation [19]). *Every POVM $M : B(O) \to \mathcal{B}(\mathcal{H})$ can be extended to a projection-valued measure, i.e., there exists a Hilbert space $\mathcal{H}'$ containing $\mathcal{H}$ and a projection-valued measure E on $\mathcal{H}'$, $E : \mathcal{B}(O) \to \mathcal{B}(\mathcal{H}')$ such that*

$$M(B) = P_{\mathcal{H}} E(B)\big|_{\mathcal{H}} \qquad \textit{for all } B \in B(O) \ ,$$

where $P_{\mathcal{H}}$ is the projection from $\mathcal{H}'$ onto $\mathcal{H}$.

From Naimark's theorem, it follows that for an arbitrary POVM M in $\mathcal{H}$, there exists a Hilbert space $\mathcal{H}_0$, a density operator ρ_0 in $\mathcal{S}(\mathcal{H}_0)$ and a projector-valued measure E in $\mathcal{H} \otimes \mathcal{H}_0$ such that

$$\mu_\rho^M(B) = \operatorname{tr}\left((\rho \otimes \rho_0)E(B)\right) \ , \qquad \text{for all } B \in B(O), \rho \in \mathcal{S}(\mathcal{H}) \ .$$

In this sense, using POVMs, which are more general than projective measurements, is a sensible way of mapping the physical concept of measurements to mathematical objects.

2.2 Quantum operations

We have introduced mathematical objects which correspond to states and measurements. As a next step, we seek a mathematical description of allowed quantum operations. Since quantum operations should map quantum systems to quantum systems, we want a quantum operation from a quantum system A to a quantum system B to map states on $\mathcal{H}_A$ (i.e., elements of $\mathcal{S}(\mathcal{H}_A)$) to states on $\mathcal{H}_B$. Since quantum mechanics is linear, the operation itself should also preserve this structure and be linear. It is easy to see that in order to map states to states, a linear map $\mathcal{E} : \mathcal{B}_1(\mathcal{H}_A) \to \mathcal{B}_1(\mathcal{H}_B)$ necessarily needs to be *positive* (i.e., $\mathcal{E}(\rho) \geq 0$ whenever $\rho \geq 0$) and *trace-preserving* (i.e., $\operatorname{tr}(\mathcal{E}(\rho)) = \operatorname{tr}(\rho)$ for all $\rho \in \mathcal{B}_1(\mathcal{H}_A)$). However, since it is possible that the system we consider is a subsystem of a larger system, a stronger notion than the preservation of positivity is necessary for a map to be a quantum operation. This notion is that of complete positivity.

Definition 2.2.1 (Complete positivity). *A linear map $\mathcal{E} : \mathcal{B}_1(\mathcal{H}_A) \to \mathcal{B}_1(\mathcal{H}_B)$ is called completely positive if the map*

$$\mathcal{E} \otimes \mathbb{1}_{\mathcal{B}(\mathbb{C}^d)} : \mathcal{B}_1(\mathcal{H}_A) \otimes \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}_1(\mathcal{H}_B) \otimes \mathcal{B}(\mathbb{C}^d)$$

is positive for all $d \in \mathbb{N}$.

The notion of complete positivity is different from the notion of positivity: For example, the transposition map $\Theta : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$, $\Theta(X) := X^T$, where we identify $\mathbb{C}^{d\times d}$ with $\mathcal{B}(\mathbb{C}^d)$, is a linear map which is positive but not completely positive.

We call linear maps with the property that they map quantum states to quantum states *quantum channels.*

Definition 2.2.2 (Quantum channel). *A linear map $\mathcal{E} : \mathcal{B}_1(\mathcal{H}_A) \to \mathcal{B}_1(\mathcal{H}_B)$ is called a* quantum channel *if it is completely positive and trace-preserving (CPTP).*

This notion is in the Schrödinger picture, where quantum channels act on states. One can also define quantum channels in the Heisenberg picture, in which they act on observables.

Note that in general, the dual space of the trace-class operators $\mathcal{B}_1(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ is isomorphic to the bounded linear operators[2] $\mathcal{B}(\mathcal{H})$, with the duality given by

$$\langle T, A\rangle = \operatorname{tr} TA \qquad \text{for } T \in \mathcal{B}_1(\mathcal{H}), A \in \mathcal{B}(\mathcal{H}) \ .$$

Then the translation to the Heisenberg picture works simply by introducing the dual map to a Schrödinger-picture quantum channel, $\mathcal{E}^* : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$:

$$\operatorname{tr}(\rho\mathcal{E}^*(A)) = \operatorname{tr}(\mathcal{E}(\rho)A) \qquad \text{for all } \rho \in \mathcal{B}_1(\mathcal{H}), A \in \mathcal{B}(\mathcal{H}) \ ,$$

which makes sure that expectation values are left unchanged. In this formulation, quantum channels $\mathcal{E}^*$ are then not CPTP, but completely positive and unital (CPU), where unitality means that $\mathcal{E}^*(\mathbb{1}_{\mathcal{H}_B}) = \mathbb{1}_{\mathcal{H}_A}$. Dual maps of CPTP maps are also *normal*, which means that $\mathcal{E}^*(A_k) \uparrow \mathcal{E}^*(A)$ if $A_k \uparrow A$. For infinite-dimensional Hilbert spaces, a general CPU map $\mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_A)$ does not have a dual CPTP map unless it is normal.

Completely positive maps have some very useful properties. A powerful description of completely positive maps is given by the *Stinespring dilation.*

Theorem 2.2.3 (Stinespring dilation [17, Chapter 9])**.** *A linear map $\mathcal{E} : \mathcal{B}_1(\mathcal{H}_A) \to \mathcal{B}_1(\mathcal{H}_B)$ is completely positive if and only if there exists a Hilbert space $\mathcal{H}_E$ and a bounded linear operator $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E$ such that*

$$\mathcal{E}(\rho) = \operatorname{tr}_E(V\rho V^\dagger) \qquad \text{for all } \rho \in \mathcal{B}_1(\mathcal{H}_A) \ .$$

Furthermore, $\mathcal{E}$ is trace-preserving if and only if $V^\dagger V = \mathbb{1}$.

Physically, the Stinespring dilation theorem tells us that any quantum channel can be realized as an isometry on a larger system. The Hilbert space $\mathcal{H}_E$ is therefore often called the environment system. If we interchange the role of the system A and the environment E in the Stinespring dilation, we obtain the so-called *complementary channel* of $\mathcal{E}$, which we call $\mathcal{E}^\# : \mathcal{B}_1(\mathcal{H}_A) \to \mathcal{B}_1(\mathcal{H}_E)$:

$$\mathcal{E}^\#(\rho) := \operatorname{tr}_A\left(V\rho V^\dagger\right) \qquad \text{for } \rho \in \mathcal{B}_1(\mathcal{H}_A) \ .$$

Another equivalent way to describe completely positive maps is by their so-called *Kraus representation*, which we formulate in the Heisenberg picture.

Theorem 2.2.4 (Kraus representation [17, Chapter 9])**.** *A normal positive linear map $\mathcal{E}$: $\mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ (where $\mathcal{H}$ is separable) is completely positive if and only if there exists a countable family of bounded operators $\{M_k\}_{k=1}^\infty$ on $\mathcal{H}$ such that*

$$\mathcal{E}(X) = \sum_{k=1}^\infty M_k^\dagger X M_k \qquad \textit{for all } X \in \mathcal{B}(\mathcal{H}) \ .$$

If $\mathcal{E}$ is unital, then $\sum_{k=1}^\infty M_k^\dagger M_k = \mathbb{1}$.

We have discussed the most basic concepts of quantum mechanics from the viewpoint of quantum information theory. We continue by asking a more information-theoretic, but fundamental question: Given a quantum channel $\mathcal{E}$, what is the maximal amount of information we can transmit via such a channel? In the next chapter, we present some tools which enable us to deal with this fundamental question and also define what we mean by the term "amount of information".

[2] Note that in the case of an infinite-dimensional Hilbert space, the converse is not true: trace-class operators are the dual of *compact operators*, not the full set of bounded linear operators. This gives rise to a number of subtleties which we are able to ignore here because we have assumed that quantum states are normal.

3 Classical communication over quantum channels

On a fundamental level, information is encoded in a physical system. Since quantum mechanics is the description of the physical world on this fundamental level, one can ask what limits quantum mechanics imposes on communication. This question lies at the heart of the field of quantum communication. We will implicitly assume some concepts from classical information theory, a detailed exposition of which can be found in [23]. We focus on the problem of communicating *classical information* via quantum systems here. First, let us briefly present the communication problem from classical information theory. Throughout this chapter, when we write log we mean the logarithm to base 2, as this makes some operational motivations clearer. In subsequent chapters, we are going to use the natural logarithm purely for convenience.

3.1 Communication via classical channels

The presentation of this section largely follows [24, Chapter 7.2]. Consider two parties which we call Alice and Bob. Alice has a set of messages $\{1,\dots,2^M\}$, $M\in\mathbb{N}$, and wants to communicate one of them to Bob via a classical channel T with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{B}$. The sets $\mathcal{A}$ and $\mathcal{B}$ are for now assumed to be finite and called the *input alphabet* and the *output alphabet*. Such a channel T is given by a conditional probability $T(b|a)$ of obtaining the output symbol b if the input symbol was a, for each $a\in\mathcal{A}$ and $b\in\mathcal{B}$.

As M might be very large, we want to allow Alice to use the channel $L\in\mathbb{N}$ times. The L-times use of the channel T then induces a channel $T^{(L)}$ with input alphabet $\mathcal{A}^L$ and output alphabet $\mathcal{B}^L$, given by

$$T^{(L)}\left((b_1,\dots,b_L)|(a_1,\dots,a_L)\right)=T(b_1|a_1)\cdots T(b_L|a_L)\ .$$

This channel is referred to as the *discrete memoryless channel without feedback*, i.e., the different uses of T act independently on their respective inputs, and the different inputs do not depend on the outputs.

A $(2^M,L)$ *code* consists of an encoding function $\mathrm{enc}_L:\{1,\dots,2^M\}\to\mathcal{A}^L$ and a decoding function $\mathrm{dec}_L:\mathcal{B}^L\to\{1,2,\dots,2^M\}$. The encoding associates a codeword to a message, and the decoding maps every channel output to a message. Write

$$\lambda_m=\sum_{(b_1,\dots b_L)\notin\mathrm{dec}_L^{-1}(\{m\})}T^{(L)}\left((b_1,\dots,b_L)|\mathrm{enc}_L(m)\right)$$

for the probability that $1\leq m\leq 2^M$ was sent over the channel but m was not received. Then the *average probability of error* $P_e^{(L)}$ of the code is given by

$$P_e^{(L)}:=\frac{1}{2^M}\sum_{m=1}^{2^M}\lambda_m\ .$$

The *rate* of a $(2^M, L)$ code is defined as $R = M/L$. Given a channel T, a given rate R is *achievable* if there exists a sequence of $(\lceil 2^{LR} \rceil, L)$ codes such that the average probability of error tends to 0 as $L \to \infty$. The *capacity* of the channel is the supremum of all achievable rates.

Let us define an information-theoretic quantity which seems, at first glance, unrelated to this problem, and link it to the capacity. Given two random variables A and B which take values in $\mathcal{A}$ and $\mathcal{B}$, respectively, we define the *mutual information* between A and B as

$$I(A:B) = H(A) + H(B) - H(AB) \ , \tag{3.1}$$

where $H(A) = -\sum_{a\in\mathcal{A}} p_a \log p_a$ is the Shannon entropy of the random variable A and $\{p_a\}_{a\in\mathcal{A}}$ is the probability distribution of A, $p_a = \Pr(A = a)$. The quantities $H(B)$ and $H(AB)$ are defined analogously, using the probability distributions $\{p_b\}_{b\in\mathcal{B}}$ of B and the joint distribution $\{p_{ab} = \Pr(A = a \text{ and } B = b)\}_{a\in\mathcal{A},b\in\mathcal{B}}$. Given a channel T, we define the *Shannon capacity* C of the channel as the supremum of all mutual informations $I(A:B)$ between the input A and the output B of the channel, over all probability distributions $\{p_a\}_{a\in\mathcal{A}}$ on the input alphabet $\mathcal{A}$:

$$C = \sup_{\{p_a\}_{a\in\mathcal{A}}} I(A:B) \ . \tag{3.2}$$

A central result of classical information theory is Shannon's noisy channel coding theorem [9], which states that the Shannon capacity of a channel is equal to its capacity.

Theorem 3.1.1 (Channel coding theorem [24, Theorem 7.2])**.** *If $R < C$ then there exists a sequence of $(\lceil 2^{LR} \rceil, L)$ codes such that the average probability of error $P_{\mathrm{e}}^{(L)}$ tends to 0 for $L \to \infty$. Conversely, if a sequence of $(\lceil 2^{LR} \rceil, L)$ codes has average probability of error tending to 0, then $R \leq C$.*

The channel coding theorem underlines that the mutual information $I(A:B)$ is a good measure for the amount of information transmitted over a channel. The proof proceeds by averaging the error probability over random codes, and arguing that if an average random code has small error probability, then there exists one particular code which has small error probability. This *random coding* argument is a powerful one which is encountered often in information theory.

3.1.1 The channel coding theorem for continuous alphabets

The channel coding theorem 3.1.1 remains valid for a channel T in the case of continuous alphabets $\mathcal{A}$ and $\mathcal{B}$, with some modifications. Such a continuous-variable channel transforms probability densities on $\mathcal{A}$ to probability densities on $\mathcal{B}$. Assume that $\mathcal{A} \subset \mathbb{R}^n$. The differential entropy [9,10] of an $\mathcal{A}$-valued random variable X with probability density function $f_X : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$H(X) = -\int_{\mathcal{A}} f_X(x) \log f_X(x) \mathrm{d}^n x \ ,$$

where we have used the convention that[1] $0 \log 0 = 0$. The differential entropy depends only on the probability density of the random variable X, and hence we will often write $H(f_X)$ instead of $H(X)$. The mutual information $I(A:B)$ is defined analogously to Eq. (3.1), by replacing the Shannon entropy by the differential entropy. The Shannon capacity C defined in the same way as in Eq. (3.2) of a classical channel T is not finite in general. In order to

[1] As an alternative to this convention, we could have defined the entropy by integrating only over the support of X.

make a meaningful statement about the capacity of a continuous-variable channel, a constraint needs to be introduced on the input. The most common such constraint is a *power constraint*, which demands that any *codeword* $(x_1, \ldots, x_L) \in f_L\left(\{1, \ldots 2^M\}\right)$ in the image of the encoding function of a code satisfies

$$\frac{1}{L}\sum_{k=1}^{L} x_k^2 \leq P \ , \tag{3.3}$$

for some $P > 0$. The (Shannon) *power-constrained capacity* of a continuous-variable channel is defined by taking the supremum of the mutual information over all input probability densities $f_A : \mathcal{A} \to \mathbb{R}$ satisfying $\mathbb{E}[A^2] \leq P$,

$$C_P = \sup_{\substack{f_A:\mathcal{A}\to\mathbb{R} \\ \mathbb{E}[A^2]\leq P}} I(A:B) \ .$$

An analog of the channel coding theorem 3.1.1 holds when replacing the capacity C with the power-constrained capacity C_P. That is, the maximal achievable rate with codes satisfying the power constraint (3.3) for $P > 0$ is equal to C_P. In the next section, we move from classical information theory to quantum information theory and start with the problem of transmitting classical information via the preparation and measurement of quantum states.

3.2 A communication problem in quantum mechanics

We formulate a general communication problem in the quantum setting. Alice wants to send classical information to Bob using a quantum system Q (described by a Hilbert space $\mathcal{H}$). Assume that Alice picks a finite alphabet $\mathcal{A}$ and chooses corresponding quantum states $\{\rho_a\}_{a\in\mathcal{A}} \subset \mathcal{S}(\mathcal{H})$. The map

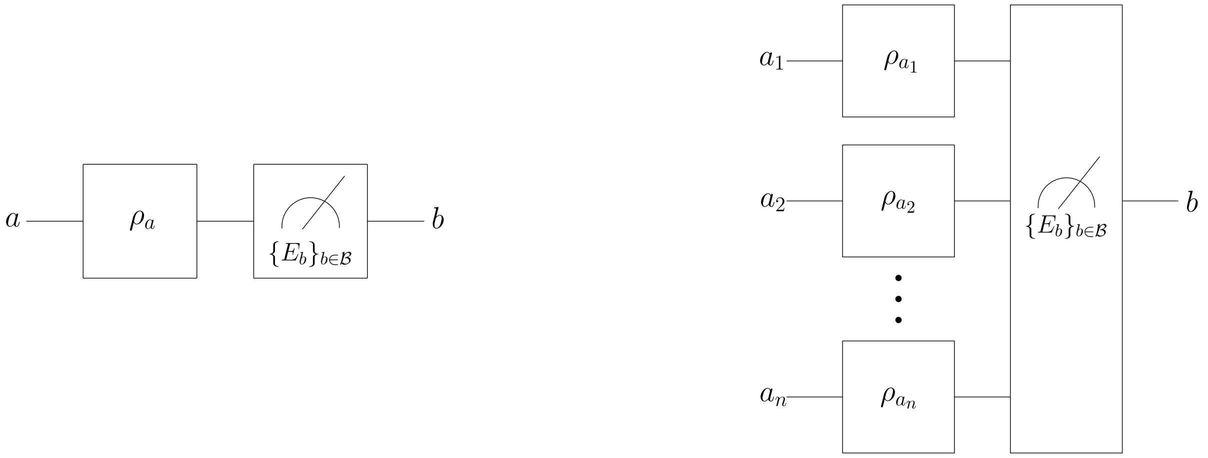
$$\begin{aligned} \mathcal{A} &\to \mathcal{S}(\mathcal{H}) \ , \\ a &\mapsto \rho_a \end{aligned}$$

is called a *classical-quantum* (cq) channel.

Suppose Alice prepares ρ_a according to the outcome of some $\mathcal{A}$-valued random variable A with probability distribution $\{p_a\}_{a\in\mathcal{A}}$. Bob then wants to find the value of A by performing a measurement (which is described by a POVM $\{E_b\}_{b\in\mathcal{B}}$) on the states ρ_a, obtaining a classical output random variable B. The probability of obtaining the output b if the input was a is equal to $T(b|a) = \mathrm{tr}(\rho_a E_b)$, corresponding to a classical channel T with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{B}$. If Alice chooses mutually orthogonal states, then Bob can distinguish them perfectly, by choosing $\mathcal{B} = \mathcal{A}$ and E_a to be the projection onto the support of ρ_a.

But since quantum states are in general non-orthogonal, Bob will in general not be able to distinguish perfectly between them. This setting is depicted in Fig. 3.1a. We can ask how much information he can obtain about the random variable A. One measure of information is the so-called *accessible information.* This is the maximum value of the classical mutual information (3.1) between the two random variables A and B over all possible measurements which Bob can perform:

$$I_{\mathsf{acc}}(\{p_a, \rho_a\}_{a\in\mathcal{A}}) = \sup_{\{E_b\}_{b\in\mathcal{B}}} I(A:B) = \sup_{\{E_b\}_{b\in\mathcal{B}}} \sum_{a\in\mathcal{A}}\sum_{b\in\mathcal{B}} p_a T(b|a) \log\left(\frac{T(b|a)}{\sum_{k\in\mathcal{A}} p_a T(b|a)}\right) \ .$$



(a) Single use of a cq channel: Alice prepares a state ρ_a according to the outcome a of a random variable A, and Bob measures the state, obtaining the outcome of a random variable b.

(b) Multiple uses of a cq channel: Alice prepares a state $\rho_{a_1} \otimes \cdots \otimes \rho_{a_n}$ according to the outcomes of multiple iid drawings of the random variable A, and Bob collectively measures the state, obtaining the outcome of a random variable B.

Figure 3.1: The definition of our first communication problem in quantum mechanics for one-time use and multiple uses of the cq channel.

The motivation of using the accessible information as our measure of information is that for every choice of states $\{\rho_a\}_{a\in\mathcal{A}}$ and Bob's POVM, our setting is described by the classical channel T introduced above. By the channel coding theorem, the capacity of each of these channels is given by the supremum of the mutual information $I(A:B)$ over all choices of probability distributions p_a. Therefore, if we maximize the accessible information with respect to all choices of ensembles $\{p_a,\rho_a\}_{a\in\mathcal{A}}$, we expect to obtain a meaningful quantity in the context of communication of classical information via quantum systems, accounting for Alice's and Bob's freedom in choosing the cq channel and POVM. In what sense this is the case is the content of the next sections.

3.2.1 The Holevo bound

The calculation of the accessible information involves a maximization over all possible POVMs and is generally a nontrivial optimization problem. However, a theorem by Holevo gives a useful upper bound on this quantity.

Theorem 3.2.1 (Holevo [25–27])**.** *Let $\{p_a,\rho_a\}_{a\in\mathcal{A}}$ be a finite ensemble of states ρ_a with probability distribution p_a. Then the accessible information is bounded by*

$$I_{\mathsf{acc}}(\{p_a,\rho_a\}_{a\in\mathcal{A}}) \leq \chi(\{p_a,\rho_a\}_{a\in\mathcal{A}}) := S(\overline{\rho}) - \sum_{a\in\mathcal{A}} p_a S(\rho_a) \ ,$$

where $S(\rho) = -\operatorname{tr}(\rho\log\rho)$ is the von Neumann entropy of the state ρ and $\overline{\rho} = \sum_{a\in\mathcal{A}} p_a\rho_a$ is the average signal state. The quantity $\chi(\{p_a,\rho_a\}_{a\in\mathcal{A}})$ is called the Holevo quantity *of the ensemble $\{p_a,\rho_a\}_{a\in\mathcal{A}}$.*

Proof (following [28]). Suppose Alice records her state in a classical register A (which we represent by a Hilbert space $\mathbb{C}^{|\mathcal{A}|}$ with orthonormal basis vectors $\{|a\rangle\}_{a\in\mathcal{A}}$). Then the joint classical-quantum state on Alice's classical register and the quantum system Q is given by

$$\rho_{AQ} = \sum_{a\in\mathcal{A}} p_a \, |a\rangle\langle a| \otimes \rho_a \ .$$

Bob's measurement, described by the POVM $\{E_b\}_{b\in\mathcal{B}}$, can be described by a quantum channel which maps Q to QB (where B is Bob's classical register with basis vectors $\{|b\rangle\}_{b\in\mathcal{B}}$) as

$$\rho_a \mapsto \sum_{b\in\mathcal{B}} M_b \rho_a M_b^\dagger \otimes |b\rangle\langle b| \ ,$$

where $M_b^\dagger M_b = E_b$. The full state on AQB after Bob's measurement is then

$$\rho'_{AQB} = \sum_{a\in\mathcal{A}, b\in\mathcal{B}} p_a \, |a\rangle\langle a| \otimes M_b \rho_a M_b^\dagger \otimes |b\rangle\langle b| \ .$$

We introduce the *quantum mutual information* $I(X:Y)_{\rho_{XY}} = S(\rho_X) - S(\rho_{XY}) + S(\rho_Y)$, which is a quantity analogous to the classical mutual information, but replaces the entropy of random variables with the von Neumann entropy of quantum states. We then have

$$I(A:B)_{\rho'} \leq I(A:QB)_{\rho'} \leq I(A:Q)_\rho \ , \tag{3.4}$$

where we have used strong subadditivity (i.e., $I(A:B) \leq I(A:BC)$ for A, B, C classical registers or quantum systems [29, Chapter 11.4]) and monotonicity under CPTP maps (i.e., $I(A:B)_{\mathcal{E}(\rho)} \leq I(A:B)_\rho$ for a CPTP map $\mathcal{E}$ [29, Theorem 11.15]), in this case applied for $\mathcal{E}$ being the channel describing Bob's measurement. Since the left-hand side of Eq. (3.4) is equal to the classical mutual information (3.1), it remains to show that we have

$$I(A:Q)_\rho \ = \ \chi(\{p_a, \rho_a\}_{a\in\mathcal{A}}) \ .$$

We calculate

$$\begin{aligned} S(AQ) &= -\operatorname{tr}\left(\left(\sum_{a\in\mathcal{A}} p_a \, |a\rangle\langle a| \otimes \rho_a\right) \log\left(\sum_{a'\in\mathcal{A}} |a'\rangle\langle a'| \otimes \rho_{a'}\right)\right) \\ &= -\sum_{a\in\mathcal{A}} \operatorname{tr}\left(p_a \rho_a \left(\log p_a + \log \rho_a\right)\right) \\ &= -\sum_{a\in\mathcal{A}} p_a \log p_a \operatorname{tr} \rho_a - \sum_{a\in\mathcal{A}} p_a \operatorname{tr}\left(\rho_a \log \rho_a\right) \\ &= H(A) + \sum_{a\in\mathcal{A}} p_a S(\rho_a) \ , \end{aligned}$$

giving $S(Q|A) = H(AQ) - H(A) = \sum_{a\in\mathcal{A}} p_a S(\rho_a)$. By definition of the quantum mutual information,

$$\begin{aligned} I(A:Q)_\rho &= S(\rho_Q) - S(\rho_{QA}) + S(\rho_A) \\ &= S(\rho_Q) - H(A) - \sum_{a\in\mathcal{A}} p_a S(\rho_a) + H(A) \\ &= S\left(\sum_{a\in\mathcal{A}} p_a \rho_a\right) - \sum_{a\in\mathcal{A}} p_a S(\rho_a) \ . \end{aligned}$$

The claim now follows with Eq. (3.4). □

The von Neumann entropy $S(\rho)$ in this theorem is a central quantity of interest in quantum information theory. It corresponds to the well-established Shannon entropy from classical information theory. It plays a central role throughout this thesis.

Theorem 3.2.1 has a striking consequence, namely that the maximal amount of information which can be retrieved from n qubits (where the Hilbert space is given by $\mathcal{H} = \left(\mathbb{C}^2\right)^{\otimes n}$, and $S(\rho) \leq \log(2^n) = n$ for any $\rho \in \mathcal{S}(\mathcal{H})$) is merely n classical bits. In this sense, qubits cannot store more information than classical bits can.

3.2.2 The converse to the Holevo bound

It is natural to ask whether the bound given by the Holevo quantity on the accessible information can be achieved by a suitable choice of Bob's measurement. Before we address this question, let us introduce some terminology. Fix an ensemble $\{p_a, \rho_a\}_{a\in\mathcal{A}}$. For $L \in \mathbb{N}$, we define the ensemble $\{P_a^{(L)}, \rho_a^{(L)}\}_{a\in\mathcal{A}^L}$ according to

$$P_a = p_{a_1} \cdots p_{a_L} \qquad \text{and} \qquad \rho_a = \rho_{a_1} \otimes \cdots \otimes \rho_{a_L} \in \mathcal{S}(\mathcal{H}^{\otimes L})$$

for all $a = (a_1, \dots a_L) \in \mathcal{A}^L$. Suppose Alice prepares the state ρ_a with probability P_a and Bob performs a measurement described by a POVM $\{E_b\}_{b\in\mathcal{B}} \subset \mathcal{B}(\mathcal{H}^{\otimes L})$. This protocol can be described by a classical channel $T^{(L)}$ with input alphabet $\mathcal{A}^L$, output alphabet $\mathcal{B}$, and conditional probabilities $T^{(L)}(b|a) = \operatorname{tr}\left(\rho_a^{(L)} E_b\right)$. The mutual information between Alice's input and Bob's output is given by

$$I^{(L)}(A:B) = \sum_{b\in\mathcal{B}} \sum_{a\in\mathcal{A}^L} P_a T^{(L)}(b|a) \log\left(\frac{T^{(L)}_{b|a}}{\sum_{k\in\mathcal{A}^L} P_a T^{(L)}_{b|k}}\right) . \tag{3.5}$$

This setting is depicted in Fig. 3.1b. By allowing Alice to use such "extended" ensembles for arbitrarily large L and Bob to collectively measure the output, the upper bound from Theorem 3.2.1 is in fact asymptotically achievable: By picking a sufficiently large L and choosing a suitable measurement POVM, we can come close to the Holevo bound for the ensemble $\{p_a, \rho_a\}_{a\in\mathcal{A}}$. This is the content of the Holevo-Schumacher-Westmoreland (HSW) theorem.

Theorem 3.2.2 (Holevo, Schumacher, Westmoreland [30,31]; [24, Theorem 7.8])**.** *Suppose we have an ensemble $\{p_a, \rho_a\}_{a\in\mathcal{A}}$ of states $\rho_a \in \mathcal{S}(\mathcal{H})$ with a priori probabilities p_a, and fix $\delta > 0$. Then, there is $L \in \mathbb{N}$ and a POVM $\{E_b\}_{b\in\mathcal{B}} \subset \mathcal{B}(\mathcal{H}^{\otimes L})$ for some finite alphabet $\mathcal{B}$ such that the mutual information from Eq.* (3.5) *satisfies*

$$\frac{I^{(L)}(A:B)}{L} \geq \chi\left(\{p_a, \rho_a\}_{a\in\mathcal{A}}\right) - \delta .$$

3.3 The classical capacity of noisy quantum channels

So far we have assumed that there is no noise affecting the system which Alice and Bob use for communication. Let us generalize our setup to the setting where Alice and Bob want to use a *memoryless quantum channel* for communication.

Alice wants to send classical information to Bob. As before, she has a set of messages $\{1, \dots, 2^M\}$, $M \in \mathbb{N}$, and wants to communicate one of them to Bob, this time via n uses of a *quantum* channel $\mathcal{E} : \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_B)$ which maps Alice's quantum system A to Bob's quantum system B. Multiple uses of the quantum channel $\mathcal{E}$ are modeled by the quantum channel $\mathcal{E}^{\otimes n} : \mathcal{S}(\mathcal{H}_A^{\otimes n}) \to \mathcal{S}(\mathcal{H}_B^{\otimes n})$. This quantum channel models the n-times use of the channel $\mathcal{E}$ in a memoryless way.

A $(2^M, n)$ *quantum code* consists of an encoding and a decoding. An encoding is a collection of *codeword states* $\{\rho_m\}_{m=1}^{2^M} \subset \mathcal{S}\left(\mathcal{H}_A^{\otimes n}\right)$. A decoding is a set of positive decoding operators $\{E_b\}_{b=1}^{2^M} \subset \mathcal{B}(\mathcal{H}^{\otimes n})$ which satisfy $\sum_{b=1}^{2^M} E_b \leq \mathbb{1}_{\mathcal{H}^{\otimes n}}$. The interpretation of this quantum code is that Alice can prepare her system (described by a Hilbert space $\mathcal{H}_A$) in a codeword state $\rho_m \in \mathcal{S}\left(\mathcal{H}_A^{\otimes M}\right)$ which depends on her choice of message $m \in \{1, \ldots, 2^M\}$. Bob receives the output $\mathcal{E}^{\otimes n}(\rho_m)$ and can perform a *decoding* measurement to decide which message was sent to him. The measurement which Bob performs on his system is described by the POVM $\{E_b\}_{b=0}^{2^M} \subset \mathcal{B}(\mathcal{H}_B^{\otimes n})$, with $E_0 = \mathbb{1} - \sum_{b=1}^{2^M} E_b$. If Bob obtains the output m for $1 \leq m \leq 2^M$, he decides that the message m was transmitted, whereas in the case of the output 0, the decoding fails. The average error probability of this code is given by

$$P_e^{(n)} = \frac{1}{2^M} \sum_{m=1}^{2^M} \left(1 - \operatorname{tr}(\mathcal{E}^{\otimes n}(\rho_m) E_m)\right) \ ,$$

and the *rate* of this code is $R = \frac{M}{n}$. As before, a rate R is achievable if there exists a sequence of $(\lceil 2^{nR} \rceil, n)$ quantum codes such that the average probability of error $P_e^{(n)}$ tends to 0 as $n \to \infty$. The classical capacity of the quantum channel $\mathcal{E}$ is then defined as the maximal achievable rate.

Definition 3.3.1 (Classical capacity of a quantum channel)**.** *The (full) classical capacity $C(\mathcal{E})$ of the quantum channel $\mathcal{E} : \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_B)$ is the supremum of all achievable rates.*

In comparison with the previous section, Alice's task here becomes finding input states ρ_a such that Bob can still reliably distinguish the output states $\mathcal{E}(\rho_a)$. We allow Alice to choose her input states in order to maximize the information which Bob can receive on the other end of the channel. Suppose Alice only uses product states as inputs, i.e., $\rho_m = \rho_{a_1(m)} \otimes \cdots \otimes \rho_{a_n(m)}$, with $\rho_{a_k(m)} \in \mathcal{S}(\mathcal{H}_A)$ for $1 \leq k \leq n$ for all $1 \leq m \leq 2^M$. In this case, Bob will also receive product states, since memoryless channels have the property that they map product states to product states. This motivates the definition of the so-called *product state capacity*, the setting of which is depicted in Fig. 3.2a.

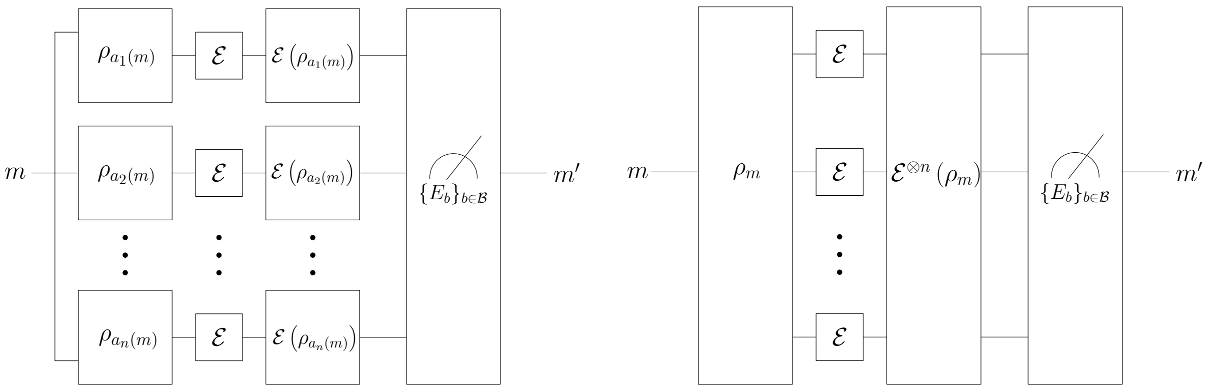
Definition 3.3.2 (Product state capacity)**.** *The product state capacity $C_1(\mathcal{E})$ of a quantum channel $\mathcal{E} : \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_B)$ is the maximal achievable rate if we only consider quantum codes whose codeword states are product states.*

With a slight modification of the proof of Theorem 3.2.2, it can be shown that the Holevo quantity of the ensemble $\{p_a, \mathcal{E}(\rho_a)\}_{a\in\mathcal{A}}$ can be achieved in rate by a suitable decoding scheme for sufficiently large codeword lengths, providing us with a formula for the product-state capacity.

Lemma 3.3.3 (Product state capacity [31])**.** *The product state capacity of a quantum channel $\mathcal{E} : \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_B)$ is given by*

$$C_1(\mathcal{E}) = \chi(\mathcal{E}) = \sup_{\{p_a, \rho_a\}_{a\in\mathcal{A}}} \chi(\{p_a, \mathcal{E}(\rho_a)\}) = \sup_{\{p_a, \rho_a\}_{a\in\mathcal{A}}} S(\mathcal{E}(\overline{\rho})) - \sum_{a\in\mathcal{A}} p_a S(\mathcal{E}(\rho_a)) \ .$$

The product state capacity is sometimes also called *one-shot capacity* of the channel. It can be shown that the product state capacity is also equal to the maximal achievable rate


m
$\rho_{a_1(m)}$
$\rho_{a_2(m)}$
$\rho_{a_n(m)}$
$\mathcal{E}$
$\mathcal{E}\left(\rho_{a_1(m)}\right)$
$\mathcal{E}\left(\rho_{a_2(m)}\right)$
$\mathcal{E}\left(\rho_{a_n(m)}\right)$
$\{E_b\}_{b\in\mathcal{B}}$
m'
m
ρ_m
$\mathcal{E}$
$\mathcal{E}^{\otimes n}\left(\rho_m\right)$
$\{E_b\}_{b\in\mathcal{B}}$
m'

(a) The product-state capacity: Alice is only allowed to use product states as codeword states, and Bob collectively measures the output of the channel $\mathcal{E}^{\otimes n}$.

(b) The full classical capacity: Alice can use generic codeword states $\rho_m \in \mathcal{S}(\mathcal{H}_A^{\otimes n})$, and Bob collectively measures the output of the channel $\mathcal{E}^{\otimes n}$.

Figure 3.2: The operational settings of the capacities $C_1(\mathcal{E})$ and $C(\mathcal{E})$.

using *separable* codeword states, i.e., convex combinations of product states [2,31]. However, a general state in $\mathcal{S}(\mathcal{H}_A^{\otimes n})$ is not separable, but *entangled*[2]. Hence if we want to find the ultimate limit on the transmission of information, we need to allow Alice to use general codeword states, as we did in Definition 3.3.1. This setting is depicted in Fig. 3.2b.

We can find a formula for the full capacity by considering the product state capacity of the channel $\mathcal{T} = \mathcal{E}^{\otimes n} : \mathcal{S}(\mathcal{H}_A^{\otimes n}) \to \mathcal{S}(\mathcal{H}_B^{\otimes n})$ for some $n \in \mathbb{N}$. As mentioned before, this channel models n parallel uses of the channel $\mathcal{E}$ (in a memoryless way). In this setting, Alice can prepare any input states in $\mathcal{S}(\mathcal{H}_A^{\otimes n})$ for the channel $\mathcal{T}$. The maximal achievable rate for this channel according to the HSW theorem is given by $\chi(\mathcal{T})$, and the maximal achievable rate *per use of the channel* $\mathcal{E}$ is given by $\chi(\mathcal{T})/n$. This is the capacity of the channel $\mathcal{E}$ if Alice is allowed to use the channel in entangled blocks of length n. Since product states are a special case of this, it is clear that

$$\frac{1}{n}\chi(\mathcal{E}^{\otimes n}) \geq \chi(\mathcal{E}) \qquad \text{for any } L \in \mathbb{N}\ . \tag{3.6}$$

The full classical capacity of the channel $\mathcal{E}$, in which quantum codes without restrictions on the codeword states or on n are considered, is then obtained by taking the limit $n \to \infty$. This is the ultimate limit on the transmission of classical information via the quantum channel $\mathcal{E}$.

Corollary 3.3.4 (Classical capacity [31])**.** *The full classical capacity $C(\mathcal{E})$ of a quantum channel $\mathcal{E}: \ \mathcal{S}(\mathcal{H}_A) \ \to \ \mathcal{S}(\mathcal{H}_B)$ is given by the limit*

$$C(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n} C_1(\mathcal{E}^{\otimes n})\ .$$

In literature, sometimes Lemma 3.3.3 and Corollary 3.3.4 are referred to as the HSW theorem, since they were conclusions originally stated in the same article as Theorem 3.2.2.

There are other capacities one can associate with a quantum channel. One of them is the *entanglement-assisted classical capacity* [32–34], in which Alice and Bob are allowed to share unlimited amounts of prior entanglement. Another important capacity is the *quantum capacity* [35,36], which measures the amount of *quantum* information which can be transmitted over

[2] A state is called entangled if it is not separable.

a quantum channel. While some of the results presented in this thesis have applications to the entanglement-assisted classical capacity, the classical capacity plays a central role in this thesis. Therefore we do not present details about other capacities here.

3.4 The additivity problem

The question whether the inequality in (3.6) is strict is referred to as the *additivity problem*. It comes from the terminology that if we have

$$\chi(\mathcal{E}^{\otimes n}) = n\chi(\mathcal{E}) \qquad \text{for all } n \in \mathbb{N} \ ,$$

then the quantity χ is said to be *additive* and we immediately obtain equality in (3.6). In this case, the use of entangled signal states gives no operational advantage over using only product states in the code. In a landmark result [37], Hastings proved that the Holevo quantity for finite-dimensional channels is in general non-additive, i.e., for sufficiently large d there exists a channel $\mathcal{T} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ for which

$$\chi(\mathcal{T}^{\otimes 2}) > 2\chi(\mathcal{T}) \ .$$

This leaves room for the classical capacity to be non-additive for some channels in the sense that

$$C(\mathcal{E}) > C_1(\mathcal{E}) \ .$$

Understanding for which channels the capacity may or may not be additive is a central problem in quantum information theory.

The discussion so far has been a fairly general treatment of the classical capacity of quantum channels. We want to focus on one particular model of quantum communication, which uses *bosonic* quantum systems. Bosonic quantum systems are a model for continuous-variable quantum systems, such as the quantum harmonic oscillator, the degrees of freedom of the electromagnetic field, and more. In the next chapter, we introduce the formalism of bosonic quantum information.

4 Bosonic quantum systems

In this chapter, we review the main concepts and the formalism of continuous-variable quantum information theory. For a more detailed exposition we refer to some of the many review papers on this topic [38–40]. A ubiquitous model for continuous-variable quantum systems is given by the bosonic harmonic oscillator of n modes, for $n \in \mathbb{N}$. We start by presenting the case for one mode, $n = 1$, which is instructive for the multimode case. For a one-mode bosonic system, we have two *quadrature operators*, referred to as the "position" and "momentum" operators (Q, P), which satisfy the canonical commutation relations

$$[Q, P] = i\mathbb{1} \ . \tag{4.1}$$

Equivalently, we can describe the system in terms of the *ladder operators*, which are a pair of bosonic field operators $a, a^\dagger$ which satisfy the bosonic commutation relations

$$[a, a^\dagger] = \mathbb{1} \ . \tag{4.2}$$

The Hilbert space of a one-mode bosonic system $\mathcal{H}_{\mathrm{osc}}$ is the separable Hilbert space spanned by an orthonormal system $\{|j\rangle\}_{j\in\mathbb{N}_0}$. The ladder operators act on the basis vectors in the following way:

$$\begin{aligned} a\,|0\rangle &= 0 \ , \\ a\,|j\rangle &= \sqrt{j}\,|j-1\rangle \qquad \text{for } j \geq 1 \ , \\ a^\dagger\,|j\rangle &= \sqrt{j+1}\,|j+1\rangle \qquad \text{for } j \in \mathbb{N}_0 \ . \end{aligned}$$

This clarifies the origin of the terminology *ladder operators*. The connection between the quadrature operators and the ladder operators is given by

$$a = \frac{1}{\sqrt{2}}(Q + iP) \ , \qquad a^\dagger = \frac{1}{\sqrt{2}}(Q - iP) \ , \tag{4.3}$$

and it can easily be checked that this relation is consistent with the commutation relations (4.1) and (4.2).

We are ready to treat the case of n modes, for $n \in \mathbb{N}$. Corresponding to each mode $k \in \{1, \ldots, n\}$, there is a set of quadrature operators (Q_k, P_k) which satisfy the canonical commutation relations:

$$[Q_j, P_k] = i\delta_{j,k}\mathbb{1} \qquad \text{for } j, k = 1, \ldots, n \ .$$

It is common to define a vector of quadrature operators $R = (Q_1, P_1, \cdots, Q_n, P_n)$ and write the canonical commutation relations in the following form:

$$[R_j, R_k] = i\Delta_{j,k}\mathbb{1} \qquad \text{for } j, k = 1, \ldots, 2n \ , \tag{4.4}$$

where $\Delta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus n}$ is the matrix of a symplectic form on $\mathbb{R}^{2n}$.

Equivalently, associated to each mode $k \in \{1, \ldots, n\}$, there is a pair of bosonic ladder operators $a_k, a_k^\dagger$, and these operators satisfy the bosonic commutation relations

$$[a_j, a_k^\dagger] = \delta_{j,k} \mathbb{1} \qquad \text{for } j, k = 1, \ldots, n \ .$$

The Hilbert space $\mathcal{H}^{\otimes n} = \mathcal{H}_{\mathrm{osc}}^{\otimes n}$ is the n-fold tensor power of the Hilbert space of a quantum harmonic oscillator of one mode. We obtain a basis of $\mathcal{H}$ from the one-mode basis vectors as $\{|j_1, \ldots, j_n\rangle = |j_1\rangle \otimes |j_2\rangle \otimes \cdots \otimes |j_n\rangle\}_{(j_1,\ldots,j_n)\in\mathbb{N}_0^n}$. The action of the ladder operators is then simply given by the action of the one-mode ladder operators after writing

$$a_k = \underbrace{\mathbb{1} \otimes \cdots \otimes \mathbb{1}}_{k-1 \text{ times}} \otimes a \otimes \underbrace{\mathbb{1} \otimes \cdots \otimes \mathbb{1}}_{n-k \text{ times}} \ .$$

The connection between the quadrature operators and the ladder operators of each mode is given by Eq. (4.3).

4.1 Phase space description

The canonical commutation relations (4.4) imply that the mode operators R_k need to be unbounded. A common realization of operators satisfying these commutation relations is the *Schrödinger representation*, where the Hilbert space of one mode is given by the square-integrable functions on the real line, $L^2(\mathbb{R}, \mathrm{d}x)$, and the quadrature operators act as

$$\begin{aligned} (Q\psi)(x) &= x\psi(x) \ , \\ (P\psi)(x) &= -i\frac{\mathrm{d}}{\mathrm{d}x}\psi(x) \qquad \text{for all } x \in \mathbb{R} \ . \end{aligned}$$

These are defined on a dense subset of $L^2(\mathbb{R}, \mathrm{d}x)$.

The vacuum state $|0\rangle$ then corresponds to the wavefunction $\psi_0(x) := \pi^{-\frac{1}{4}} e^{-\frac{x^2}{2}}$, and the other basis vectors can be obtained by applying the ladder operators using Eq. (4.3). Unfortunately, the Schrödinger representation allows for subtle ambiguities regarding the domains of the involved operators, a phenomenon which is illustrated by a counterexample from [41]. By moving to a description in terms of certain exponentials of the field operators, which are bounded, this problem can be overcome.

We introduce the so-called *phase space*, which is the vector space of $\mathbb{R}^{2n}$ equipped with the symplectic form $(x, y) \mapsto x^T \Delta^{-1} y$. The *Weyl displacement operators* are defined by

$$D(\xi) := \exp\left(i\xi^T(\Delta^{-1}R)\right) \qquad \text{for } \xi \in \mathbb{R}^{2n} \ .$$

They satisfy commutation relations on their own, namely

$$\begin{aligned} D(\xi)D(\eta) &= \exp\left(-\frac{i}{2}\xi^T(\Delta^{-1}\eta)\right) D(\xi + \eta) \ , \\ D(\xi)D(\eta) &= \exp\left(-i\xi^T(\Delta^{-1}\eta)\right) D(\eta)D(\xi) \qquad \text{for } \xi, \eta \in \mathbb{R}^{2n} \ . \end{aligned} \tag{4.5}$$

These commutation relations are called the *Weyl relations.* The reason why the Weyl operators are called “displacements” becomes apparent when investigating their action on the quadrature operators, which is

$$D(\xi)^\dagger R_j D(\xi) = R_j + \xi_j \mathbb{1} \qquad \text{for all } \xi \in \mathbb{R}^{2n} \ , j = 1, \ldots, 2n \ .$$

It is possible to regard the Weyl relations (as opposed to the canonical commutation relations) as a starting point. The generators R_k of a family of unitary operators which satisfy the Weyl relations themselves satisfy the canonical commutation relations[1] [41]. For this description, we introduce some terminology. A map $D : \mathbb{R}^{2n} \to \mathcal{B}(\mathcal{H})$ which satisfies the Weyl relations (4.5) and for which $D(\xi)$ are unitary for all $\xi \in \mathbb{R}^{2n}$ we call a *Weyl system*. A Weyl system is called *strongly continuous* if for all $\psi \in \mathcal{H}$, we have $\lim_{\xi \to 0} \|\psi - D(\xi)\psi\| = 0$. It is called *irreducible* if the only subspaces of $\mathcal{H}$ which are invariant under all $D(\xi)$ are $\{0\}$ and $\mathcal{H}$.

There is an advantage in using the Weyl description: For a finite-dimensional phase space, any two strongly continuous irreducible Weyl systems are unitarily equivalent. This is the content of the *Stone-von Neumann theorem*.

Theorem 4.1.1 (Stone-von Neumann [20, 41, 42])**.** *Let $D^{(1)} : \mathbb{R}^{2n} \to \mathcal{B}(\mathcal{H})$ and $D^{(2)} : \mathbb{R}^{2n} \to \mathcal{B}(\mathcal{H})$ be two strongly continuous irreducible Weyl systems over a finite-dimensional phase space. Then there exists a unitary operator U such that $D^{(1)}(\xi) = U^\dagger D^{(2)}(\xi) U$ for all $\xi \in \mathbb{R}^{2n}$.*

Hence if we only consider strongly continuous and irreducible Weyl systems, all these systems are equivalent to the Schrödinger representation and there is no ambiguity left. From here on, we write D for any strongly continuous and irreducible Weyl system and R_k for the associated generators.

4.1.1 The characteristic function and the Wigner function

The Weyl operators implement a form of a non-commutative Fourier transform, which gives us a duality between operators and complex functions on phase space. An n-mode quantum state ρ can be represented by its *characteristic function* $\chi_\rho \in L^2(\mathbb{R}^{2n}, \mathrm{d}^{2n}\xi)$, defined by

$$\chi_\rho(\xi) = \mathrm{tr}(D(\xi)\rho) \qquad \text{for } \xi \in \mathbb{R}^{2n} .$$

The characteristic function is well-defined for trace-class operators ρ. However, the map $\rho \mapsto \chi_\rho$ can be extended to the Hilbert-Schmidt class by continuity. Extended this way, it becomes an isometry between the Hilbert-Schmidt class and $L^2(\mathbb{R}^{2n}, \mathrm{d}^{2n}\xi)$, due to the Parseval relation [43]

$$\mathrm{tr}(\rho^\dagger \sigma) = (2\pi)^{-n} \int_{\mathbb{R}^{2n}} \overline{\chi_\rho(\xi)} \chi_\sigma(\xi) \, \mathrm{d}^{2n}\xi .$$

Given a characteristic function, the state can be reconstructed by taking the so-called *Weyl transform* of the characteristic function χ_ρ,

$$\rho = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^{2n}} \chi_\rho(\xi) D(\xi)^\dagger \, \mathrm{d}^{2n}\xi .$$

The characteristic function, being a function in $L^2(\mathbb{R}^{2n}, \mathrm{d}^{2n}\xi)$, can of course itself be viewed as the classical Fourier transform of a function instead of a non-commutative Fourier transform. This gives rise to the *Wigner function*: The characteristic function χ_ρ is the Fourier transform of the Wigner function W_ρ,

$$W_\rho(\xi) = \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} e^{i\xi^T(\Delta^{-1}\eta)} \chi_\rho(\eta) \, \mathrm{d}^{2n}\eta \qquad \text{for all } \xi \in \mathbb{R}^{2n} .$$

[1] The converse is not true: there exist operators R_k which satisfy the canonical commutation relations, but whose exponentials do not satisfy the Weyl relations [20, Example 14.5].

The Wigner function of a state is normalized, but not positive in general, hence it is a quasiprobability distribution.

Important quantities which give information about the characteristic function of a quantum state are its first and second moments. The vector of first moments of the quantum state ρ is called the *displacement vector*, which we denote by $d(\rho)$. Its entries are given by

$$d_k(\rho) := \mathrm{tr}[R_k \rho] \qquad \text{for } k = 1, \ldots, 2n \ .$$

The second moments make up the entries of the so-called covariance matrix $\Gamma(\rho)$,

$$\Gamma_{kl}(\rho) := \mathrm{tr}[\{R_k - d_k(\rho)\mathbb{1}, R_l - d_l(\rho)\mathbb{1}\}\rho] \qquad k, l = 1, \ldots, 2n \ ,$$

where $\{X, Y\} := XY + YX$ is the anticommutator. The covariance matrix is a $2n \times 2n$ real symmetric matrix which satisfies the uncertainty principle [44]

$$\Gamma(\rho) + i\Delta \geq 0 \ .$$

It turns out there is a particular class of quantum states for which the first and second moments capture all information about the state. These states are the so-called Gaussian states.

4.2 Gaussian states

Gaussian states are n-mode quantum states which have the property that their characteristic function is Gaussian. Since the first and second moments capture all information about these states, we will write $\rho_{\mathrm{G}}(d, \Gamma)$ for the Gaussian state with displacement vector $d \in \mathbb{R}^{2n}$ and covariance matrix $\Gamma \in \mathbb{R}^{2n \times 2n}$. This state has the characteristic function

$$\chi_{\rho_{\mathrm{G}(d,\Gamma)}}(\xi) = \exp\left[-\frac{1}{4}(\Delta^{-1}\xi)^T \Gamma(\Delta^{-1}\xi) + i\xi(\Delta^{-1}d)\right] \qquad \text{for all } \xi \in \mathbb{R}^{2n} \ .$$

A prime example of Gaussian states are the so-called *thermal states* whose covariance matrix is proportional to the identity. In the one-mode case, a thermal state is characterized by its average number of photons $N = \mathrm{tr}(a^\dagger a \rho)$ (also called the *average energy* or the *mean photon number*) and has the form

$$\rho_{\mathrm{th},N} = \frac{1}{N+1} \sum_{k=0}^{\infty} \left(\frac{N}{N+1}\right)^k |k\rangle\langle k| \ .$$

It has displacement vector $d = 0$ and covariance matrix $\Gamma = (2N+1)\,\mathbb{1}_2$. The von Neumann entropy of a one-mode thermal state is given by

$$S(\rho_{\mathrm{th},N}) = g(N) := (N+1)\log(N+1) - N\log N \ . \tag{4.6}$$

Another example of one-mode Gaussian states is given by the so-called *coherent states* $\rho_{\mathrm{G}}(\xi, \mathbb{1}_2)$ for $\xi \in \mathbb{R}^{2n}$: These states are pure and can be written as $|\xi\rangle\langle\xi|$, with

$$|\xi\rangle := D(\xi)\,|0\rangle \ .$$

It is easy to see that these states satisfy

$$a\,|\xi\rangle = \frac{\xi_1 + i\xi_2}{\sqrt{2}}\,|\xi\rangle \ ,$$

hence the coherent states are eigenstates of the annihilation operator a. Coherent states form a so-called *overcomplete basis*, and it is sometimes useful to describe states in terms of their coherent state expansion. For instance, the thermal state $\rho_{\mathrm{th},N}$ can be written as

$$\rho_{\mathrm{th},N} = \frac{1}{2\pi N}\int_{\mathbb{R}^2} e^{-\frac{|\xi|^2}{2N}}\ |\xi\rangle\langle\xi|\ \mathrm{d}^2\xi\ .$$

The following theorem is useful in the analysis of Gaussian states.

Theorem 4.2.1 (Williamson [45])**.** *Let A be a symmetric and positive $2n \times 2n$ matrix. Then A can be diagonalized by a symplectic transformation S (i.e., a $2n \times 2n$ matrix satisfying the equation $S^T\ \Delta\ S\ =\ \Delta$) such that*

$$SAS^T = \bigoplus_{j=1}^{n} \nu_j \mathbb{1}_2\ ,$$

where $\nu_j \geq 0$. The values ν_j are called the symplectic eigenvalues of A and are equal to the absolute values of the eigenvalues of $i\Delta^{-1}A$.

Applying Williamson's theorem to an arbitrary covariance matrix $\Gamma \in \mathbb{R}^{2n}$, we see that there exists a symplectic matrix S such that

$$S^T\Gamma S = \bigoplus_{k=1}^{n} \nu_k \mathbb{1}_2\ ,$$

with $\nu_k \geq 0$ for $k = 1,\ldots n$ (in fact, due to the uncertainty relation, we have $\nu_k \geq 1$). The symplectic transformation S induces a unitary U_S which realizes the action of S via conjugation:

$$U_S^\dagger R_k U_S = \sum_{j=1}^{2n} S_{kj} R_j\ . \tag{4.7}$$

We call such transformations U_S *Gaussian unitaries* (see Section 4.3). The unitary U_S transforms a state $\rho_{\mathrm{G}}(0,\Gamma)$ with zero displacement vector into a product of one-mode thermal states:

$$U_S \rho_{\mathrm{G}}(0,\Gamma) U_S^\dagger = \bigotimes_{k=1}^{n} \rho_{\mathrm{th},\frac{\nu_k-1}{2}}\ .$$

A general Gaussian state $\rho_{\mathrm{G}}(d,\Gamma)$ with nonzero displacement vector then has the decomposition

$$\rho_{\mathrm{G}}(d,\Gamma) = D(d)U_S^\dagger \left(\bigotimes_{k=1}^{n} \rho_{\mathrm{th},\frac{\nu_k-1}{2}}\right) U_S D(d)^\dagger\ .$$

This is called the *thermal decomposition* of the Gaussian state $\rho_{\mathrm{G}}(d,\Gamma)$. Since the von Neumann entropy is invariant under conjugation with unitaries, the entropy of a Gaussian state $\rho_{\mathrm{G}}(d,\Gamma)$ is given by

$$S(\rho_{\mathrm{G}}(d,\Gamma)) = \sum_{k=1}^{n} g\left(\frac{\nu_k - 1}{2}\right)\ .$$

The numbers $\frac{\nu_k-1}{2}$ are sometimes also referred to as the numbers of *thermal photons*, since they are equal to the mean photon numbers of the thermal states in the thermal decomposition of $\rho_{\mathrm{G}}(d,\Gamma)$.

4.2.1 Maximum entropy principle

Among all quantum states with given fixed covariance matrix and displacement vector, the Gaussian state maximizes the entropy [18, 46]. This is a useful fact in the context of bosonic information theory, and can be seen from the following Lemma.

Lemma 4.2.2 (Maximum entropy principle [18, Lemma 12.25])**.** *The Gaussian state $\rho_{\mathrm{G}}(d, \Gamma)$ has the largest entropy among all states with given first moments $d \in \mathbb{R}^{2n}$ and covariance matrix $\Gamma \in \mathbb{R}^{2n \times 2n}$. That is, for any n-mode quantum state ρ on a bosonic system with first and second moments given by d and Γ, we have*

$$S(\rho) \leq S(\rho_{\mathrm{G}}(d, \Gamma)) \ .$$

The proof proceeds by showing that for any n-mode quantum state ρ whose first and second moments are given by d and Γ, we have

$$S(\rho_{\mathrm{G}}(d, \Gamma)) = S(\rho) + D(\rho || \rho_{\mathrm{G}}(d, \Gamma)) \ , \tag{4.8}$$

where $D(\rho||\sigma) = \operatorname{tr}(\rho(\log \rho - \log \sigma))$ is the relative entropy between the two states ρ and σ. Because the relative entropy is nonnegative [18, Proposition 7.3], $D(\rho||\sigma) \geq 0$ for all states ρ, σ, it follows immediately from Eq. (4.8) that the Gaussian state maximizes the entropy among all states with given first and second moments.

4.3 Gaussian channels

Gaussian channels are quantum channels which map Gaussian states to Gaussian states, that is, a channel $\mathcal{E} : \mathcal{S}(\mathcal{H}^{\otimes n}) \to \mathcal{S}(\mathcal{H}^{\otimes m})$ is called Gaussian if $\mathcal{E}(\rho)$ is Gaussian whenever ρ is. The action of a Gaussian channel can be described by a triple (X, Y, η), where $\eta \in \mathbb{R}^{2m}$ is an arbitrary vector, and $X \in \mathbb{R}^{2n \times 2m}$, $Y \in \mathbb{R}^{2m \times 2m}$ are real matrices which satisfy $Y^T = Y$ and $Y + i(\Delta_{2m} - X^T \Delta_{2n} X) \geq 0$. Here, the index in Δ_{2m} is used to indicate the size of the matrix Δ, which we will suppress from here on. On a Gaussian state $\rho_{\mathrm{G}}(d, \Gamma)$, the Gaussian channel then acts as

$$\mathcal{E}(\rho_{\mathrm{G}}(d, \Gamma)) = \rho_{\mathrm{G}}(Xd + \eta, X^T \Gamma X + Y) \ .$$

On a general state ρ, the action of $\mathcal{E}$ can be described on the level of characteristic functions:

$$\chi_{\mathcal{E}(\rho)}(\xi) = \chi_\rho(X\xi) e^{-\frac{1}{4}(\Delta^{-1}\xi)^T Y (\Delta^{-1}\xi) + i\xi\Delta^{-1}\eta} \qquad \text{for all } \xi \in \mathbb{R}^{2n} \ .$$

A first example for a Gaussian channel for $n > m$ is the partial trace over the last $n - m$ modes, for which $Y = 0, \eta = 0$, and

$$X = \begin{pmatrix} \mathbb{1}_{2m} \\ 0_{(2n-2m) \times 2m} \end{pmatrix} \ .$$

On the level of characteristic functions, this amounts to evaluating the characteristic function while setting the phase space coordinates of the last $n - m$ modes to zero.

In the case where $n = m$, $Y = 0$, and $X = S$ is symplectic, in light of Eq. (4.7) we have $\mathcal{E}(\rho) = U_S \rho U_S^\dagger$ and the channel is described by a Gaussian unitary U_S.

4.3.1 The beamsplitter and squeezing

Let us consider a system of $2n$ modes, which we want to consider as the composition of two n-mode systems A and B. The quadratures are labelled by

$$R = (R_A, R_B) = (Q_{1,A}, P_{1,A}, \ldots, Q_{n,A}, P_{n,A}, Q_{1,B}, P_{1,B}, \ldots, Q_{n,B}, P_{n,B}) \ .$$

We define a Gaussian unitary $U_\lambda = U_{S_\lambda}$ via a symplectic transformation S_λ, for $\lambda \geq 0$, given by

$$S_\lambda = \begin{cases} \begin{pmatrix} \sqrt{\lambda}\mathbb{1}_{2n} & \sqrt{1-\lambda}\mathbb{1}_{2n} \\ -\sqrt{1-\lambda}\mathbb{1}_{2n} & \sqrt{\lambda}\mathbb{1}_{2n} \end{pmatrix} & \text{for } 0 \leq \lambda \leq 1 \ , \\ \begin{pmatrix} \sqrt{\lambda}\mathbb{1}_{2n} & \sqrt{\lambda-1}Z_{2n} \\ \sqrt{\lambda-1}Z_{2n} & \sqrt{\lambda}\mathbb{1}_{2n} \end{pmatrix} & \text{for } \lambda > 1 \ , \end{cases} \tag{4.9}$$

where $Z_{2n} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}^{\oplus n}$. The unitary U_λ implements the *beamsplitter* of *transmissivity* λ for $0 \leq \lambda \leq 1$ and the (two-mode) *squeezing* for $\lambda > 1$. The quantum channel $\rho \mapsto U_\lambda \rho U_\lambda^\dagger$ implements a Gaussian channel acting on a $2n$-mode system, with $X = S_\lambda$, $Y = 0$, $\eta = 0$.

For a quantum state $\sigma \in \mathcal{S}(\mathcal{H}_B)$ (the *environment state*) and $\lambda \geq 0$, we define the quantum channel

$$\mathcal{E}_{\lambda,\sigma}(\rho) = \mathrm{tr}_B \left[U_\lambda(\rho \otimes \sigma) U_\lambda^\dagger \right] \qquad \text{for all } \rho \in \mathcal{S}(\mathcal{H}_A) \ . \tag{4.10}$$

The channel $\mathcal{E}_{\lambda,\sigma}$ is a Gaussian channel if the environment σ is a Gaussian state. This follows immediately from the fact that the channels given by $\rho \mapsto U_\lambda \rho U_\lambda^\dagger$ and the partial trace are Gaussian. As an example, if $n = 1$, $0 \leq \lambda \leq 1$, and $\sigma = \rho_{\mathrm{th},N}$ is a thermal state with mean photon number N, then $\mathcal{E}_{\lambda,\rho_{\mathrm{th},N}}$ is a Gaussian channel with $X = \sqrt{\lambda}\mathbb{1}_2, Y = (1-\lambda)(2N+1)\mathbb{1}_2$, and $\eta = 0$. This channel is commonly referred to as the *thermal noise channel.* For $N = 0$ this channel is often called the *attenuation channel.* As another example, if $n = 1, \lambda > 1$, and $\sigma = |0\rangle\langle 0|$ is the vacuum state, then $\mathcal{E}_{\lambda,|0\rangle\langle 0|}$ implements a Gaussian channel with $X = \sqrt{\lambda}\mathbb{1}_2$, $Y = (\lambda - 1)\mathbb{1}_2$, and $\eta = 0$. This channel is also called the *amplification channel.*

4.3.2 The classical noise channel

The examples of channels we have seen so far coupled the system to an environment which was described by a quantum state. We can also consider noise which has a classical description acting on the system. For a probability density function $f : \mathbb{R}^{2n} \to \mathbb{R}$ on phase space and $t > 0$, we define

$$\mathcal{F}_{t,f}(\rho) := \int_{\mathbb{R}^{2n}} f(\xi) D(\sqrt{2\pi t}\xi) \rho D(\sqrt{2\pi t}\xi)^\dagger \, \mathrm{d}^{2n}\xi \qquad \text{for all } \rho \in \mathcal{S}(\mathcal{H}^{\otimes n}) \ . \tag{4.11}$$

This channel is called the *classical noise channel* because it can be understood to add classical noise to the system: The quantum state is displaced across phase space according to a classical probability distribution f (the parameter t is introduced purely for convenience). If f is Gaussian, then the channel $\mathcal{F}_{t,f}$ is a Gaussian channel.

If the classical "noise function" f is Gaussian, then the corresponding classical noise function is Gaussian. In the case when $f(\xi) = f_Z(\xi) = (2\pi\sigma^2)^{-n} e^{-\|\xi\|_2^2/(2\sigma^2)}$ is equal to a centered Gaussian whose covariance matrix is proportional to the identity, $\mathcal{F}_{t,f}$ implements a Gaussian channel with $X = \mathbb{1}_{2n}, Y = 4\pi t\sigma^2\mathbb{1}_{2n}$, $\eta = 0$.

The channels defined by the beamsplitter and squeezing unitaries as well as the classical noise channel are general models for bosonic quantum channels which cover many applications. In fact, in the one-mode case, they are among the main building blocks for any Gaussian quantum channel [47].

4.3.3 Non-Gaussian bosonic channels

It is easy to see that if the environment state σ of the channel $\mathcal{E}_{\lambda,\sigma}$ is not Gaussian, then the channel is not a Gaussian channel. Similarly, if the function f is not Gaussian, then the channel $\mathcal{F}_{t,f}$ is not a Gaussian channel. However, these channels still have more structure than a generic channel: Their Stinespring dilations (see Theorem 2.2.3) are given by Gaussian unitaries. Such channels are called *Gaussian-dilatable* [48]. Using this fact makes it possible to understand some of their properties and to use techniques from Gaussian information theory to prove statemens about these non-Gaussian channels. On the other hand, the study of generic bosonic channels, which do not have a Gaussian dilation, is out of the scope of this work.

We have introduced some basic concepts of continuous quantum information theory and quantum communication, and we can start with a discussion of particular topics which are central to the work presented in this thesis.

5 Entropic inequalities for bosonic quantum systems

Entropy is a key quantity of interest in all of information theory. Hence it is a central task of information theory to understand how entropy behaves when the system is subjected to noise. Entropic inequalities give bounds on the entropy production of channels and therefore make up important tools in the understanding of noise. In this chapter we give an overview of entropic inequalities for bosonic quantum systems. We start with one of the most fundamental examples of entropic inequalities, namely the data processing inequalities of classical and quantum information theory. Next we give a short introduction to additional entropic inequalities in classical information theory, which have inspired similar inequalities in the quantum setting. We then give an overview of a number of key results regarding entropic inequalities for bosonic systems, and show where the work of the present thesis fits in.

5.1 Data processing

Let us recall the definitions of the Shannon entropy and the corresponding mutual information from Eq. (3.1). A fundamental inequality governing these quantities is the data processing inequality, which deals with the behavior of information under certain operations.

Theorem 5.1.1 (Data processing inequality [23, Theorem 2.8.1 and Corollary])**.** *Let X, Y, Z be $\mathcal{A}$-valued random variables for a finite alphabet $\mathcal{A}$ which form a Markov chain, that is $X \to Y \to Z$. Then,*

$$I(X : Y) \geq I(X : Z) \ .$$

In particular, if $f : \mathcal{A} \to \mathcal{A}$ and $Z = f(Y)$ is a function of Y, we have

$$I(X : Y) \geq I(X : f(Y)) \ .$$

A common interpretation of the data processing inequality is that no manipulation of data whatsoever can improve the inferences which can be made from the data [23].

Similarly, the quantum data processing inequality states that the application of CPTP maps to each part of a bipartite quantum system cannot increase the mutual information between the two parts of the system.

Theorem 5.1.2 (Quantum data processing inequality [49, Theorem 11.9.4])**.** *Suppose that $\rho_{AB} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B)$ is a quantum state, where $\dim(\mathcal{H}_A), \dim(\mathcal{H}_B) < \infty$. Then, for all CPTP maps $\mathcal{N} : \mathcal{B}_1(\mathcal{H}_A) \to \mathcal{B}_1(\mathcal{H}_{A'})$ and $\mathcal{M} : \mathcal{B}_1(\mathcal{H}_B) \to \mathcal{B}_1(\mathcal{H}_{B'})$, we have*

$$I(A : B)_{\rho_{AB}} \geq I(A' : B')_{\rho'_{A'B'}} \ ,$$

where $\rho'_{A'B'} = (\mathcal{N} \otimes \mathcal{M}) (\rho_{AB})$.

There are many different formulations of data processing inequalities in a number of different settings, the common theme being that some measure of information cannot be increased by the processing of information. Such inequalities are among the most fundamental entropic inequalities in information theory. Let us now focus on the particular entropic inequalities which are important for the work presented in this thesis.

5.2 Entropic inequalities in classical information theory

For two $\mathbb{R}^n$-valued random variables X and Y, the sum $X+Y$ is defined as the random variable with probability density given by the convolution of the probability densities of X and Y, i.e.,

$$f_{X+Y}(z) := \int_{\mathbb{R}^n} f_X(x) f_Y(z-x) \mathrm{d}^n x \qquad \text{for } z \in \mathbb{R}^n \ .$$

For $t > 0$, the rescaled random variable $\sqrt{t}X$ is defined as the random variable with probability density given by $f_{\sqrt{t}X}(x) = t^{-\frac{n}{2}} f_X\left(\frac{x}{\sqrt{t}}\right)$ for $x \in \mathbb{R}^n$.

In his seminal 1948 paper [9,10], Shannon identified a central entropic inequality which gives a lower bound on the entropy of the sum of two random variables in terms of the individual entropies of the two random variables. It is the *entropy power inequality* (EPI).

Theorem 5.2.1 (Entropy power inequality [10,50–52])**.** *Let X and Y be two independent $\mathbb{R}^n$-valued random variables with finite second moments. Then we have*

$$e^{2H(X+Y)/n} \geq e^{2H(X)/n} + e^{2H(Y)/n} \ , \tag{5.1}$$

with equality if and only if X and Y are Gaussian random variables with proportional covariance matrices, that is, $\mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])^T] \propto \mathbb{E}[(Y-\mathbb{E}[Y])(Y-\mathbb{E}[Y])^T]$.

Remark 5.2.2. *Sometimes the entropy power inequality is stated in the form*

$$e^{2H(\sqrt{\lambda}X+\sqrt{1-\lambda}Y)/n} \geq \lambda e^{2H(X)/n} + (1-\lambda) e^{2H(Y)/n} \ ,$$

for $\lambda \in (0,1)$, where the random variable $\sqrt{\lambda}X + \sqrt{1-\lambda}Y$ has probability density

$$f_{\sqrt{\lambda}X+\sqrt{1-\lambda}Y}(z) = \frac{1}{\lambda^{\frac{n}{2}}(1-\lambda)^{\frac{n}{2}}} \int_{\mathbb{R}^n} f_X\left(\frac{x}{\sqrt{\lambda}}\right) f_Y\left(\frac{z-x}{\sqrt{1-\lambda}}\right) \mathrm{d}^n x \qquad \textit{for } z \in \mathbb{R}^n \ .$$

The quantity $e^{2H(X)/n}$ as a function of the random variable X is called the *entropy power* of X. The choice of name can be explained by noticing that a probability density function $f_Z : \mathbb{R}^n \to \mathbb{R}$, $f_Z(z) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\|z\|_2^2/(2\sigma^2)}$ which is the product distribution of n i.i.d. Gaussians with variance σ^2, has entropy $H(Z) = \frac{n}{2}\log 2\pi e\sigma^2$. The entropy power of Z is then (up to a prefactor) given by the variance σ^2, which is also referred to as the average energy or *power* of Z. The entropy power of a generic random variable X is thus equal to the power of the Gaussian random variable which has the same entropy as X. It is remarkable that the entropy power inequality (5.1) holds for all random variables which satisfy certain regularity assumptions (specifically, it is enough to assume that the random variables have finite variance [52]), but are otherwise arbitrary. It is a very powerful inequality in classical information theory: Shannon [10] has used it to calculate the capacity of the additive Gaussian noise channel and to bound the capacity of an arbitrary additive noise channel. Later works have applied it in a variety of different settings, for instance, in [52] the EPI was used to obtain bounds on the convergence rate in the central limit theorem, and applications to multi-terminal information theory can be found in [53].

There is a wide array of important information-theoretic inequalities which concern entropic quantities. We want to highlight a particular one and refer to [54] for an extensive review of other information-theoretic inequalities. This is the isoperimetric inequality for entropies.

Theorem 5.2.3 (Isoperimetric inequality for entropies [54]). *Let X be an $\mathbb{R}^n$-valued random variable with finite second moments and let*

$$J(X) = J(f_X) = \int_{\mathbb{R}^n} \nabla f_X(x)^T \nabla f_X(x) \frac{\mathrm{d}^n x}{f_X(x)} \tag{5.2}$$

be the Fisher information of X, where $\nabla f(x) = \left(\frac{\partial}{\partial x_1} f(x), \ldots, \frac{\partial}{\partial x_n} f(x)\right)^T$ is the gradient of f. Then we have

$$\frac{1}{n} J(X) e^{H(X)/n} \geq 1 \ . \tag{5.3}$$

The isoperimetric inequality for entropies is a direct consequence of the entropy power inequality via the de Bruijn identity [50, 51], which is also a crucial ingredient in the proof of the EPI:

Theorem 5.2.4 (De Bruijn's identity [23, 50, 52]). *Let X be an $\mathbb{R}^n$-valued random variable with finite second moments. Then we have, for any $\epsilon > 0$*

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} H(X + \sqrt{\epsilon} Z) = \frac{1}{2} J(X + \sqrt{\epsilon} Z) \ ,$$

where Z has the probability density function $f_Z(z) = (2\pi)^{-n/2} e^{-||z||_2^2/2}$. In particular, if the limits $\lim_{\epsilon' \downarrow 0} \big|_{\epsilon=\epsilon'} \frac{\mathrm{d}}{\mathrm{d}\epsilon} H(X + \sqrt{\epsilon} Z)$ and $\lim_{\epsilon \downarrow 0} J(X + \sqrt{\epsilon} Z)$ exist, we have

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\bigg|_{\epsilon=0} H(X + \sqrt{\epsilon} Z) = \frac{1}{2} J(X) \ . \tag{5.4}$$

The isoperimetric inequality for entropies is closely related to Gross's Log-Sobolev inequality [55], which has applications in quantum field theory.

There is a striking connection between the aforementioned entropic inequalities and geometric analysis: The EPI and the related isoperimetric inequality are formally similar to the Brunn-Minkowski inequality from geometric analysis [56]: For two compact subsets $A, B \subset \mathbb{R}^n$, we have

$$\mu(A + B)^{1/n} \geq \mu(A)^{1/n} + \mu(B)^{1/n} \ ,$$

where μ is the Lebesgue volume and $A+B$ is the Minkowski sum $A+B := \{a+b \mid a \in A, b \in B\}$. The Brunn-Minkowski inequality implies the isoperimetric inequality. The latter states that for a subset $A \subset \mathbb{R}^n$, we have

$$n\omega_n^{1/n} \mu(A)^{(n-1)/n} \leq \mathrm{area}(A) \ ,$$

where $\mathrm{area}(A) = \lim_{\epsilon \downarrow 0} \frac{\mu(A + B_\epsilon) - \mu(A)}{\epsilon}$ is the surface area of A if the limit exists [57–59], B_ϵ is an ϵ-ball in $\mathbb{R}^n$, and $\omega_n = \mu(B_1)$ is the volume of the unit ball $B_1 \subset \mathbb{R}^n$. Comparing with Eq. (5.3), the entropy power plays the role of volume and the Fisher information plays the role of inverse surface area. One can also view the isoperimetric inequality for entropies using this connection: Gaussian states have the smallest entropy power for a fixed Fisher information. In this sense, Gaussian states play the role of balls, which have maximal volume for a fixed area by the isoperimetric inequality. This connection between geometric analysis and information theory has been shown to be very fruitful. In fact, there is a proof of the entropy power inequality based on the Brunn-Minkowski inequality [60].

While Shannon did not provide a rigorous proof of the entropy power inequality (5.1), a series of subsequent work has put the inequality on a rigorous foundation [50–52]. We give a sketch of a proof of the classical entropy power inequality here because it is of central importance for the quantum setting. The idea for this proof is due to Blachman and Stam [50, 51] and makes extensive use of the so-called heat semigroup. An alternative proof which is based on the sharp Young inequality for convolutions and Rényi entropies has been given by Lieb in [61]. The first proof can be translated to the quantum setting, which is why it is more important for our discussion here.

Our proof sketch shows how to prove a *linear* version of the EPI, which reads

$$H(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda H(X) + (1-\lambda)H(Y) \ , \tag{5.5}$$

for any two independent $\mathbb{R}^n$-valued random variables X, Y and $0 \leq \lambda \leq 1$. The derivation of the EPI is similar, but slightly more involved. However, this linear version can be shown to imply the classical EPI, and vice versa [61, 62]. The simplified version of this proof sketch was presented in [63].

Proof sketch of Eq. (5.5). The first key ingredient in the proof is given by the *Fisher information inequality*, which states that

$$\lambda J(X) + (1-\lambda)J(Y) - J\left(\sqrt{\lambda}X + \sqrt{1-\lambda}Y\right) \geq 0 \ , \tag{5.6}$$

for any two independent $\mathbb{R}^n$-valued random variables X, Y and $0 \leq \lambda \leq 1$. The second key ingredient is the already stated De Bruijn identity (5.4). Now we consider the following map on random variables, which we call the classical *heat semigroup*, for $t \geq 0$:

$$\mathcal{N}_{\mathrm{cl}}(t)(X) := X_t := X + \sqrt{t}Z \ ,$$

where $X + \sqrt{t}Z$ has the probability density function as in Theorem 5.2.4. This map acts as a one-parameter semigroup on random variables, i.e., $\mathcal{N}_{\mathrm{cl}}(t_1 + t_2) = \mathcal{N}_{\mathrm{cl}}(t_1) \circ \mathcal{N}_{\mathrm{cl}}(t_2)$ for all $t_1, t_2 \geq 0$. For $f_X \in \mathcal{C}^2(\mathbb{R}^n)$, the family of probability density functions $f_t = f_{\mathcal{N}_{\mathrm{cl}}(t)(X)}$ satisfies the heat equation

$$\frac{\partial}{\partial t} f_t = \Delta f_t \ , \qquad \Delta = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j^2}$$

with initial condition $f_0(x) = f_X(x)$ for all $x \in \mathbb{R}^n$, that is, $f_t = e^{t\Delta}(f_0)$. The probability distribution functions f_t have the explicit form

$$f_t(x) = \frac{1}{(2\pi t)^{n/2}} \int_{\mathbb{R}^n} e^{-\|y\|_2^2/(2t)} f_0^{(y)}(x) \, \mathrm{d}^n y \qquad \text{for } \ x \in \mathbb{R}^n \ .$$

For consistency with later notation, we have introduced the translation of a function, denoted by $f^{(y)}(x) := f(x - y)$, for $y \in \mathbb{R}^n$. It is a crucial fact that evolution under the heat semigroup makes the function f_t approach a Gaussian for $t \to \infty$. Furthermore, the asymptotic scaling of its entropy is a function of t only, independently of the initial distribution f_X, i.e., $\lim_{t\to\infty} H(X_t) - g(t) = 0$ for some function g which does not depend on f_X.

Moreover, the heat semigroup is compatible with convolution: We have, for any $0 \leq \lambda \leq 1$ and any $t \geq 0$,

$$(\lambda X_t + (1-\lambda)Y_t) = (\lambda X + (1-\lambda)Y)_t \ .$$

As a last crucial ingredient, we note that translations of probability distribution functions are also compatible with convolution: Denote by $X^{(\theta)}$ the random variable which has probability distribution function $f_X^{(\theta)}$, for $\theta \in \mathbb{R}^n$. Then

$$\left(\sqrt{\lambda}X + \sqrt{1-\lambda}Y\right)^{(\theta)} = \sqrt{\lambda}X^{(\sqrt{\lambda}\theta)} + \sqrt{1-\lambda}Y^{(\sqrt{1-\lambda}\theta)} \ .$$

This means that adding random variables and then translating them is the same as adding appropriately translated versions of X and Y.

We can now move to the proof sketch of the EPI itself. Consider the given random variables X and Y and apply the heat semigroup to them, obtaining random variables X_t and Y_t. Define the quantity

$$\delta(t) := H(\sqrt{\lambda}X_t + \sqrt{1-\lambda}Y_t) - \lambda H(X_t) - (1-\lambda)H(Y_t) \ .$$

If we can prove $\delta(0) \geq 0$, we have proven the linear entropy power inequality. In the limit $t \to \infty$, we have

$$\lim_{t\to\infty} \delta(t) = 0 \ , \tag{5.7}$$

as the three entropy quantities have the same scaling as a function of t (and because of compatibility of convolution with the heat semigroup). By de Bruijn's identity, we can calculate the derivative of δ in terms of the Fisher information, and obtain

$$\dot{\delta}(0) = \frac{1}{2}\left(J(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) - \lambda J(X) - (1-\lambda)J(Y)\right) \leq 0 \ ,$$

because of the Fisher information inequality (5.6). By the semigroup property of the heat semigroup, this implies that

$$\dot{\delta}(t) \leq 0 \qquad \text{for all } t \geq 0 \ ,$$

hence the function δ is nonincreasing. Combining this with Eq. (5.7), we obtain $\delta(0) \geq 0$, concluding the proof sketch. □

An alternative proof of the entropy power inequality which also uses Fisher information but does not involve the scaling of the entropy for $t \to \infty$ can be found in [54].

5.3 Quantum channels as convolutions

As we have seen, the sum of random variables is modeled by convolution of their probability densities. One can ask if there is an analogous operation for quantum states, or for classical noise acting on a quantum system. Considering the action of the channels from Sections 4.3.1 and 4.3.2 on the level of Wigner functions, the channels described by beamsplitter and squeezing unitaries and the classical noise channel are convolutions. In particular, for $0 < \lambda < 1$ and $t = \frac{1}{2\pi}$, we have for $\xi \in \mathbb{R}^{2n}$

$$W_{\mathcal{E}_{\lambda,\sigma}(\rho)}(\xi) = \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} \frac{1}{\lambda^n(1-\lambda)^n} W_\rho\left(\frac{\eta}{\sqrt{\lambda}}\right) W_\sigma\left(\frac{\xi-\eta}{\sqrt{1-\lambda}}\right) \mathrm{d}^{2n}\eta \ ,$$

and

$$W_{\mathcal{F}_{\frac{1}{2\pi},f}(\rho)}(\xi) = \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} W_\rho(\eta) f\left(\xi - \eta\right) \mathrm{d}^{2n}\eta \ , \tag{5.8}$$

for quantum states ρ, σ and a classical probability density $f : \mathbb{R}^{2n} \to \mathbb{R}$. Therefore the channels $\mathcal{E}_{\lambda,\sigma}$ and $\mathcal{F}_{t,f}$ are candidates for quantum analogs to the sum of random variables. The next sections will deal with entropic inequalities for these "quantum-quantum" and "classical-quantum" convolution operations.

5.4 The quantum entropy power inequality

The concepts of Section 5.2 can be translated to the quantum information setting. The role of random variables is played by quantum states, the role of the differential entropy is played by the von Neumann entropy, and the role of addition is played by the beamsplitter/squeezing interaction U_λ which is defined in Eq. (4.9).

Theorem 5.4.1 (Quantum entropy power inequality [13, 63, 64])**.** *Let ρ_X, ρ_Y be two n-mode bosonic states with finite second moments. Then we have*

$$\exp \frac{S\left(\mathcal{E}_{\lambda,\rho_Y}(\rho_X)\right)}{n} \geq \lambda \exp \frac{S(\rho_X)}{n} + (1-\lambda) \exp \frac{S(\rho_Y)}{n} \ , \tag{5.9}$$

for $\lambda \in [0,1]$ and the channel $\mathcal{E}_{\lambda,\rho_Y}$ defined in Eq. (4.10)*.*

The quantum entropy power inequality was first proven in the case $\lambda = \frac{1}{2}$ by König and Smith [63]. Subsequent work by De Palma et al. [64] lifted this restriction on λ, and the sufficiency of finite second moments follows from the work carried out in [13].

The entropy power inequality stated above is not the only conceivable way to translate the entropy power inequality to the quantum setting: Instead of taking the formal definition of entropy power and replacing the differential entropy by the von Neumann entropy, there is another way to generalize the notion of entropy power to the quantum setting. This is by translating the notion that the entropy power of a random variable X is the power of the Gaussian random variable Z which has the same entropy as X: For a generic n-mode state ρ, one then considers the average energy per mode of a Gaussian thermal state which has the same entropy as ρ [65]. This quantity is called the *mean number of thermal photons* of ρ and is given by $g^{-1}(S(\rho)/n)$, where g is the function from Eq (4.6). This leads to the so-called *Entropy Photon-Number Inequality* (EPNI) [65, 66].

Conjecture 5.4.2 (Entropy Photon Number Inequality [65,66])**.** *Let ρ_X, ρ_Y be n-mode bosonic states with finite second moments. Then we have*

$$g^{-1}\left(\frac{S(\mathcal{E}_{\lambda,\rho_Y}(\rho_X))}{n}\right) \geq \lambda g^{-1}\left(\frac{S(\rho_X)}{n}\right) + (1-\lambda) g^{-1}\left(\frac{S(\rho_Y)}{n}\right) \ ,$$

for $\lambda \in [0,1]$.

The EPNI is an arguably more natural way to translate the entropy power inequality to the quantum setting than Eq. (5.9). For instance, if the EPNI holds true, it is saturated by thermal states, while the quantum EPI is not saturated by Gaussian states with proportional covariance matrices unless they have the same entropy [66]. However, despite considerable research efforts, the EPNI remains an unproven conjecture. The EPNI has only been shown to hold true in very few special cases: It is known to be true in the one-mode case when one of the input states is thermal [67, 68] and when the two inputs are Gaussian states [69].

5.5 The entropy power inequality for classical noise channels

Motivated by the study of the EPI in the classical and quantum setting, one can wonder whether there are more ways to formulate an entropy power inequality. In light of Eq. (5.8), we define a convolution operation between a classical probability density function $f : \mathbb{R}^{2n} \to \mathbb{R}$ and an n-mode quantum state ρ as in Eq. (4.11) via

$$(f, \rho) \mapsto f \star_t \rho := \mathcal{F}_{f,t}(\rho) = \int_{\mathbb{R}^{2n}} f(\xi) D(\sqrt{2\pi t}\xi) \rho D^\dagger(\sqrt{2\pi t}\xi) \, \mathrm{d}^{2n}\xi \ . \tag{5.10}$$

The parameter $t \geq 0$ is introduced as a "tuning parameter" purely for convenience. Sometimes we will omit it and simply write $f \star \rho$ in the case $t = 1$. The factor of $\sqrt{2\pi}$ in the argument of the Weyl displacement operators is also chosen purely for convenience[1]. This convolution operation (for $t = 1$) was introduced by Werner in his seminal paper on quantum harmonic analysis on phase space [6], which established a form of Young's inequality for this convolution. The EPI for this convolution operation reads

Theorem 5.5.1 (Quantum entropy power inequality for classical noise channels [1,3])**.** *Let ρ be an n-mode bosonic state and $f : \mathbb{R}^{2n} \to \mathbb{R}$ a probability density function, each with finite second moments. Then we have, for any $t \geq 0$,*

$$\exp \frac{S(f \star_t \rho)}{n} \geq \exp \frac{S(\rho)}{n} + t \exp \frac{H(f)}{n} \ . \tag{5.11}$$

A proof outline of this inequality is one of the main contributions of Core Article I [1]. The assumptions on ρ and f used here were proven sufficient in Core Article III [3]. It enables us to study the output entropy of classical noise channels, which add classical noise to a quantum system, in the case of general, possibly non-Gaussian, noise. This was a key step to establish the capacity bounds on additive classical noise channels which are presented in Chapter 6.

Similarly to the classical EPI, the study of the inequality (5.10) has produced numerous other interesting information-theoretic inequalities. One notable example is the isoperimetric inequality for quantum entropies, which was derived in Core Article I [1] and which can be stated as

$$\frac{1}{2n} J(\rho) e^{S(\rho)/n} \geq 2\pi e \ . \tag{5.12}$$

where $J(\rho)$ is the quantum Fisher information defined as the trace of the Fisher information matrix

$$\left(\frac{\partial^2}{\partial\theta_j \partial\theta_k}\bigg|_{\theta=0} D\left(\rho \middle\| \rho^{(\theta)}\right) \right)_{j,k=1}^{2n} \ , \tag{5.13}$$

where $D(\cdot||\cdot)$ is the relative entropy defined in Lemma 4.2.2 and we denote the translation of an n-mode state ρ by a parameter $\theta \in \mathbb{R}^{2n}$ by $\rho^{(\theta)} = D(\theta)\rho D(\theta)^\dagger$. The quantum Fisher information is connected to the entropy production under a quantum version of the heat semigroup in a way closely resembling the classical de Bruijn identity. This gives additional motivation for the study of the convolution operation (5.10): While an isoperimetric inequality for quantum entropies can easily be stated, it is not implied by the quantum entropy power inequality (5.9) in the way the classical isoperimetric inequality is implied by the classical EPI. This is because the covariance matrices of quantum states have to satisfy the Heisenberg uncertainty relation and there is no meaningful analog of taking the limit of vanishing variance. However, the

[1] We note that in Article III, we have omitted this factor of $\sqrt{2\pi}$ in the Weyl displacement operators and have rescaled the classical differential entropy instead.

inequality (5.11) can be applied to Gaussian distributions f_Z, and then we can take the limit of vanishing variance to obtain Eq. (5.12). For more information-theoretic inequalities in this spirit, we refer to Core Article I [1].

5.6 Conditional information-theoretic inequalities

Many applications of the classical EPI consider a setting with *side information* and use a different version of the inequality, which involves the *conditional entropy.* Let X, Y be $\mathbb{R}^n$-valued random variables. The *joint probability density* $f_{XY} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ of X and Y is defined via

$$\Pr(X \in A, Y \in B) = \int_A \int_B f_{XY}(x,y) \mathrm{d}^n y \mathrm{d}^n x \ ,$$

for measurable $A, B \subset \mathbb{R}^n$. The *conditional probability density* of X conditioned on Y taking the value $y \in \mathbb{R}^n$ is defined when $f_Y(y) \neq 0$ as

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \qquad \text{for } x \in \mathbb{R}^n \ .$$

The *conditional entropy* of X given Y is then defined as the expectation value over Y of the entropy of X given the value of Y,

$$H(X|Y) = \int_{\mathbb{R}^n} H(X|Y=y) f_Y(y) \mathrm{d}^n y \ , \tag{5.14}$$

where $H(X|Y=y)$ is the entropy of the conditional probability density $f_{X|Y}(\cdot|y)$ for fixed $y \in \mathbb{R}^n$.

Given three $\mathbb{R}^n$-valued random variables $X, Y,$ and M, we say that X and Y are *conditionally independent* given M if the joint probability density is of the form

$$f_{XYM}(x,y,m) = f_M(m) f_{X|M}(x|m) f_{Y|M}(y|m) \ , \tag{5.15}$$

for all $x, y, m \in \mathbb{R}^n$. This condition is equivalent to the condition that the *conditional mutual information* of X and Y given M vanishes [70], which is

$$I(X:Y|M) := H(X|M) + H(Y|M) - H(XY|M) = 0 \ . \tag{5.16}$$

Given a joint probability density f_{XYM} such that X and Y are conditionally independent given M, we can define the notion of the sum $Z = X + Y$ via

$$f_{Z|M=m}(z) = \int_{\mathbb{R}^n} f_{XY|M=m}(x, z-x) \mathrm{d}^n x, \qquad \text{for all } z \in \mathbb{R}^n \ .$$

The *conditional entropy power inequality* is then an immediate consequence of the EPI (5.1).

Corollary 5.6.1 (Conditional entropy power inequality)**.** *Let X, Y, M be $\mathbb{R}^n$-valued random variables with finite second moments such that X and Y are conditionally independent given M. We then have*

$$e^{2H(Z|M)/n} \geq e^{2H(X|M)/n} + e^{2H(Y|M)/n} \ .$$

Proof (following [13, Appendix A]). From the classical entropy power inequality (5.1), we have for any fixed $m \in \mathbb{R}^n$ that

$$H(Z|M=m) \geq \frac{n}{2} \log \left(\exp \frac{2H(A|M=m)}{n} + \exp \frac{2H(B|M=m)}{n} \right) . \tag{5.17}$$

Note that the function $q : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, defined by

$$(a,b) \mapsto \frac{n}{2} \log \left(\exp \frac{2a}{n} + \exp \frac{2b}{n} \right)$$

is convex. Hence we have from the definition of the conditional entropy and Eq. (5.17)

$$\begin{aligned} H(Z|M) &\geq \int_{\mathbb{R}^n} \frac{n}{2} \log \left(\exp \frac{2H(A|M=m)}{n} + \exp \frac{2H(B|M=m)}{n} \right) f_M(m) \mathrm{d}^n m \\ &\geq \frac{n}{2} \log \left(\exp \frac{2}{n} \int_{\mathbb{R}^n} H(A|M=m) f_M(m) \mathrm{d}^n m + \exp \frac{2}{n} \int_{\mathbb{R}^n} H(B|M=m) f_M(m) \mathrm{d}^n m \right) \\ &= \frac{n}{2} \log \left(\exp \frac{2H(A|M)}{n} + \exp \frac{2H(B|M)}{n} \right) , \end{aligned}$$

where we have used Jensen's inequality [71, 72] in the second step. □

In light of the conditional EPI following immediately from the EPI, it is natural to ask whether a conditional version of the quantum EPI (5.9) also holds. The *quantum conditional entropy* of a quantum system A given a quantum system M is defined as

$$S(A|M) := S(\rho_{AM}) - S(\rho_M) .$$

The quantum conditional mutual information is defined accordingly, by replacing the classical conditional entropy with the quantum conditional entropy $I(A:B|M)$ in Eq. (5.16). If the n-mode bosonic systems A, B, and M are in a state ρ_{ABM}, we say that A and B are conditionally independent given M if the quantum conditional mutual information $I(A:B|M)$ vanishes. Regarding Eq. (5.15), we may ask how such states look like. Indeed, if $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_M$ are finite-dimensional, if we decompose $\mathcal{H}_M$ into a direct sum of tensor products of the form

$$\mathcal{H}_M = \bigoplus_j \mathcal{H}_{m_j^L} \otimes \mathcal{H}_{m_j^R} , \tag{5.18}$$

then states of the form

$$\rho_{ABM} = \bigoplus_j q_j \rho_{Am_j^L} \otimes \rho_{Bm_j^R}$$

with states $\rho_{Am_j^L}$ on $\mathcal{H}_A \otimes \mathcal{H}_{m_j^L}$ and $\rho_{Bm_j^R}$ on $\mathcal{H}_B \otimes \mathcal{H}_{m_j^R}$ for a probability distribution $\{q_j\}$ are such that A and B are conditionally independent given M. In fact, any conditionally independent state is of this form for some decomposition (5.18) [70]. However, in the infinite-dimensional case, this condition is sufficient, but it is not known whether it is necessary for conditional independence.

The *quantum conditional entropy power inequality* can be proven for states ρ_{ABM} for which the n-mode bosonic systems A and B are conditionally independent given the quantum system M.

Theorem 5.6.2 (Conditional quantum entropy power inequality [13])**.** *Let A, B be n-mode bosonic systems and let M be a quantum system. Let $\rho_{ABM} \in \mathcal{S}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_M)$ be such that*

$$\mathrm{tr}_{AB}\left[\left(\sum_{k=1}^{n} a_{k,A}^{\dagger} a_{k,A} + a_{k,B}^{\dagger} a_{k,B}\right) \rho_{AB}\right] < \infty, \qquad S(\rho_M) < \infty \ ,$$

and let $I(A:B|M)_{\rho_{ABM}} = 0$. Then we have for any $\lambda \geq 0$

$$\exp \frac{S(C|M)_{\rho_{CM}}}{n} \geq \lambda \exp \frac{S(A|M)_{\rho_{AM}}}{n} + (1-\lambda) \exp \frac{S(B|M)_{\rho_{BM}}}{n} \ .$$

The state ρ_{CM} is obtained from ρ_{ABM} by

$$\rho_{CM} = (\mathcal{E}_\lambda \otimes \mathbb{1}_M)(\rho_{ABM}) \ ,$$

where $\mathcal{E}_\lambda$ is defined as

$$\mathcal{E}_\lambda(\rho_A \otimes \rho_B) := \mathcal{E}_{\lambda,\rho_B}(\rho_A) \ ,$$

and linearly extended.

Unlike in the classical setting, the quantum conditional EPI is not an immediate consequence of the quantum EPI because the quantum conditional entropy cannot be written as an expectation value in the spirit of Eq. (5.14). A linear version of the quantum conditional EPI was first proven for Gaussian states in [12], and the full inequality was proven in [13]. The conditional EPI has implications on the entanglement-assisted classical capacity of bosonic channels, as was first discussed in [12].

One can also formulate a conditional version of the EPI for classical noise channels (5.11). The proof of this inequality was the main contribution of Core Article III [3]. For details, we refer to the presentation of the results from that article. We have given an overview of entropic inequalities for bosonic systems and the state of research in this field, with focus on entropy power inequalities. We are going to discuss applications of these entropic inequalities in the next chapter.

6 The classical capacity of bosonic quantum channels

We have introduced the classical capacity of quantum channels in a general setting in Chapter 3. When treating continuous-variable systems, some subtleties arise. We want to consider one-mode bosonic channels and use a continuous alphabet $\mathcal{A} = \mathbb{R}^2$, which leads to ensembles $\{p(\xi)\mathrm{d}^2\xi, \rho_\xi\}_{\xi\in\mathbb{R}^2}$, where $\mathrm{d}^2\xi$ is the Lebesgue measure on $\mathbb{R}^2$ and $p : \mathbb{R}^2 \to \mathbb{R}$ is a probability density function. The classical capacity from Corollary 3.3.4 is not well-defined for a bosonic channel. To obtain a meaningful quantity, we need to introduce a constraint on the signal states $\{\rho_\xi\}_{\xi\in\mathbb{R}^2}$ and their distribution p, which is similar to the power constraint (3.3) commonly used in classical information theory. Such a constraint typically reads

$$\mathrm{tr}\left(a^\dagger a\overline{\rho}\right) \leq N \ ,$$

and physically means that Alice can, on average, only use a finite amount of energy N. Here the average signal state is given by

$$\overline{\rho} = \int_{\mathbb{R}^2} p(\xi)\rho_\xi \ \mathrm{d}^2\xi \ .$$

Theorem 6.0.1 (Energy-constrained classical capacity)**.** *The* energy-constrained classical capacity *of a bosonic quantum channel $\mathcal{E} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ is given by*

$$C_N(\mathcal{E}) = \lim_{n\to\infty} \frac{1}{n}\chi_{nN}\left(\mathcal{E}^{\otimes n}\right) \ ,$$

where $\chi_{nN}\left(\mathcal{E}^{\otimes n}\right)$ is the Holevo quantity with average energy constraint $N > 0$ per mode of the channel $\mathcal{E}^{\otimes n} : \mathcal{B}\left(\mathcal{H}^{\otimes n}\right) \to \mathcal{B}\left(\mathcal{H}^{\otimes n}\right)$. This Holevo quantity is defined as

$$\chi_{nN}\left(\mathcal{E}^{\otimes n}\right) = \sup_{\{p(\xi)\mathrm{d}^{2n}\xi,\rho_\xi\}_{\xi\in\mathbb{R}^{2n}}} S\left(\mathcal{E}^{\otimes n}(\overline{\rho})\right) - \int_{\mathbb{R}^{2n}} p(\xi)S\left(\mathcal{E}^{\otimes n}(\rho_\xi)\right) \ \mathrm{d}^{2n}\xi \ , \tag{6.1}$$

where the optimization is to be carried out over all ensembles $\{p(\xi)\mathrm{d}^{2n}\xi, \rho_\xi\}_{\xi\in\mathbb{R}^{2n}}$ of states on $\mathcal{H}^{\otimes n}$ with average signal state $\overline{\rho} = \int_{\mathbb{R}^{2n}} p(\xi)\rho_\xi \ \mathrm{d}^{2n}\xi$ which satisfy the average energy constraint $\mathrm{tr}\left(\sum_{j=1}^n a_j^\dagger a_j\overline{\rho}\right) \leq nN$.

As in the discussion in Chapter 3, a trivial lower bound[1] on the classical capacity of a quantum channel can be found by restricting to unentangled signal states, giving the so-called *one-shot capacity* or *product state capacity.* This capacity is equal to the Holevo quantity and we have

$$C_N(\mathcal{E}) \geq \chi_N(\mathcal{E}) \ .$$

[1] We note that it is not at all trivial to calculate $\chi_N(\mathcal{E})$ for a given channel $\mathcal{E}$, it is merely trivial to see that the quantity is a lower bound on the full capacity $C_N(\mathcal{E})$.

Naturally, the question whether the Holevo quantity is additive, i.e. whether $\chi_{nN}(\mathcal{E}^{\otimes n}) = n\chi_N(\mathcal{E})$ for all $n \in \mathbb{N}$, is a central question of quantum information processing. We have discussed that the Holevo quantity is in general not additive for finite-dimensional systems, but there is no counterexample for its additivity known in the bosonic setting.

The classical capacity of a special subclass of bosonic quantum channels, the so called single-mode phase-insensitive Gaussian channels, has been found by Giovannetti et al. in [8, 73] - in these cases, the classical capacity is additive, i.e. $C_{nN}\left(\mathcal{E}^{\otimes n}\right) = nC_N\left(\mathcal{E}\right)$ for all $n \in \mathbb{N}$. In this chapter, we give an overview of this remarkable result and how entropic inequalities in the spirit of the ones presented in Chapter 5 are related to capacity questions.

6.1 The classical capacity of one-mode phase-insensitive Gaussian channels

A phase-insensitive one-mode bosonic channel is a quantum channel $\Phi : \mathcal{S}(\mathcal{H}_{\text{osc}}) \to \mathcal{S}(\mathcal{H}_{\text{osc}})$ which is either phase-covariant or phase-contravariant under phase shift operations $e^{i\varphi a^\dagger a}$, that is, for all $\rho \in \mathcal{S}(\mathcal{H}_{\text{osc}})$ and any $\varphi \in \mathbb{R}$, we have

$$\Phi\left(e^{i\varphi a^\dagger a}\rho e^{-i\varphi a^\dagger a}\right) = \begin{cases} e^{i\varphi a^\dagger a}\Phi(\rho)e^{-i\varphi a^\dagger a} & (\text{phase} - \text{covariance}) \\ e^{-i\varphi a^\dagger a}\Phi(\rho)e^{i\varphi a^\dagger a} & (\text{phase} - \text{contravariance}) \ . \end{cases}$$

Many channels of practical importance are phase-insensitive. For instance, the beamsplitter channel $\mathcal{E}_{\lambda,\sigma}$ defined in Eq. (4.10) for Gaussian environment state σ is a phase-insensitive Gaussian channel. The classical noise channel $\mathcal{F}_{t,f}$ from Eq. (4.11) is also a phase-insensitive Gaussian channel if the probability density function f is Gaussian. Phase-insensitive Gaussian channels have been the subject of extensive research for decades [11, 38, 65, 67, 68, 74–87], and one of the landmark results in this context is the proof of additivity of the classical capacity for one-mode phase-insensitive bosonic Gaussian channels [8]. For the most important particular cases, this result is stated in the following theorem.

Theorem 6.1.1 (Classical capacity of fundamental phase-insensitive Gaussian channels [8])**.** *The classical capacity of the one-mode beamsplitter channel $\mathcal{E}_{\lambda,\sigma}$ with an environment state σ equal to a thermal state ρ_{th,N_E} is given by*

$$C_N(\mathcal{E}_{\lambda,\rho_{\text{th},N_E}}) = g\left(\lambda N + (1-\lambda)N_E\right) - g\left((1-\lambda)N_E\right) \ .$$

The classical capacity of the one-mode classical noise channel $\mathcal{F}_{t,f}$ with noise probability density f equal to a unit-variance centered Gaussian $f_Z(\xi) = (2\pi)^{-1}e^{-\|\xi\|^2/2}$ is given by

$$C_N\left(\mathcal{F}_{t,f_Z}\right) = g\left(N + 2\pi t\right) - g(2\pi t) \ .$$

These capacities are achieved with Gaussian encodings, in this case via the ensemble

$$\{p(\xi)\mathrm{d}^2\xi, \rho_\xi\}_{\xi\in\mathbb{R}^2} = \{(2\pi N)^{-1}e^{-\|\xi\|^2/(2N)}\mathrm{d}^2\xi, |\xi\rangle\langle\xi|\}_{\xi\in\mathbb{R}^2}$$

of coherent states with Gaussian probability density. The average signal state of such an ensemble is the thermal state $\rho_{\text{th},N}$. This corresponds to codewords which are tensor products of coherent states.

The proof of Theorem 6.1.1 works by proving the so-called *minimum output entropy conjecture* for these channels.

Theorem 6.1.2 (Minimum output entropy of fundamental phase-insensitive Gaussian channels [84, 88])**.** *For any $n \in \mathbb{N}$, the vacuum state $|0\rangle^{\otimes n} \in \mathcal{H}_{osc}^{\otimes n}$ minimizes the output entropy for the n-mode channels $\mathcal{E}_{\lambda,\rho_{\mathrm{th},N_E}}^{\otimes n}$ and $\mathcal{F}_{t,f_Z}^{\otimes n}$, i.e., for any n-mode quantum state ρ, $N_E \geq 0$, $t \geq 0$, and $0 \leq \lambda \leq 1$, we have*

$$S\left(\mathcal{E}_{\lambda,\rho_{\mathrm{th},N_E}}^{\otimes n}(\rho)\right) \geq S\left(\mathcal{E}_{\lambda,\rho_{\mathrm{th},N_E}}^{\otimes n}(|0\rangle\langle 0|^{\otimes n})\right) = nS\left(\mathcal{E}_{\lambda,\rho_{\mathrm{th},N_E}}(|0\rangle\langle 0|)\right) \ ,$$

$$S\left(\mathcal{F}_{t,f_Z}^{\otimes n}\right) \geq S\left(\mathcal{F}_{t,f_Z}^{\otimes n}(|0\rangle\langle 0|^{\otimes n})(\rho)\right) = nS\left(\mathcal{F}_{t,f_Z}(|0\rangle\langle 0|)\right) \ .$$

In general, lower bounds on the output entropy of quantum channels can be used to estimate the second term appearing in the definition of the Holevo quantity (6.1). This is commonly used to establish upper bounds on the classical capacity. The upper bounds on the capacity obtained from Theorem 6.1.2 in the case of the channels $\mathcal{E}_{\lambda,\rho_{\mathrm{th},N_E}}$ and $\mathcal{F}_{t,f_Z}$ are, in turn, achievable by the aforementioned Gaussian modulation of coherent states. For these channels, we then have a lower and an upper bound on the capacity which coincide, which settles the question of classical capacity.

This general procedure of obtaining lower bounds on the output entropy to establish upper bounds on the capacity had already been used before Theorem 6.1.2 was proven. If the minimum output entropy result is unavailable, one can use an Entropy Power Inequality to find lower bounds on the output entropy. This strategy has been applied to the channel $\mathcal{E}_{\lambda,\rho_{\mathrm{th},N_{\mathrm{E}}}}$ in [11, 79], and led to the best upper bounds known on the capacity of these channels before the landmark result from [8]. In these cases, due to the fact that quantum EPIs are typically not tight, there is still a gap between the lower bound on the capacity achievable by Gaussian modulation of coherent states and the upper bound. In the case of the channel $\mathcal{E}_{\lambda,\rho_{\mathrm{th},N_{\mathrm{E}}}}$, this gap is independent of the input energy and hence EPIs can be used to establish an absolute bound on the additivity violation of the capacity of this channel. This means that if an additivity violation exists at all, it is small, and Gaussian modulation of coherent states is a practically useful encoding strategy. Ultimately, Theorem 6.1.1 rules out an additivity violation for the channels considered in [11, 79] and supersedes the results therein.

6.2 Consequences of the EPI & EPNI on the classical capacity of non-Gaussian channels

When it comes to non-Gaussian channels, little is known about their classical capacity. Suppose $\Phi : \mathcal{S}(\mathcal{H}_{\mathrm{osc}}) \to \mathcal{S}(\mathcal{H}_{\mathrm{osc}})$ is a general one-mode bosonic channel, which is not necessarily Gaussian. Lower bounds can be obtained by employing a Gaussian encoding like in the previous section, while for upper bounds we need to upper bound the Holevo quantity of the n-mode channels $\Phi^{\otimes n}$, for any $n \in \mathbb{N}$. Such bounds can be established if we have both upper and lower bounds on the output entropy of the channels $\Phi^{\otimes n}$. This is a very difficult problem in general, but in the case of the non-Gaussian channels introduced in Section 4.3.3, which are Gaussian-dilatable, Entropy Power Inequalities come to the rescue. The work from Core Article II [2] employs the Entropy Power Inequalities from Core Article I [1] and [63], together with the maximum entropy principle 4.2.2 to establish upper and lower bounds on the classical capacity of the channels $\mathcal{E}_{\lambda,\sigma}$ and $\mathcal{F}_{t,f}$ for general, non-Gaussian σ and f. The gap between the upper and lower bound is bounded by a constant independent of the input energy. Hence, for these non-Gaussian channels, we have established a bound on the magnitude of additivity violations, if such additivity violations exist.

Theorem 6.2.1 (Capacity bounds for non-Gaussian channels [2, Lemmas 1A and 1B])**.** *The classical capacity of the single-mode beamsplitter channel $\mathcal{E}_{\lambda,\sigma_E}$ for a general environment state $\sigma_E \in \mathcal{S}(\mathcal{H}_{\mathrm{osc}})$ with finite first and second moments satisfies*

$$g\left(\lambda N + (1-\lambda)N_E^{ep}\right) \leq C_N\left(\mathcal{E}_{\lambda,\sigma_E}\right) \leq g\left(\lambda N + (1-\lambda)N_E\right) - \log\left(\lambda + (1-\lambda)e^{S(\sigma_E)}\right) \ ,$$

where $N_E = \mathrm{tr}(a^\dagger a \sigma_E)$ is the average energy of the environment state, and $N_E^{ep} = g^{-1}\left(S(\sigma_E)\right)$ is the average number of thermal photons in the environment. The lower bound is achievable with a coherent state ensemble. The difference between the upper and lower bound is bounded by $2g\left((1-\lambda)N_E\right) - g\left((1-\lambda)N_E^{ep}\right) - \log\left(\lambda + (1-\lambda)e^{S(\sigma_E)}\right)$, independently of the input energy constraint N. For the classical capacity of the single-mode classical noise channel $\mathcal{F}_{t,f}$ for a general probability density function $f : \mathbb{R}^2 \to \mathbb{R}$ with finite first and second moments satisfies

$$\log\left(e^{g(N)} + te^{H(f)}\right) - g\left(\pi t \mathrm{E}(f)\right) \leq C_N\left(\mathcal{F}_{t,f}\right) \leq g\left(N + \pi t E(f)\right) - \log\left(1 + te^{H(f)}\right) \ ,$$

where $E(f) = \sum_{i=1}^{2} \int_{\mathbb{R}^2} \xi_i^2 f(\xi) \mathrm{d}^2\xi$ is the energy of f. The lower bound is achievable with a coherent state ensemble, and the difference between this upper and lower bound is bounded by $2g\left(\pi t \mathrm{E}(f)\right) - \log\left(1 + te^{H(f)}\right)$, independently of the input energy constraint N.

To our knowledge, this is the first result on the capacity of non-Gaussian bosonic channels. We have given an overview of crucial results in the fields of entropic inequalities for bosonic channels and classical capacities of bosonic channels, and how some of the results from the Core Articles presented in this thesis contribute to the literature in these fields. Many questions remain, such as the question of validity of the Entropy Photon-Number Inequality [65] or the question of validity of the constrained minimum output entropy conjecture for multiple modes [69], both of which are intimately related to Entropy Power Inequalities. We conclude with the presentation of the contributed articles of this thesis.

Bibliography

[1] S. Huber, R. König, and A. Vershynina. Geometric inequalities from phase space translations. *Journal of Mathematical Physics*, 58(1):012206, 2017. doi:10.1063/1.4974224.

[2] S. Huber and R. König. Coherent state coding approaches the capacity of non-gaussian bosonic channels. *Journal of Physics A: Mathematical and Theoretical*, 51(18):184001, 2018. doi:10.1088/1751-8121/aab7ff.

[3] G. De Palma and S. Huber. The conditional entropy power inequality for quantum additive noise channels. *Journal of Mathematical Physics*, 59(12):122201, 2018. doi:10.1063/1.5027495.

[4] J. M. Renes, V. B. Scholz, and S. Huber. Uncertainty relations: An operational approach to the error-disturbance tradeoff. *Quantum*, 1:20, 2017. doi:10.22331/q-2017-07-25-20.

[5] S. Huber, R. König, and M. Tomamichel. Jointly constrained bilinear semidefinite programming with an application to Dobrushin curves. *arXiv preprint arXiv:1808.03182 [quant-ph]*, 2018.

[6] R. Werner. Quantum harmonic analysis on phase space. *Journal of Mathematical Physics*, 25(5):1404–1411, 1984. doi:10.1063/1.526310.

[7] E. A. Carlen and J. Maas. Gradient flow and entropy inequalities for quantum markov semigroups with detailed balance. *Journal of Functional Analysis*, 273(5):1810–1869, 2017. doi:10.1016/j.jfa.2017.05.003.

[8] V. Giovannetti, R. García-Patrón, N. J. Cerf, and A. S. Holevo. Ultimate classical communication rates of quantum optical channels. *Nature Photonics*, 8:796–800, 2014. doi:10.1038/nphoton.2014.216.

[9] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[10] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:623–656, 1948.

[11] R. König and G. Smith. Limits on classical communication from quantum entropy power inequalities. *Nature Photonics*, 7(2):142–146, 2013. doi:10.1038/nphoton.2012.342.

[12] R. König. The conditional entropy power inequality for Gaussian quantum states. *Journal of Mathematical Physics*, 56(2):022201, 2015. doi:10.1063/1.4906925.

[13] G. De Palma and D. Trevisan. The conditional entropy power inequality for bosonic quantum systems. *Communications in Mathematical Physics*, 360:639–662, 2018. doi:10.1007/s00220-017-3082-8.

[14] D. Kretschmann, D. Schlingemann, and R. F. Werner. A continuity theorem for stinespring's dilation. *Journal of Functional Analysis*, 255(8):1889 – 1904, 2008. `doi:10.1016/j.jfa.2008.07.023`.

[15] J. M. Renes and V. B. Scholz. Operationally-Motivated Uncertainty Relations for Joint Measurability and the Error-Disturbance Tradeoff. *arXiv preprint arXiv:1402.6711 [quant-ph]*, 2014.

[16] F. A. Al-Khayyal and J. E. Falk. Jointly constrained biconvex programming. *Mathematics of Operations Research*, 8(2):273–286, 1983. `doi:10.1287/moor.8.2.273`.

[17] E. B. Davies. *Quantum Theory of Open Systems.* Academic Press London, 1976.

[18] A. S. Holevo. *Quantum Systems, Channels, Information: A Mathematical Introduction.* De Gruyter, 2013.

[19] A. S. Holevo. *Statistical Structure of Quantum Theory.* Lecture Notes in Physics: N.s M, Monographs, 67. Springer, 2001.

[20] B. C. Hall. *Quantum Theory for Mathematicians*, volume 267 of *Graduate Texts in Mathematics.* Springer, New York, 2013. `doi:10.1007/978-1-4614-7116-5`.

[21] R. Renner. Ergänzendes Material zur Vorlesung Quantenmechanik I, 2011. lecture notes as used by R. Renner in the Quantum Mechanics I course at ETH Zurich in the fall of 2011.

[22] I. Gelfand and M. Neumark. On the imbedding of normed rings into the ring of operators in Hilbert space. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):197–217, 1943.

[23] T. M. Cover and J. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing).* Wiley-Interscience, 2006.

[24] D. Petz. *Channels and Their Capacity*, pages 91–107. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. `doi:10.1007/978-3-540-74636-2_7`.

[25] J. P. Gordon. Quantum Electronics and Coherent Light. In P. A. Miles, editor, *Proceedings of the International School of Physics "Enrico Fermi", Course XXXI*, pages 156–181. Academic, New York, 1964.

[26] L. B. Levitin. *Information, Complexity, and Control in Quantum Physics.* Springer, Vienna, 1987.

[27] A. S. Holevo. Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel. *Probl. Peredachi Inf.*, pages 177–183, 1973.

[28] J. Preskill. Quantum information, 2016. Lecture notes as used by J. Preskill in the Quantum Information class ph219 at CALTECH.

[29] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[30] A. S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, 1998. `doi:10.1109/18.651037`.

[31] B. Schumacher and M. D. Westmoreland. Sending classical information via noisy quantum channels. *Phys. Rev. A*, 56:131–138, 1997. `doi:10.1103/PhysRevA.56.131`.

[32] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-Assisted Classical Capacity of Noisy Quantum Channels. *Phys. Rev. Lett.*, 83:3081–3084, 1999. `doi:10.1103/PhysRevLett.83.3081`.

[33] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal. Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem. *IEEE Transactions on Information Theory*, 48(10):2637–2655, 2002. `doi:10.1109/TIT.2002.802612`.

[34] A. S. Holevo. On entanglement-assisted classical capacity. *Journal of Mathematical Physics*, 43(9):4326–4333, 2002. `doi:10.1063/1.1495877`.

[35] S. Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, 55:1613–1622, Mar 1997. URL: `http://link.aps.org/doi/10.1103/PhysRevA.55.1613`, `doi:10.1103/PhysRevA.55.1613`.

[36] B. Schumacher and M. A. Nielsen. Quantum data processing and error correction. *Phys. Rev. A*, 54:2629–2635, Oct 1996. URL: `https://link.aps.org/doi/10.1103/PhysRevA.54.2629`, `doi:10.1103/PhysRevA.54.2629`.

[37] M. B. Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5(4):255–257, 2009.

[38] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd. Gaussian quantum information. *Rev. Mod. Phys.*, 84:621–669, 2012. `doi:10.1103/RevModPhys.84.621`.

[39] S. L. Braunstein and P. van Loock. Quantum information with continuous variables. *Reviews of Modern Physics*, 77(2):513–577, jun 2005. `doi:10.1103/revmodphys.77.513`.

[40] C. Navarrete-Benlloch. An Introduction to the Formalism of Quantum Information. *arXiv preprint arXiv:1504.05270 [quant-ph]*, 2015.

[41] M. Reed and B. Simon. Viii - unbounded operators. In M. Reed and B. Simon, editors, *Methods of Modern Mathematical Physics*, pages 249 – 317. Academic Press, 1972. `doi:10.1016/B978-0-12-585001-8.50014-3`.

[42] O. Krüger. *Quantum Information Theory with Gaussian Systems*. PhD thesis, Technische Unversität Carolo-Wilhelmina zu Braunschweig, 2006.

[43] A. S. Holevo. Quantum Characteristic Functions. *Probl. Peredachi Inf.*, 6:42–48, 1970.

[44] J. Eisert and M. M. Wolf. *Gaussian quantum channels*, pages 23–42. Quantum Information with Continuous Variables of Atoms and Light. Imperial College Press, London, 2007.

[45] J. Williamson. On the algebraic problem concerning the normal forms of linear dynamical systems. *American Journal of Mathematics*, 58(1):pp. 141–163, 1936.

[46] M. M. Wolf, G. Giedke, and J. I. Cirac. Extremality of Gaussian quantum states. *Phys. Rev. Lett.*, 96:080502, 2006. `doi:10.1103/PhysRevLett.96.080502`.

[47] A. S. Holevo. One-mode quantum gaussian channels: Structure and quantum capacity. *Problems of Information Transmission*, 43(1):1–11, 2007. doi:10.1134/S0032946007010012.

[48] K. K. Sabapathy and A. Winter. Non-gaussian operations on bosonic modes of light: Photon-added gaussian channels. *Phys. Rev. A*, 95:062309, 2017. doi:10.1103/PhysRevA.95.062309.

[49] M. M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2017.

[50] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101 – 112, 1959. doi:10.1016/S0019-9958(59)90348-1.

[51] N. Blachman. The convolution inequality for entropy powers. *Information Theory, IEEE Transactions on*, 11(2):267 – 271, 1965. doi:10.1109/TIT.1965.1053768.

[52] A. R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14(1):336–342, 1986.

[53] A. E. Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2012.

[54] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *Information Theory, IEEE Transactions on*, 37(6):1501 –1518, 1991. doi:10.1109/18.104312.

[55] L. Gross. Logarithmic Sobolev Inequalities. *American Journal of Mathematics*, 97:1061–1083, 1975.

[56] R. Schneider. *Convex Bodies: The Brunn–Minkowski Theory*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2 edition, 2013. doi:10.1017/CBO9781139003858.

[57] M. Kneser. Über den Rand von Parallelkörpern. *Math. Nachr.*, 5(3-5):241–251, 1951. doi:10.1002/mana.19510050309.

[58] L. L. Stachó. On the volume function of parallel sets. *Acta Sci. Math.*, 38:365–374, 1976.

[59] J. Rataj and S. Winter. On volume and surface area of parallel sets. *Indiana University Mathematics Journal*, 59(5):1661–1686, 2010. doi:10.1512/iumj.2010.59.4165.

[60] S. J. Szarek and D. Voiculescu. *Shannon's entropy power inequality via restricted Minkowski sums*, volume 1745 of *Lecture Notes in Math.*, pages 257–262. Springer Berlin Heidelberg, 2000. URL: http://dx.doi.org/10.1007/BFb0107219, doi:10.1007/BF02108815.

[61] E. H. Lieb. Proof of an entropy conjecture of Wehrl. *Comm. Math. Phys.*, 62(1):35–41, 1978.

[62] S. Verdu and D. Guo. A simple proof of the entropy-power inequality. *Information Theory, IEEE Transactions on*, 52(5):2165–2166, 2006. doi:10.1109/TIT.2006.872978.

[63] R. König and G. Smith. The entropy power inequality for quantum systems. *Information Theory, IEEE Transactions on*, 60(3):1536–1548, 2014. doi:10.1109/TIT.2014.2298436.

[64] G. D. Palma, A. Mari, S. Lloyd, and V. Giovannetti. Multimode quantum entropy power inequality. *Phys. Rev. A*, 91:032320, 2015. doi:10.1103/PhysRevA.91.032320.

[65] S. Guha, B. Erkmen, and J. Shapiro. The entropy photon-number inequality and its consequences. In *Information Theory and Applications Workshop, 2008*, pages 128–130, 2008. doi:10.1109/ITA.2008.4601037.

[66] G. D. Palma, A. Mari, and V. Giovannetti. A generalization of the entropy power inequality to bosonic quantum systems. *Nature Photonics*, 8:958–964, 2014. doi:10.1038/nphoton.2014.252.

[67] G. D. 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, 2017. doi:10.1109/tit.2016.2621748.

[68] G. D. 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. doi:10.1103/physrevlett.118.160503.

[69] G. D. Palma. *Gaussian optimizers and other topics in quantum information.* PhD thesis, Scuola Normale Superiore, Pisa (Italy), 2016. Supervisor: Prof. Vittorio Giovannetti; arXiv:1710.09395.

[70] P. Hayden, R. Jozsa, D. Petz, and A. Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics*, 246(2):359–374, 2004. doi:10.1007/s00220-004-1049-z.

[71] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Math.*, 30:175–193, 1906. doi:10.1007/BF02418571.

[72] M. D. Perlman. Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *Journal of Multivariate Analysis*, 4(1):52 – 65, 1974. doi:10.1016/0047-259X(74)90005-0.

[73] V. Giovannetti, A. S. Holevo, and R. García-Patrón. A Solution of Gaussian Optimizer Conjecture for Quantum Channels. *Communications in Mathematical Physics*, 334(3):1553–1571, 2014. doi:10.1007/s00220-014-2150-6.

[74] A. S. Holevo and R. F. Werner. Evaluating capacities of bosonic Gaussian channels. *Phys. Rev. A*, 63:032312, 2001. doi:10.1103/PhysRevA.63.032312.

[75] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro. Minimum output entropy of bosonic channels: A conjecture. *Phys. Rev. A*, 70:032315, 2004. doi:10.1103/PhysRevA.70.032315.

[76] V. Giovannetti, A. S. Holevo, S. Lloyd, and L. Maccone. Generalized minimal output entropy conjecture for one-mode Gaussian channels: definitions and some exact results. *J. Phys. A*, 43(032315), 2010. doi:10.1088/1751-8113/43/41/415305.

[77] R. García-Patrón, C. Navarrete-Benlloch, S. Lloyd, J. H. Shapiro, and N. J. Cerf. Majorization theory approach to the gaussian channel minimum entropy conjecture. *Physical Review Letters*, 108(11), 2012. doi:10.1103/physrevlett.108.110505.

[78] R. García-Patrón, C. Navarrete-Benlloch, S. Lloyd, J. H. Shapiro, and N. J. Cerf. The holy grail of quantum optical communication. *AIP Conference Proceedings*, 1633(1):109–112, 2014. doi:10.1063/1.4903108.

[79] R. König and G. Smith. Classical capacity of quantum thermal noise channels to within 1.45 bits. *Phys. Rev. Lett.*, 110:040501, 2013. doi:10.1103/PhysRevLett.110.040501.

[80] V. Giovannetti, S. Lloyd, L. Maccone, and J. H. Shapiro. Electromagnetic channel capacity for practical purposes. *Nature Photonics*, 7(10):834–838, 2013. doi:10.1038/nphoton.2013.193.

[81] V. Giovannetti and S. Lloyd. Additivity properties of a gaussian channel. *Physical Review A*, 69(6), 2004. doi:10.1103/physreva.69.062307.

[82] V. Giovannetti, A. Holevo, and R. García-Patron. A Solution of Gaussian Optimizer Conjecture for Quantum Channels. *Communications in Mathematical Physics*, 334(3):1553–1571, 2015. doi:10.1007/s00220-014-2150-6.

[83] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen. Classical Capacity of the Lossy Bosonic Channel: The Exact Solution. *Phys. Rev. Lett.*, 92:027902, 2004. doi:10.1103/PhysRevLett.92.027902.

[84] S. Lloyd, V. Giovannetti, L. Maccone, N. J. Cerf, S. Guha, R. Garcia-Patron, S. Mitter, S. Pirandola, M. B. Ruskai, J. H. Shapiro, and H. Yuan. The bosonic minimum output entropy conjecture and Lagrangian minimization, 2009. arXiv:0906.2758.

[85] S. Guha, J. H. Shapiro, and B. I. Erkmen. Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture. *Phys. Rev. A*, 76:032303, 2007. doi:10.1103/PhysRevA.76.032303.

[86] G. D. Palma, D. Trevisan, and V. Giovannetti. One-mode quantum-limited Gaussian channels have Gaussian maximizers. *arXiv preprint arXiv:1610.09967 [quant-ph]*, 2016.

[87] G. D. Palma, D. Trevisan, and V. Giovannetti. Passive states optimize the output of bosonic gaussian quantum channels. *IEEE Transactions on Information Theory*, 62(5):2895–2906, 2016. doi:10.1109/tit.2016.2547426.

[88] G. De Palma, D. Trevisan, V. Giovannetti, and L. Ambrosio. Gaussian optimizers for entropic inequalities in quantum information. *Journal of Mathematical Physics*, 59(8):081101, 2018. doi:10.1063/1.5038665.

[89] F. A. Al-Khayyal. Generalized bilinear programming: Part I. Models, applications and linear programming relaxation. *European Journal of Operational Research*, 60(3):306–314, 1992. doi:10.1016/0377-2217(92)90082-K.

[90] H. Konno. A cutting plane algorithm for solving bilinear programs. *Mathematical Programming*, 11(1):14–27, 1976. doi:10.1007/BF01580367.

[91] H. D. Sherali and C. M. Shetty. A finitely convergent algorithm for bilinear programming problems using polar cuts and disjunctive face cuts. *Mathematical Programming*, 19(1):14–31, 1980. doi:10.1007/BF01581626.

[92] J. E. Falk. A linear max—min problem. *Mathematical Programming*, 5(1):169–188, 1973. doi:10.1007/BF01580119.

[93] A. V. Cabot and R. L. Francis. Solving certain nonconvex quadratic minimization problems by ranking the extreme points. *Operations Research*, 18(1):82–86, 1970. doi:10.1287/opre.18.1.82.

[94] H. Vaish and C. M. Shetty. The bilinear programming problem. *Naval Research Logistics Quarterly*, 23(2):303–309, 1976. doi:10.1002/nav.3800230212.

[95] Y. Polyanskiy and Y. Wu. Dissipation of information in channels with input constraints. *IEEE Transactions on Information Theory*, 62(1):35–55, 2016. doi:10.1109/TIT.2015.2482978.

A Core Articles

A.1 Geometric inequalities from phase space translations

Geometric inequalities from phase space translations

Stefan Huber, Robert König, and Anna Vershynina

We establish a quantum version of the classical isoperimetric inequality relating the Fisher information and the entropy power of a quantum state. The key tool is a Fisher information inequality for a state which results from a certain convolution operation: the latter maps a classical probability distribution on phase space and a quantum state to a quantum state. We show that this inequality also gives rise to several related inequalities whose counterparts are well-known in the classical setting: in particular, it implies an entropy power inequality for the mentioned convolution operation as well as the isoperimetric inequality, and establishes concavity of the entropy power along trajectories of the quantum heat diffusion semigroup. As an application, we derive a Log-Sobolev inequality for the quantum Ornstein-Uhlenbeck semigroup, and argue that it implies fast convergence towards the fixed point for a large class of initial states.

A.1.1 Main Results

In the following, denote by $f, g : \mathbb{R}^{2n} \to \mathbb{R}$ probability density functions with finite second moments and by ρ, σ quantum states of an n-mode bosonic system with finite energies $\operatorname{tr}\left(\sum_{k=1}^n a_k^\dagger a_k \rho\right), \operatorname{tr}\left(\sum_{k=1}^n a_k^\dagger a_k \sigma\right) < \infty$.

Lemma A.1.1 (Data processing inequality for convolution). *Let $f, g : \mathbb{R}^{2n} \to \mathbb{R}$ be probability density functions with full support. Then*

$$D(f \star_t \rho \| g \star_t \sigma) \leq D(f\|g) + D(\rho\|\sigma) \ .$$

Theorem A.1.2 (Quantum Stam inequality). *Let $\omega_q, \omega_c \in \mathbb{R}$, and $t \geq 0$. Then the Fisher information from Eqs.* (5.13) *and* (5.2) *satisfies*

$$\omega^2 J(f \star_t \rho) \leq \omega_q^2 J(\rho) + \omega_c^2 J(f) \ ,$$

where $\omega = \omega_q + \sqrt{t}\omega_c$. In particular,

$$J(f \star_t \rho)^{-1} - J(\rho)^{-1} - tJ(f)^{-1} \geq 0 \ .$$

Let $f_Z : \mathbb{R}^{2n} \to \mathbb{R}$ be the probability density of a centered Gaussian with variance 1, i.e., $f_Z(\xi) = (2\pi)^{-n} e^{-\|\xi\|^2/2}$.

Lemma A.1.3 (Quantum Fisher information isoperimetric inequality). *The following inequality holds, where $t \geq 0$:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \left[\frac{1}{2n} J(f_Z \star_t \rho)\right]^{-1} \geq 1 \ .$$

Let $N(\rho) := e^{S(\rho)/n}$ be the entropy power of an n-mode state ρ and let $N_{\mathrm{cl}}(f) := e^{H(f)/n}$ be the entropy power of a probability density function $f : \mathbb{R}^{2n} \to \mathbb{R}$. A *trajectory of the diffusion semigroup* is given by the map $t \mapsto f_Z \star_t \rho$ for a fixed initial state ρ.

Theorem A.1.4 (Concavity of the quantum entropy power)**.** *The entropy power along trajectories of the diffusion semigroup $t \mapsto f_Z \star_t \rho$ is concave, i.e.,*

$$\left.\frac{\mathrm{d}^2}{\mathrm{d}t^2}\right|_{t=0} N\left(f_Z \star_t \rho\right) \leq 0 \ .$$

Theorem A.1.5 (Entropy Power Inequality)**.** *For $t \geq 0$, the following inequality holds:*

$$N(f \star_t \rho) \geq N(\rho) + tN_{\mathrm{cl}}(f) \ .$$

In particular, choosing $f = f_Z$ as the distribution of a unit-variance centered Gaussian, we have

$$N\left(f_Z \star_t \rho\right) \geq N(\rho) + t2\pi e \ .$$

Theorem A.1.6 (Isoperimetric inequality for entropies)**.** *We have*

$$\frac{1}{n} J(\rho) N(\rho) \geq 4\pi e \ .$$

A.1.2 Application: The quantum Ornstein-Uhlenbeck semigroup

Consider a one-mode bosonic quantum system A and the quantum Ornstein-Uhlenbeck semigroup $\{\mathcal{P}^{(\mu,\lambda)}(t) = e^{t\mathcal{L}_{\mu,\lambda}}\}_{t\geq 0}$ defined by the Liouvillian generator.

$$\mathcal{L}_{\mu,\lambda} = \mu^2 \mathcal{L}_- + \lambda^2 \mathcal{L}_+ \qquad \text{for } \mu > \lambda > 0 \ , \tag{A.1}$$

where

$$\mathcal{L}_+(\rho) = a^\dagger \rho a - \frac{1}{2}\{aa^\dagger, \rho\} \qquad \text{and} \qquad \mathcal{L}_-(\rho) = a\rho a^\dagger - \frac{1}{2}\{a^\dagger a, \rho\} \ .$$

This semigroup has the thermal state $\sigma_{\mu,\lambda} = \rho_{\mathrm{th},\frac{\lambda^2}{\mu^2-\lambda^2}}$ with mean photon number $\frac{\lambda^2}{\mu^2-\lambda^2}$ as fixed point. We conjecture that in relative entropy, the quantum Ornstein-Uhlenbeck semigroup converges exponentially fast to its fixed point, for any input state.

Conjecture A.1.7 (Fast convergence of the quantum Ornstein-Uhlenbeck semigroup in relative entropy)**.** *We have for any one-mode quantum state ρ and for all $t \geq 0$*

$$D\left(\mathcal{P}^{(\mu,\lambda)}(t)(\rho) \middle\| \sigma_{\mu,\lambda}\right) \leq e^{-(\mu^2-\lambda^2)t} D\left(\rho \middle\| \sigma_{\mu,\lambda}\right) \ .$$

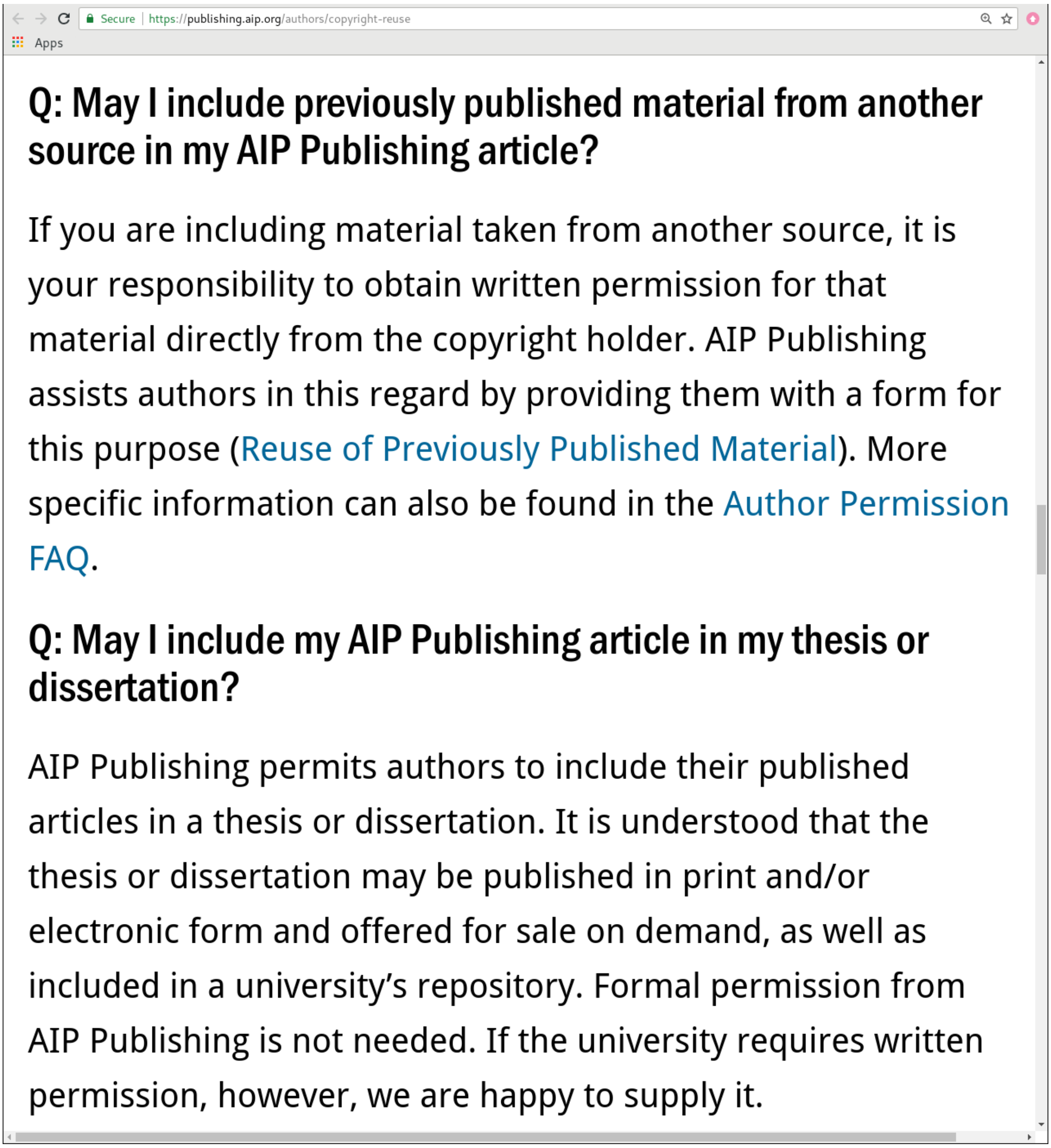
Using the isoperimetric inequality for quantum entropies, we give evidence for this conjecture by establishing a bound on the entropy production rate of the quantum Ornstein-Uhlenbeck semigroup for states with entropy larger than a certain threshold value, and for states with mean photon number smaller than a certain threshold value. Furthermore, we show that the statement holds for Gaussian states and that the exponent $\mu^2 - \lambda^2$ is optimal for Gaussian states. This conjecture has been subsequently proven in [7] using gradient flow methods, and in Core Article III using the entropy power inequality.

A.1.3 Individual Contribution

I was significantly involved in finding the ideas and carrying out the scientific work of all parts of this article, and I was in charge of writing the article, with the exception of Section V A and Lemma 8.

Permission to include:

Stefan Huber, Robert König, and Anna Vershynina
Geometric inequalities from phase space translations.
J. Math. Phys. 58, 012206 (2017).



Q: May I include previously published material from another source in my AIP Publishing article?

If you are including material taken from another source, it is your responsibility to obtain written permission for that material directly from the copyright holder. AIP Publishing assists authors in this regard by providing them with a form for this purpose (Reuse of Previously Published Material). More specific information can also be found in the Author Permission FAQ.

Q: May I include my AIP Publishing article in my thesis or dissertation?

AIP Publishing permits authors to include their published articles in a thesis or dissertation. It is understood that the thesis or dissertation may be published in print and/or electronic form and offered for sale on demand, as well as included in a university's repository. Formal permission from AIP Publishing is not needed. If the university requires written permission, however, we are happy to supply it.

JOURNAL OF MATHEMATICAL PHYSICS **58**, 012206 (2017)

Geometric inequalities from phase space translations

Stefan Huber,[1,2] Robert König,[1,2] and Anna Vershynina[2,3]
[1]*Institute for Advanced Study Technische Universität München, 85748 Garching, Germany*
[2]*Zentrum Mathematik Technische Universität München, 85748 Garching, Germany*
[3]*BCAM-Basque Center for Applied Mathematics, 48009 Bilbao, Spain*

(Received 22 July 2016; accepted 4 January 2017; published online 31 January 2017)

We establish a quantum version of the classical isoperimetric inequality relating the Fisher information and the entropy power of a quantum state. The key tool is a Fisher information inequality for a state which results from a certain convolution operation: the latter maps a classical probability distribution on phase space and a quantum state to a quantum state. We show that this inequality also gives rise to several related inequalities whose counterparts are well-known in the classical setting: in particular, it implies an entropy power inequality for the mentioned convolution operation as well as the isoperimetric inequality and establishes concavity of the entropy power along trajectories of the quantum heat diffusion semigroup. As an application, we derive a Log-Sobolev inequality for the quantum Ornstein-Uhlenbeck semigroup and argue that it implies fast convergence towards the fixed point for a large class of initial states. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4974224]

I. INTRODUCTION

The convolution operation $(X, Y) \mapsto X + Y$ between two real- (respectively vector-) valued independent random variables X and Y plays a central role in classical information theory. The operation is defined in terms of the action on probability density functions as

$$(f_X, f_Y) \mapsto f_{X+Y}, \qquad \text{where} \qquad f_{X+Y}(z) := \int f_X(z-x) f_Y(x) dx. \tag{1}$$

The convolution models a general class of additive noise channels, and thus provides a natural framework for the study of information processing and associated capacities. The operation (1) also is a central element in many functional analytic inequalities, most notably Young's inequality,[1] the Brascamp-Lieb inequalities,[2] de Bruijn's identity,[3] the Fisher information inequality,[3] and the entropy power inequality.[4,5] Such inequalities have wide use in information theory, yielding bounds on communication capacities, as observed by Shannon.[4] They can also provide, for example, bounds on the convergence rate in the central limit theorem.[6] Beyond these applications, these results are appealing from a conceptual, geometric viewpoint: many inequalities can be regarded as information-theoretic counterparts of related statements about convex bodies. For example, identifying entropy power with volumes reveals a formal similarity between the Brunn-Minkowski inequality and the entropy power inequality.[7] Indeed, there is even a proof of the latter guided by this intuition.[8] We refer to Refs. 9 and 10 for detailed accounts of this wealth of inequalities and their interrelationships.

Our work is guided by the question of whether a similar array of inequalities exists in a quantum setting. A key first step in this direction—one which is directly relevant to our work—was taken by Werner.[11] He introduced the convolution operation

$$(f, \rho) \mapsto f \star_t \rho, \qquad \text{where} \qquad f \star_t \rho = \int f(\xi) W(\sqrt{t}\xi) \rho W(\sqrt{t}\xi)^\dagger d\xi, \tag{2}$$

which involves a probability density function f on phase space as well as a state ρ (of a bosonic system). The operation results in the average state $f \star_t \rho$ when displacing ρ according to f using the (Weyl) displacement operators $W(\xi)$. Here we introduce the parameter $t \geq 0$ for convenience, the case $t = 1$ was considered in Ref. 11. Treating the convolutions (1) and (2) on the same algebraic footing, Werner established (among other results characterizing (2)) a form of Young's inequality. It is worth mentioning that Carlen and Lieb have recently established generalizations of the latter in a fermionic context.[12]

More recent work[13] has centered around a convolution operation of the form

$$(\rho_X, \rho_Y) \mapsto \rho_{X \boxplus_\lambda Y} = \mathrm{tr}_2(U_\lambda(\rho_X \otimes \rho_Y)U_\lambda^\dagger), \tag{3}$$

where U_λ is a (d-mode) beamsplitter of transmissivity $\lambda \in [0, 1]$. It is worth pointing out that the action of this operation is formally similar to (1) when expressed in terms of the Wigner functions describing the quantum states ρ_X and ρ_Y. The map (3) describes a process where two states interact. It captures, in particular, the situation where one of the states is transmitted through an additive (bosonic) noise channel. In Ref. 13, the authors established an entropy power inequality of the form

$$e^{S(\rho_X \boxplus_\lambda \rho_Y)/d} \geq \lambda e^{S(\rho_X)/d} + (1 - \lambda)e^{S(\rho_Y)/d}, \tag{4}$$

for the convolution (3) and for $\lambda = 1/2$, where $S(\rho_X) = -\mathrm{tr}(\rho_X \log \rho_X)$ denotes the von Neumann entropy. Subsequent work[14] managed to lift the restriction on λ, and generalized this result to more general (Gaussian) unitaries in place of U_λ. A related inequality of the form $S(\rho_X \boxplus_\lambda \rho_Y) \geq \lambda S(\rho_X) + (1 - \lambda)S(\rho_Y)$ for $\lambda \in [0, 1]$ was also shown in Ref. 13, generalizing classical results[15] (see also Ref. 16 for a discussion of the relationship between the two). A generalization to conditional entropies was proposed in Ref. 17, and an application to channel capacities was discussed in Ref. 18.

A key tool in establishing these results is the quantum Fisher information $J(\rho)$, defined for a state ρ as the divergence-based Fisher information of the family $\{\rho^{(\theta)}\}_\theta$ obtained by displacing ρ along a phase space direction (see Section III for a precise definition). It was shown in Ref. 13 for $\lambda = 1/2$ and in Ref. 14 for general $\lambda \in [0, 1]$ that this quantity satisfies the Fisher information inequality[3]

$$J(\rho_X \boxplus_\lambda \rho_Y)^{-1} \geq \lambda J(\rho_Y)^{-1} + (1 - \lambda)J(\rho_X)^{-1}.$$

This identity is a consequence of the strong subadditivity (data processing) inequality for relative entropy and lies at the heart of the proof of (4).

Following the theme of entropy power inequalities for quantum systems, Audenaert, Datta, and Ozols[19] have obtained strong majorization-type results for the finite-dimensional case. These center around an operation of the form (3), but with ρ_X and ρ_Y being the states on a finite-dimensional Hilbert space, and a family $U_\lambda = e^{i\lambda H}$ of unitaries generated by a Hamiltonian H realizing the SWAP-operation of the two systems. As argued by Audenaert *et al.*, these results imply several entropy-power-type inequalities. Very recently, Carlen, Lieb, and Loss[20] have provided an elegant short proof of these statements. Guha, Shapiro, and García-Patrón have discussed alternative definitions of the quantum entropy power and corresponding entropy power inequalities.[21]

II. OUR CONTRIBUTION

A. New geometric inequalities for bosonic systems

Here we focus on the convolution operation (2). We find that this operation satisfies similar properties as the convolutions (1) and (3). In particular, we establish a Stam inequality for the Fisher information of the form

$$J(f \star_t \rho)^{-1} \geq J(\rho)^{-1} + tJ(f)^{-1}, \qquad \text{for all } t \geq 0. \tag{5}$$

This classical-quantum version of Stam's inequality has several immediate consequences, all of which follow the reasoning used in establishing classical results.[9] For example, we use (5) to establish an entropy power inequality of the form

$$\exp\left(S(f \star_t \rho)/d\right) \geq \exp\left(S(\rho)/d\right) + t \exp\left(H(f)/d\right). \tag{6}$$

Taking $f = f_Z$ to be a unit-variance centered Gaussian, we find a quantum isoperimetric inequality of the form

$$\frac{d}{dt}\Big|_{t=0}\left(\frac{1}{2d}J(f_Z \star_t \rho)\right)^{-1} \geq 1.$$

Note that $f_Z \star_t \rho$ is the result of applying a classical noise channel to ρ, where the variance of the displacement goes to 0 in the limit $t \to 0$. This family of maps constitutes a semigroup, which we call the heat diffusion semigroup: it is generated by a Liouvillian $\mathcal{L}_{\mathsf{heat}}$ such that

$$f_Z \star_t \rho = e^{t\mathcal{L}_{\mathsf{heat}}}(\rho), \qquad \text{for an initial state } \rho \text{ and } t \geq 0.$$

We find that the entropy power along trajectories generated by this Liouvillian is concave, i.e.,

$$\frac{d^2}{dt^2}\Big|_{t=0}\exp\left(\frac{1}{d}S(e^{t\mathcal{L}_{\mathsf{heat}}}(\rho))\right) \leq 0. \tag{7}$$

Eq. (7) generalizes a celebrated result[22,23] concerning the classical heat equation. The entropy power inequality (6) for Gaussian f implies the lower bound $\exp\left(\frac{1}{d}S(e^{t\mathcal{L}_{\mathsf{heat}}}(\rho))\right) \geq \exp\left(\frac{1}{d}S(\rho)\right) + (2\pi e)t$ and establishes the isoperimetric inequality for the Fisher information

$$J(\rho)\exp\left(\frac{1}{d}S(\rho)\right) \geq 4\pi e\, d, \tag{8}$$

for states ρ of d bosonic modes. We find that for $d = 1$, this statement is tight: Gaussian thermal states achieve equality in (8) in the limit of large mean photon numbers.

B. Application to the Ornstein-Uhlenbeck semigroup

We apply our results, in particular Eq. (8), to the quantum Ornstein-Uhlenbeck (qOU) semigroup for a one-mode bosonic system. This is a one-parameter group of CPTP maps $\{e^{t\mathcal{L}_{\mu,\lambda}}\}_{t\geq 0}$ generated by a linear combination of Liouvillians of a quantum amplifier and an attenuator channel, respectively,

$$\mathcal{L}_{\mu,\lambda} = \mu^2\mathcal{L}_- + \lambda^2\mathcal{L}_+, \qquad \text{for } 0 < \lambda < \mu, \text{ where} \tag{9}$$
$$\mathcal{L}_+(\rho) = a^\dagger\rho a - \frac{1}{2}\{aa^\dagger,\rho\} \qquad \text{and} \qquad \mathcal{L}_-(\rho) = a\rho a^\dagger - \frac{1}{2}\{a^\dagger a,\rho\},$$

where $a^\dagger$ and a are the creation and annihilation operators (i.e., $[a,a^\dagger] = \mathsf{id}$). The qOU semigroup (9) is a natural counterpart of the classical semigroup generated by the Fokker-Planck equation (see Appendix A). The unique fixed point of the semigroup $\{e^{t\mathcal{L}_{\mu,\lambda}}\}_{t\geq 0}$ is the state

$$\sigma_{\mu,\lambda} = (1-\nu)\sum_{n=0}^{\infty}\nu^n|n\rangle\langle n| \qquad \text{with} \qquad \nu = \lambda^2/\mu^2,$$

i.e., it is diagonal in the number state basis $\{|n\rangle\}_{n\in\mathbb{N}_0}$ with a geometric distribution, hence a Gaussian thermal state.

We conjecture that for an arbitrary initial state ρ, this semigroup converges to the fixed point at an exponential rate given by the exponent $\mu^2 - \lambda^2$, that is,

$$D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \leq e^{-(\mu^2-\lambda^2)t}D(\rho\|\sigma_{\mu,\lambda}), \qquad \text{for all } t \geq 0, \tag{10}$$

where $D(\rho\|\sigma) = \mathsf{tr}(\rho\log\rho - \rho\log\sigma)$ is the relative entropy. In Appendix D, we show that (10) holds for all Gaussian states ρ, and the exponent $\mu^2 - \lambda^2$ is optimal. In other words, our conjecture amounts to the statement that certain Gaussian thermal states converge "most slowly" to the fixed point, and the Log-Sobolev-1-constant, defined as the largest constant $\alpha_1 > 0$ such that $D(e^{t\mathcal{L}}(\rho)\|\sigma) \leq e^{-2\alpha_1 t}D(\rho\|\sigma)$ for any state ρ and all $t \geq 0$, is given by $\alpha_1 = \frac{1}{2}(\mu^2 - \lambda^2)$.

The quantum isoperimetric inequality (8) provides evidence for this conjecture: Taking as a specific example the case where $\lambda = 1$ and $\mu = \sqrt{2}$ (and thus $\mu^2 - \lambda^2 = 1$), we can show (see

Example 2 in Section VI) that

$$\frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\sqrt{2},1}}(\rho)\|\sigma_{\sqrt{2},1}) \leq -D(\rho\|\sigma_{\sqrt{2},1}), \qquad \text{for all states } \rho \text{ with } \mathrm{tr}(\rho a^\dagger a) \lesssim 0.67. \tag{11}$$

We also show (see Example 1 in Section VI) that recent work by De Palma, Trevisan, and Giovannetti[24] similarly implies exponential convergence of the form (10) for initial states ρ having large entropy: the isoperimetric inequality of Ref. 24 (which we discuss in more detail below) implies that

$$\frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\sqrt{2},1}}(\rho)\|\sigma_{\sqrt{2},1}) \leq -D(\rho\|\sigma_{\sqrt{2},1}), \qquad \text{for all states } \rho \text{ with } S(\rho) \gtrsim \; 2.4. \tag{12}$$

While this does not establish the scaling (10) for all states ρ, it illustrates the use of the quantum isoperimetric inequality in a concrete context. We clarify the relationship between our case and the one in the classical setting (see Section II C), where the Log-Sobolev-1 constant can be obtained from the isoperimetric inequality for the classical Fisher information, following work by Carlen.[25] In fact, this is the main motivation for our conjecture: Gaussian distributions give the optimal convergence rate in the classical problem.

Whether conjecture (10) holds is a challenging open question. We remark that it would significantly strengthen the known results about the qOU semigroup. Specifically, Carbone *et al.*[26] established that the qOU semigroup is hypercontractive. In Ref. 26, Proposition 4.2, the following inequality was shown for the Log-Sobolev-2 constant[43] α_2 of $\mathcal{L}_{\mu,\lambda}$:

$$\alpha_C^{-1} \leq \alpha_2^{-1} \leq \frac{4(5-\log(1-\nu))}{\mu^2(1-\nu)} + (3\log 3)\alpha_C^{-1}, \qquad \text{for} \qquad \nu = \lambda^2/\mu^2.$$

In this expression, α_C is the Log-Sobolev-2 constant of the associated classical birth-and-death process (which is unknown), but for which the bounds

$$\frac{\log \nu^{-1}}{5\sqrt{5}\mu^2(1-\nu)^{3/2}} \leq \alpha_C^{-1} \leq \frac{255}{4}\,\frac{(1+\log 2)(1-\nu)+\log \nu^{-1}}{\mu^2(1-\nu)^3}$$

were shown (see Ref. 26, Proposition 4.1 where α_C is denoted α_0). Following Refs. 27–29, this implies that for all states ρ, we have $D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \leq e^{-2\alpha_2 t}D(\rho\|\sigma_{\mu,\lambda})$ (respectively, we have $D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \leq e^{-\alpha_2 t}D(\rho\|\sigma_{\mu,\lambda})$), if the semigroup can be shown to be strongly (respectively weakly) L_p-regular.

The derivation of our fast convergence results from the isoperimetric inequality (8) follows, to some extent, a well-known line of reasoning considered in the classical context. Indeed, Carlen[25] has shown that the isoperimetric inequality gives rise to a Log-Sobolev inequality, which in turn provides bounds on the convergence of the classical Ornstein-Uhlenbeck (cOU) semigroup to the fixed point (see Appendix A). However, in our case we find that, while a Log-Sobolev inequality again easily follows from (8), the quantities appearing in it are not easily estimated. This concerns, in particular, the entropy rate of the attenuator

$$J_-(\rho) = 2\frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_-}(\rho)),$$

given an arbitrary initial state ρ (the factor 2 is chosen for convenience only). This is in sharp contrast to the classical case: here the trivial identity $H(e^t X) = H(X) + t$, for a scalar $t > 0$, satisfied by the differential entropy $H(X)$ of a random variable X, is sufficient for the purpose of establishing fast convergence of the cOU semigroup to the fixed point (see Appendix A).

De Palma *et al.*[24] showed that the infimum $\inf_{\rho: S(\rho)=S} J_-(\rho)$ over all states ρ with a given entropy is achieved by a Gaussian thermal state. This statement, combined with a lower bound on the corresponding quantity $J_+(\rho)$ for the amplifier from Ref. 30, valid for all states ρ, immediately yields inequality (12).

To establish (11), we prove another lower bound on $J_-(\rho)$: more precisely, we show that the infimum

$$\inf_{\rho:\, \mathrm{tr}(\rho a^\dagger a) \leq \mathbf{n}} J_-(\rho)$$

over all states ρ with mean photon number bounded by $\mathbf{n}$ is achieved by a Gaussian thermal state. The proof proceeds by reduction to the classical case using recent majorization-type results.[31,32] The latter can be treated using the results from Ref. 24. This, then, provides the required lower bound and yields statement (11). It is not, however, tight enough to establish Conjecture (10).

Understanding the relationship between entropy production rates along trajectories of the qOU semigroup, i.e., different quantities of the form $J_{\mu,\lambda}(\rho) = 2\frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_{\mu,\lambda}}(\rho))$ for $(\mu,\lambda) \neq (1,0)$, remains an open problem. We believe that progress in this direction could help provide further evidence for (or indeed lead to a proof of) the validity of conjecture (10).

C. Prior work in the classical setting

Our work in the quantum setting follows a long sequence of well-known existing arguments applicable to classical probability distributions. All geometric inequalities established here have classical counterparts, and their proofs are inspired by (and directly generalize) corresponding classical proofs. This raises the question of whether other analogues of classical results exist: for example, one may conjecture that there is a quantum counterpart of Young's inequality for the convolution operation (3).

It is not our intention to provide a complete review of this assortment of classical results: it is hardly possible to do justice to the many important developments in this area. We refer to the article Ref. 9 for a survey of many known connections. Instead, we briefly review some of the basic definitions and seminal results which are directly relevant to our work.

In the classical setting, we assume to have $\mathbb{R}^d$-valued random variables with absolutely continuous density functions. For such a random variable X with density function f (which we often assume to be non-vanishing everywhere for simplicity), a fundamental information measure of interest is the *Shannon (differential) entropy*

$$H(f) = -\int_{\mathbb{R}^d} f(\boldsymbol{x}) \log f(\boldsymbol{x}) d\boldsymbol{x},$$

also denoted as $H(X)$. The *entropy power* is given by

$$N(f) = \exp\left(2H(f)/d\right),$$

and also written as $N(X)$. Up to a factor, this quantity coincides with the variance of a Gaussian distribution having the same entropy. The *divergence*, or *relative entropy*, of two density functions f, g is defined as

$$D\left(f \| g\right) = \int_{\mathbb{R}^d} f(\boldsymbol{x}) \log \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}\, d\boldsymbol{x}. \tag{13}$$

The *Fisher information* of a random variable X with density function f is defined as the following quantity:[44]

$$J(f) = \int (\nabla f(\boldsymbol{x}))^T \cdot (\nabla f(\boldsymbol{x})) \cdot \frac{1}{f(\boldsymbol{x})} d\boldsymbol{x}, \tag{14}$$

which is associated with the family of translated probability density functions

$$f^{(\boldsymbol{\theta})}(\boldsymbol{x}) = f(\boldsymbol{x} - \boldsymbol{\theta}), \qquad \text{for } \boldsymbol{\theta} \in \mathbb{R}^d. \tag{15}$$

The quantity $J(f)$ is also often denoted as $J(X)$, to emphasize that it is the Fisher information of a random variable X which has density function f.

For two densities f and g describing random variables X and Y, the convolution operation of interest is given by (1) and describes the addition of the two random variables. Of particular interest is the case where one of the random variables is a centered normal distribution with unit variance. Such a random variable is denoted as Z below. Key results in this setting are as follows:

1. *The classical de Bruijn identity*

$$\frac{\partial}{\partial t} H\left(X + \sqrt{t} Z\right) = \frac{1}{2} J\left(X + \sqrt{t} Z\right). \tag{16}$$

This result was established by de Bruijn and gives an important relation between the Fisher information and the entropy when a random variable X is perturbed under an additive Gaussian noise channel. It is a key ingredient in proofs of many information-theoretic inequalities. A simple proof can be found in Ref. 7.

2. *The Fisher information inequality*

$$J\left(\sqrt{\lambda} X + \sqrt{1-\lambda} Y\right) \le \lambda J(X) + (1-\lambda) J(Y), \text{ for } \lambda \in [0,1], \tag{17}$$

as well as the related inequality

$$J(X+Y)^{-1} - J(X)^{-1} - J(Y)^{-1} \ge 0.$$

Proofs of these inequalities are given in Refs. 5, 7, and 33. Zamir[33] gives a particularly useful proof which relies on the information-processing inequality. This inequality states that the application of a channel cannot increase the Fisher information. Zamir's proof of the Fisher information inequality can be generalized to the quantum case.

3. *The Fisher information isoperimetric inequality*

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} \left[\frac{1}{d} J(X + \sqrt{\epsilon} Z)\right]^{-1} \ge 1.$$

This inequality implies that for Gaussian states, the inverse of the Fisher information has minimal sensitivity to additive Gaussian noise.[34]

4. *The concavity of the entropy power*

$$\frac{d^2}{d\epsilon^2}\bigg|_{\epsilon=0} N(X + \sqrt{\epsilon} Z) \le 0.$$

This celebrated result establishes that the entropy power is a concave function along trajectories of the heat flow semigroup. A proof is given in Ref. 22, and some shorter ones are presented in Refs. 23 and 34.

5. *The entropy power inequality*

$$N(X+Y) \ge N(X) + N(Y).$$

Stam[3] gave a proof of the entropy power inequality which relies on the de Bruijn identity and the Fisher information inequality. The proof was later simplified by Blachman[5] and others.[6,9]

6. *The isoperimetric inequality for entropies*

$$\frac{1}{d} J(X) N(X) \ge 2\pi e,$$

as given in Ref. 7. The isoperimetric inequality for entropies implies that Gaussians have minimal entropy power among random variables with fixed Fisher information and can be used to derive Log-Sobolev inequalities for the Ornstein-Uhlenbeck semigroup,[25] as we review in Appendix A.

III. PRELIMINARIES

A. States and information measures of interest

We consider a d-mode bosonic system with "position" and "momentum" operators (Q_k, P_k) of the kth mode satisfying the canonical commutation relations $[Q_j, P_k] = i\delta_{j,k} I$. Denoting the vector of position- and momentum-operators by $\boldsymbol{R} = (Q_1, P_1, \ldots, Q_d, P_d)$, the Weyl displacement operators are defined as

$$W(\boldsymbol{\xi}) = e^{i\sqrt{2\pi}\,\boldsymbol{\xi}\cdot(\sigma\boldsymbol{R})}, \qquad \text{for } \boldsymbol{\xi} \in \mathbb{R}^{2d}. \tag{18}$$

Here $\sigma = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus d}$ is the matrix defining the symplectic inner product. The factor $\sqrt{2\pi}$ in the definition (18) is for convenience only. From the commutation relations of position and momentum operators and the Campbell-Baker-Hausdorff formula, it is straightforward to check that the Weyl operators satisfy

$$W(\boldsymbol{\xi})W(\boldsymbol{\eta}) = e^{-i\pi\boldsymbol{\xi}\cdot(\sigma\boldsymbol{\eta})} W(\boldsymbol{\xi} + \boldsymbol{\eta}), \qquad \text{for } \boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^{2d}. \tag{19}$$

Consider a state ρ on d modes. Quantities of interest are the von Neumann entropy $S(\rho) = -\mathrm{tr}(\rho \log \rho)$, as well as the relative entropy $D(\rho\|\sigma) = \mathrm{tr}\,(\rho \log \rho - \rho \log \sigma)$. The latter expression is defined for positive operators ρ, σ, and we will assume without further comments that the states ρ, σ have full rank.

For a multi-parameter family $\{\rho^{(\boldsymbol{\theta})}\}_{\boldsymbol{\theta}\in\mathbb{R}^D}$ of states depending smoothly on the parameters $\boldsymbol{\theta}$, the divergence-based quantum Fisher information is defined as the trace of the Fisher information matrix

$$J\left(\{\rho^{(\boldsymbol{\theta})}\}; \boldsymbol{\theta}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} = \left(\frac{\partial^2}{\partial\theta_j \partial\theta_k}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} D\left(\rho^{(\boldsymbol{\theta_0})} \middle\| \rho^{(\boldsymbol{\theta})}\right) \right)_{j,k=1}^{D}.$$

This definition quantifies the dependence of the states on the parameter $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta} = \boldsymbol{\theta_0}$.

In the following, we will apply this definition to the family $\{\rho^{(\boldsymbol{\theta})}\}_{\boldsymbol{\theta}\in\mathbb{R}^{2d}}$ of states obtained by translating a given d-mode state ρ: Analogously to (15), where we translated a given probability density function f, we define the translated states

$$\rho^{(\boldsymbol{\theta})} := W(\boldsymbol{\theta})\rho W(\boldsymbol{\theta})^\dagger, \qquad \text{for } \boldsymbol{\theta} \in \mathbb{R}^{2d}. \tag{20}$$

Here translation by the parameter $\boldsymbol{\theta}$ on phase space is achieved by means of the Weyl operators. The corresponding quantity

$$J(\rho) = \mathrm{tr}\left(J\left(\{\rho^{(\boldsymbol{\theta})}\}; \boldsymbol{\theta}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{0}} \right), \tag{21}$$

will simply be called the Fisher information of ρ. Note that this definition matches that of the classical Fisher information (cf. Equation in Ref. 44). We emphasize that the concept of quantum Fisher information is non-unique (see Ref. 35), but the use of (21) is sufficient for our purposes.

B. The quantum diffusion semigroup and the de Brujin identity

Consider the Liouvillian defined on d modes as

$$\mathcal{L}_{\mathsf{heat}}(\rho) = -\pi \sum_{j=1}^{2d} [R_j, [R_j, \rho]]. \tag{22}$$

(The factor π differs from the convention used in Ref. 13 but turns out to be convenient in the proof of Theorem 3, as explained later.) The one-parameter semigroup $\{e^{t\mathcal{L}_{\mathsf{heat}}}\}_{t\geq 0}$ of completely positive trace-preserving maps (CPTPMs) generated by $\mathcal{L}_{\mathsf{heat}}$ will be called the quantum (heat) diffusion semigroup. It has various nice properties: for example, as shown in Ref. 13, a quantum version of the de Bruijn identity (16) reads

$$\frac{d}{dt}\Big|_{t=0} S\left(e^{t\mathcal{L}_{\mathsf{heat}}}(\rho)\right) = \frac{1}{2} J(\rho). \tag{23}$$

We remark that the proof of (23), which has subsequently been applied, e.g., in Ref. 21 and generalized in Ref. 14, involves certain formal manipulations whose rigorous justification remains an interesting mathematical problem: as a quantum counterpart of partial integration, for example, arguments under the trace need to be cyclically permuted. It is clear that such manipulations should be valid for sufficiently regular families of states (and indeed, are established for Gaussian states), but corresponding conditions are currently unknown. We believe that the recent introduction of Schwartz operators in Ref. 36 provides an appropriate framework to shed light on this aspect. In the following, we will assume that our states under consideration satisfy the required regularity assumptions. We hope that this issue will eventually be resolved in a similar manner as in the classical setting, where initial work by Shannon[4] was followed by a long sequence of papers with increasing rigor. In the case of the classical de Bruijn identity (16), Barron,[6] based on Stam's work,[3] has shown validity for all random variables X with finite variance.

The map $e^{t\mathcal{L}_{\text{heat}}}$ can be explicitly written as (see Refs. 13 and 37)

$$e^{t\mathcal{L}_{\text{heat}}}(\rho) = \frac{1}{(2\pi)^d}\int e^{-\|\xi\|^2/2} W(\sqrt{t}\xi)\,\rho\,W(\sqrt{t}\xi)^\dagger d\xi. \tag{24}$$

We may interpret this as the result of applying a certain convolution operation to a Gaussian distribution and a quantum state. More precisely, for $t \geq 0$, we define a convolution operation $\star_t$ between a probability density function $f : \mathbb{R}^{2d} \to \mathbb{R}$ and a d-mode state ρ by

$$(f,\rho) \mapsto f \star_t \rho := \int f(\xi) W(\sqrt{t}\xi)\,\rho\,W(\sqrt{t}\xi)^\dagger d\xi. \tag{25}$$

In this terminology, Eq. (24) becomes

$$f_Z \star_t \rho = e^{t\mathcal{L}_{\text{heat}}}(\rho),$$

where f_Z is a unit-variance centered Gaussian distribution, that is,

$$f_Z(\xi) = (2\pi)^{-d} e^{-\|\xi\|^2/2}. \tag{26}$$

To close this section, we list two elementary properties of the convolution (25) which can be checked by straightforward calculation. If $f = f_X$ is the probability density function of a random variable X and $t \geq 0$, then

$$f_X \star_t \rho = f_{\sqrt{t}X} \star_1 \rho, \tag{27}$$

where the probability density function of the rescaled random variable $\sqrt{t}X$ is given by $f_{\sqrt{t}X}(\xi) = f(\xi/\sqrt{t})/\sqrt{t}^{2d}$. Addition of random variables corresponds to convolution in the following sense:

$$f_{X_1} \star_1 (f_{X_2} \star_1 \rho) = f_{X_1+X_2} \star_1 \rho, \tag{28}$$

where $f_{X_1+X_2}$ is defined as in (1).

IV. QUANTUM GEOMETRIC INEQUALITIES

In this section, we present several statements about the convolution operation (25) and the quantum Fisher information (21). The key idea in establishing these results is the fact that the convolution operation (25) constitutes data processing, and hence provides an inequality because of the monotonicity of relative entropy. This is expressed in the following lemma. In the classical setting, the analogous argument for obtaining the Fisher information inequality (17) was first emphasized by Zamir.[33]

Lemma 1. (Data processing inequality for convolution) Let $f, g : \mathbb{R}^{2d} \to \mathbb{R}$ be probability density functions with full support. Then

$$D(f \star_t \rho \| g \star_t \sigma) \leq D(f\|g) + D(\rho\|\sigma), \tag{29}$$

for any states ρ, σ.

Proof. The quantum relative entropy satisfies the following scaling property for scalars $\lambda, \mu > 0$:

$$D(\lambda\rho \| \mu\sigma) = \lambda D(\rho \| \sigma) - \lambda \mathrm{tr}(\rho) \log \frac{\mu}{\lambda}. \tag{30}$$

Defining $h(\boldsymbol{\xi}) = \frac{g(\boldsymbol{\xi})}{f(\boldsymbol{\xi})}$, we obtain (using the translated states defined in Eq. (20))

$$\begin{aligned} D(f \star_t \rho \| g \star_t \sigma) &= D(f \star_t \rho \| (f \cdot h) \star_t \sigma) \\ &= D\left(\int f(\boldsymbol{\xi}) \rho^{(\sqrt{t}\boldsymbol{\xi})}\, \mathrm{d}\boldsymbol{\xi} \middle\| \int f(\boldsymbol{\xi}) h(\boldsymbol{\xi}) \sigma^{(\sqrt{t}\boldsymbol{\xi})}\, \mathrm{d}\boldsymbol{\xi} \right) \\ &\leq \int f(\boldsymbol{\xi}) D\left(\rho^{(\sqrt{t}\boldsymbol{\xi})} \middle\| h(\boldsymbol{\xi}) \sigma^{(\sqrt{t}\boldsymbol{\xi})} \right) \mathrm{d}\boldsymbol{\xi} \\ &= \int f(\boldsymbol{\xi}) \left(D\left(\rho^{(\sqrt{t}\boldsymbol{\xi})} \middle\| \sigma^{(\sqrt{t}\boldsymbol{\xi})} \right) - \mathrm{tr}\left(\rho^{(\sqrt{t}\boldsymbol{\xi})} \right) \log h(\boldsymbol{\xi}) \right) \mathrm{d}\boldsymbol{\xi} \\ &= D(\rho \| \sigma) - \int f(\boldsymbol{\xi}) \log \frac{g(\boldsymbol{\xi})}{f(\boldsymbol{\xi})} \\ &= D(\rho \| \sigma) + D(f \| g). \end{aligned}$$

Here the inequality we used is the joint convexity of the relative entropy (see Ref. 38, Theorem 1). For the third equality we used property (30), and for the fourth equality we used unitary invariance of the relative entropy and the trace as well as the fact that f is a probability distribution. The last equality follows from the definition (13) of the divergence. □

To convert Lemma 1 into a statement about Fisher information, we need the following covariance property of the convolution operation (25): it breaks down translations of the state $f \star_t \rho$ into translations of the function f and the state ρ, respectively.

Lemma 2. Let $\omega_q, \omega_c > 0$ and $t \geq 0$. Then

$$(f \star_t \rho)^{(\omega\boldsymbol{\theta})} = f^{(\omega_c\boldsymbol{\theta})} \star_t \rho^{(\omega_q\boldsymbol{\theta}),} \qquad \textit{for all } \boldsymbol{\theta} \in \mathbb{R}^{2d}, \tag{31}$$

where $\omega = \omega_q + \sqrt{t}\omega_c$.

Proof. According to Definitions (25) and (19) we have

$$\begin{aligned} (f \star_t \rho)^{(\omega\boldsymbol{\theta})} &= \int f(\boldsymbol{\xi}) W(\omega\boldsymbol{\theta}) W(\sqrt{t}\boldsymbol{\xi}) \rho W(\sqrt{t}\boldsymbol{\xi})^\dagger W(\omega\boldsymbol{\theta})^\dagger d\boldsymbol{\xi} \\ &= \int f(\boldsymbol{\xi}) W(\omega\boldsymbol{\theta} + \sqrt{t}\boldsymbol{\xi}) \rho W(\omega\boldsymbol{\theta} + \sqrt{t}\boldsymbol{\xi})^\dagger d\boldsymbol{\xi}. \end{aligned} \tag{32}$$

On the other hand, we similarly have

$$\begin{aligned} f^{(\omega_c\boldsymbol{\theta})} \star_t \rho^{(\omega_q\boldsymbol{\theta})} &= \int f(\boldsymbol{\xi} - \omega_c\boldsymbol{\theta}) W(\sqrt{t}\boldsymbol{\xi}) W(\omega_q\boldsymbol{\theta}) \rho W(\omega_q\boldsymbol{\theta})^\dagger W(\sqrt{t}\boldsymbol{\xi})^\dagger d\boldsymbol{\xi} \\ &= \int f(\boldsymbol{\xi} - \omega_c\boldsymbol{\theta}) W(\omega_q\boldsymbol{\theta} + \sqrt{t}\boldsymbol{\xi}) \rho W(\omega_q\boldsymbol{\theta} + \sqrt{t}\boldsymbol{\xi})^\dagger d\boldsymbol{\xi}. \end{aligned} \tag{33}$$

Through a simple change of variables and recalling that $\omega = \omega_q + \sqrt{t}\omega_c$, the claim (31) follows from (32) and (33). □

Combining the data processing inequality of Lemma 1 with Lemma 2, we prove an inequality which may be seen as a classical-quantum version of the Stam inequality. It relates the Fisher information of the state $f \star_t \rho$ to the Fisher informations of f and ρ, respectively.

Theorem 1. *(Quantum Stam inequality) Let $\omega_q, \omega_c \in \mathbb{R}$, and $t \geq 0$. Then*

$$\omega^2 J(f \star_t \rho) \leq \omega_q^2 J(\rho) + \omega_c^2 J(f), \tag{34}$$

where $\omega = \omega_q + \sqrt{t}\omega_c$. In particular,

$$J(f \star_t \rho)^{-1} - J(\rho)^{-1} - t J(f)^{-1} \geq 0. \tag{35}$$

Proof. Let $\boldsymbol{\theta_0} = (\theta_0^{(1)},\dots,\theta_0^{(2d)}) \in \mathbb{R}^{2d}$. For $j = 1,\dots 2d$ and $\theta_j \in \mathbb{R}$, introduce the vector

$$\tilde{\boldsymbol{\theta}}_{\boldsymbol{j}} = \tilde{\boldsymbol{\theta}}_{\boldsymbol{j}}(\theta_j) = (\theta_0^{(1)},\dots,\theta_0^{(j-1)},\theta_j,\theta_0^{(j+1)},\dots,\theta_0^{(2d)}).$$

Define the functions

$$f(\theta_j) := D\left(f^{(\omega_c \boldsymbol{\theta_0})} \middle\| f^{(\omega_c \tilde{\boldsymbol{\theta}}_{\boldsymbol{j}})} \right) + D\left(\rho^{(\omega_q \boldsymbol{\theta_0})} \middle\| \rho^{(\omega_q \tilde{\boldsymbol{\theta}}_{\boldsymbol{j}})} \right)$$

and

$$g(\theta_j) := D\left(f^{(\omega_c \boldsymbol{\theta_0})} \star_t \rho^{(\omega_q \boldsymbol{\theta_0})} \middle\| f^{(\omega_c \tilde{\boldsymbol{\theta}}_{\boldsymbol{j}})} \star_t \rho^{(\omega_q \tilde{\boldsymbol{\theta}}_{\boldsymbol{j}})} \right).$$

From the definition of the relative entropy and the data processing inequality (29), for every θ_j, we have

$$0 \le g(\theta_j) \le f(\theta_j),$$
$$0 = f(\theta_0^{(j)}) = g(\theta_0^{(j)}).$$

The second derivative of g can be written as the limit

$$\left.\frac{d^2}{d\theta_j^2}\right|_{\theta_j=\theta_0^{(j)}} g(\theta_j) = \lim_{\epsilon\to 0} \frac{g(\theta_0^{(j)}+\epsilon) - 2g(\theta_0^{(j)}) + g(\theta_0^{(j)}-\epsilon)}{\epsilon^2},$$

and, therefore, it is bounded

$$0 \le \left.\frac{d^2}{d\theta_j^2}\right|_{\theta_j=\theta_0^{(j)}} g(\theta_j) \le \left.\frac{d^2}{d\theta_j^2}\right|_{\theta_j=\theta_0^{(j)}} f(\theta_j).$$

Since

$$\operatorname{tr}\left(J(\{f^{(\omega_c\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \right) + \operatorname{tr}\left(J(\{\rho^{(\omega_q\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \right) = \sum_{j=1}^{2d} \left.\frac{d^2}{d\theta_j^2}\right|_{\theta_j=\theta_0^{(j)}} f(\theta_j)$$

and

$$\operatorname{tr}\left(J(\{f^{(\omega_c\boldsymbol{\theta})} \star_t \rho^{(\omega_q\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \right) = \sum_{j=1}^{2d} \left.\frac{d^2}{d\theta_j^2}\right|_{\theta_j=\theta_0^{(j)}} g(\theta_j),$$

we conclude that

$$\operatorname{tr}\left(J(\{f^{(\omega_c\boldsymbol{\theta})} \star_t \rho^{(\omega_q\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \right) \le \operatorname{tr}\left(J(\{f^{(\omega_c\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \right) + \operatorname{tr}\left(J(\{\rho^{(\omega_q\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} \right). \tag{36}$$

We remark that above inequality can also be derived as a matrix inequality without tracing both sides. However, this is not required for our purposes and our definition of the Fisher information.

Using Lemma 2, the left-hand side of this equation for $\boldsymbol{\theta_0} = 0$ can be written as

$$\operatorname{tr}\left(J(\{f^{(\omega_c\boldsymbol{\theta})} \star_t \rho^{(\omega_q\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{0}} \right) = \operatorname{tr}\left(J(\{(f \star_t \rho)^{(\omega\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{0}} \right) = \omega^2 J(f \star_t \rho). \tag{37}$$

The right-hand side of Eq. (36) can be simplified by noticing that both the classical and quantum Fisher information matrices satisfy reparametrization formulas (see Ref. 13, Lemma IV.1)

$$J(\{f^{(\omega_c\boldsymbol{\theta})}\};\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta_0}} = \omega_c^2 J(\{f^{(\boldsymbol{\theta})}\};\boldsymbol{\theta}) \qquad \text{and} \qquad J(\{\rho^{(\omega_q\boldsymbol{\theta})}\};\boldsymbol{\theta}) = \omega_q^2 J(\{\rho^{(\boldsymbol{\theta})}\};\boldsymbol{\theta}).$$

Therefore, taking $\boldsymbol{\theta} = 0$ leads to

$$\operatorname{tr}\left(J(\{f^{(\omega_c\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{0}} \right) + \operatorname{tr}\left(J(\{\rho^{(\omega_q\boldsymbol{\theta})}\};\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{0}} \right) = \omega_c^2 J(f) + \omega_q^2 J(\rho). \tag{38}$$

With (36)–(38), we arrive at the desired result (34). Finally, setting $\omega_q = \frac{J(\rho)^{-1}}{J(\rho)^{-1}+tJ(f)^{-1}}$ and $\omega_c = \frac{\sqrt{t}J(f)^{-1}}{J(\rho)^{-1}+tJ(f)^{-1}}$, we obtain (35). □

In the next lemma, we show how the quantum Stam inequality implies an isoperimetric inequality for the quantum Fisher information.

Lemma 3. (Quantum Fisher information isoperimetric inequality) The following inequality holds:

$$\frac{d}{dt}\Big|_{t=0}\Big[\frac{1}{2d}J(e^{t\mathcal{L}_{\text{heat}}}(\rho))\Big]^{-1} \geq 1. \tag{39}$$

Proof. Recall that we have $e^{t\mathcal{L}_{\text{heat}}}(\rho) = f_Z \star_t \rho$ for a Gaussian random variable Z (26). In the quantum Stam inequality (35), take $f = f_Z$, then

$$\frac{1}{t}\left(J(f_Z \star_t \rho)^{-1} - J(\rho)^{-1}\right) \geq J(f_Z)^{-1} = (2d)^{-1}.$$

Taking the limit $t \to 0$, we arrive at the desired inequality. □

The isoperimetric inequality is tight for one mode ($d = 1$) and saturated by the Gaussian thermal state

$$\omega_{\mathbf{n}} = \frac{1}{\mathbf{n}+1}\sum_{j=0}^{\infty}\left(\frac{\mathbf{n}}{\mathbf{n}+1}\right)^{j}|j\rangle\langle j|, \tag{40}$$

with mean-photon number $\mathbf{n}$: As shown in Appendix B, we have

$$\frac{d}{dt}\Big|_{t=0}\Big[\frac{1}{2}J(e^{t\mathcal{L}_{\text{heat}}}(\omega_{\mathbf{n}}))\Big]^{-1} = \frac{1}{\mathbf{n}(\mathbf{n}+1)}\log^{-2}\left(1+\frac{1}{\mathbf{n}}\right) \to 1, \qquad \text{as } \mathbf{n}\to\infty.$$

As in classical information theory, the isoperimetric inequality for the quantum Fisher information implies concavity of the entropy power under diffusion as an immediate consequence. We define the entropy power as

$$N(\rho) = \exp\left(S(\rho)/d\right). \tag{41}$$

Theorem 2. *(Concavity of the quantum entropy power) The entropy power along trajectories of the diffusion semigroup (22) is concave, i.e.,*

$$\frac{d^2}{dt^2}\Big|_{t=0}N(e^{t\mathcal{L}_{\text{heat}}}(\rho)) \leq 0.$$

Proof. Two applications of de Bruijn identity (23) yield

$$\frac{d^2}{dt^2}\Big|_{t=0}N(e^{t\mathcal{L}_{\text{heat}}}(\rho)) = N(\rho)\left(\Big[\frac{1}{2d}J(\rho)\Big]^2 + \frac{1}{2d}\frac{d}{dt}\Big|_{t=0}J(e^{t\mathcal{L}_{\text{heat}}}(\rho))\right).$$

The quantum Fisher information isoperimetric inequality (39) is equivalent to

$$\frac{1}{2d}J(\rho)^2 + \frac{d}{dt}\Big|_{t=0}J(e^{t\mathcal{L}_{\text{heat}}}(\rho)) \leq 0.$$

This completes the proof. □

To proceed, we establish bounds on the asymptotic scaling of the entropy power for large times. The following lemma follows directly from Refs. 5 and 13, Corollary III.4 (see also Ref. 39).

Lemma 4 (Asymptotic scaling of the entropy power under the heat flow). Let

$$C_c = \frac{1}{2}\sum_{j=1}^{2d}\frac{\partial^2}{\partial\xi_j^2} \tag{42}$$

be the generator of the classical heat diffusion semigroup on $\mathbb{R}^{2d}$. In the limit $t \to \infty$, we have

$$\exp\left(H(e^{tC_c}(f))/d\right) = (2\pi e)t + O(1),$$
$$\exp\left(S(e^{t\mathcal{L}_{\text{heat}}}(\rho))/d\right) = (2\pi e)t + O(1),$$

independent of the probability density function f on $\mathbb{R}^{2d}$ and the d-mode state ρ, respectively.

Having the same scaling for classical and quantum heat flows motivated the choice of constants in (18) and (22).

Note that if X is a random variable with probability density function f, then $e^{tC_c}(f)$ is the probability density function of the random variable $X + \sqrt{t}Z$ obtained by adding a centered Gaussian random variable with unit variance, see Eq. (26).

Lemma 5. We have

$$e^{\xi \mathcal{L}_{\text{heat}}}(f \star_t \rho) = e^{\nu C_c}(f) \star_t e^{\mu \mathcal{L}_{\text{heat}}}(\rho),$$

whenever $\xi = \mu + t\nu$.

Proof. Observe that writing $\star$ for $\star_1$, we get

$$\begin{aligned} e^{\xi \mathcal{L}_{\text{heat}}}(f_X \star_t \rho) &= f_Z \star_\xi (f_{\sqrt{t}X} \star \rho) \\ &= f_{\sqrt{\xi}Z} \star (f_{\sqrt{t}X} \star \rho) \\ &= f_{\sqrt{\xi}Z + \sqrt{t}X} \star \rho. \end{aligned}$$

On the other hand,

$$\begin{aligned} e^{\nu C_c}(f) \star_t e^{\mu \mathcal{L}_{\text{heat}}}(\rho) &= f_{X+\sqrt{\nu}Z_1} \star_t \left(e^{\mu \mathcal{L}_{\text{heat}}}(\rho)\right) \\ &= f_{\sqrt{t}(X+\sqrt{\nu}Z_1)} \star \left(f_{\sqrt{\mu}Z_2} \star \rho\right) \\ &= f_{\sqrt{t}X+\sqrt{t\nu}Z_1+\sqrt{\mu}Z_2} \star \rho \\ &= f_{\sqrt{t}X+\sqrt{\mu+t\nu}Z} \star \rho \\ &= f_{\sqrt{t}X+\sqrt{\xi}Z} \star \rho. \end{aligned}$$

We have used properties (27) and (28). In the penultimate step, we have used that for independent unit-variance centered Gaussian random variables Z_1 and Z_2, we have $aZ_1 + bZ_2 = \sqrt{a^2 + b^2}Z$. Hence the two expressions are equal and the statement follows. □

The next theorem presents the entropy power inequality for both the convolution operation (25) and the heat diffusion semigroup (22).

Theorem 3. *(Entropy power inequality) For* $t \geq 0$*, the following inequality holds:*

$$N(f \star_t \rho) \geq N(\rho) + tN(f).$$

In particular, choosing $f = f_Z$ *as the distribution of a unit-variance centered Gaussian defined in Eq. (26), we have*

$$N(e^{t\mathcal{L}_{\text{heat}}}(\rho)) \geq N(\rho) + t\, 2\pi e. \tag{43}$$

Proof. The proof is inspired by the proof of the entropy power inequality in Ref. 13, which itself is inspired by the proof for classical random variables by Blachman.[5] Here we provide all necessary details modified to the present situation. For $\mu, \nu, \xi \geq 0$, define functions

$$\begin{aligned} \mu &\to E_A(\mu) := \exp\left(S(e^{\mu \mathcal{L}_{\text{heat}}}(\rho))/d\right), \\ \nu &\to E_B(\nu) := \exp\left(H(e^{\nu \mathcal{L}_{\text{heat,cl}}}(f))/d\right), \\ \xi &\to E_C(\xi) := \exp\left(S(e^{\xi \mathcal{L}_{\text{heat}}}(f \star_t \rho)))/d\right), \end{aligned}$$

where C_c is the generator of the classical heat semigroup as defined in Eq. (42).

The initial value problems

$$\begin{aligned} \dot{\mu}(s) &= E_A(\mu(s)), \;\; \mu(0) = 0, \\ \dot{\nu}(s) &= E_B(\nu(s)), \;\; \nu(0) = 0, \end{aligned} \tag{44}$$

have solutions $\mu(\cdot), \nu(\cdot)$. Fix a pair of such solutions $(\mu(\cdot), \nu(\cdot))$ and $t \geq 0$. Define

$$\xi(s) := \mu(s) + t\nu(s). \tag{45}$$

These functions diverge, i.e.,

$$\lim_{s\to\infty} \mu(s) = \lim_{s\to\infty} \nu(s) = \lim_{s\to\infty} \xi(s) = \infty, \tag{46}$$

because of (44) and $E_{A,B} \geq 1$.

Consider the function

$$\delta(s) := \frac{E_A(\mu(s)) + tE_B(\nu(s))}{E_C(\xi(s))}.$$

With the initial conditions (44), it follows that the claim of the theorem is equivalent to

$$\delta(0) \leq 1.$$

This inequality follows from two facts: first, the fact that

$$\lim_{s\to\infty} \delta(s) = 1,$$

as follows from the asymptotic scaling shown in Lemma 4, the divergence (46), and the choice (45) of $\xi(s)$; second, the fact that

$$\dot{\delta}(s) \geq 0, \qquad \text{for all } s \geq 0. \tag{47}$$

It remains to show identity (47). Computing the derivative of δ leads to the following equality:

$$\dot{\delta}(s) = \frac{\dot{E}_A(\mu(s))\dot{\mu}(s) + t\dot{E}_B(\nu(s))\dot{\nu}(s)}{E_C(\xi(s))} - \frac{E_A(\mu) + tE_B(\nu)}{E_C(\xi)^2}\dot{E}_C(\xi(s))\dot{\xi}(s). \tag{48}$$

Define the Fisher informations

$$J_A(\mu) := J(e^{\mu\mathcal{L}_{\mathrm{heat}}}(\rho)),$$
$$J_B(\nu) := J(e^{\nu\mathcal{C}_{\mathrm{c}}}(f)),$$
$$J_C(\xi) := J(e^{\xi\mathcal{L}_{\mathrm{heat}}}(f \star_t \rho)),$$

for $\mu, \nu, \xi \geq 0$. From the quantum de Bruijn identity (23) and the classical de Bruijn identity (16) we obtain

$$\dot{E}_V(\zeta) = \frac{1}{2d}E_V(\zeta)J_V(\zeta), \qquad \text{where } V \in \{A, B, C\}.$$

With these identities and Eq. (44), Eq. (48) is equivalent to

$$2d\dot{\delta}(s) = \frac{E_A^2 J_A + tE_B^2 J_B}{E_C} - \frac{(E_A + tE_B)^2}{E_C^2}E_C J_C, \tag{49}$$

where we have used the shorthand notation $E_A = E_A(\mu(s))$, $E_B = E_B(\nu(s))$, and $E_C = E_C(\xi(s))$, and similarly for $J_{A,B,C}$.

Recall that by Lemma 5, $e^{\xi\mathcal{L}_{\mathrm{heat}}}(f \star_t \rho) = e^{\nu\mathcal{C}_{\mathrm{c}}}(f) \star_t e^{\mu\mathcal{L}_{\mathrm{heat}}}(\rho)$, and by the quantum Stam inequality (35), we have the bound

$$J_C \leq \frac{J_A J_B}{tJ_A + J_B}.$$

Inserting this upper bound into (49), we obtain

$$2d\dot{\delta}E_C \geq E_A^2 J_A + tE_B^2 J_B - (E_A + tE_B)^2 \frac{J_A J_B}{tJ_A + J_B} = \frac{t(E_A J_A - E_B J_B)^2}{tJ_A + J_B} \geq 0.$$

This proves (47). □

As the last statement in this section, we derive an isoperimetric inequality for entropies from the entropy power inequality.

Theorem 4. *(Isoperimetric inequality for entropies) We have*

$$\frac{1}{d} J(\rho) N(\rho) \geq 4\pi e. \tag{50}$$

Proof. Applying the de Bruijn identity (23) to the definition (41) of the entropy power $N(\rho)$, we obtain

$$\frac{d}{dt}\Big|_{t=0} N(e^{t\mathcal{L}_{\mathrm{heat}}}(\rho)) = \frac{1}{2d} J(\rho) N(\rho).$$

On the other hand, for $t \geq 0$, the entropy power inequality (43) reduces to

$$\frac{1}{t}[N(e^{t\mathcal{L}_{\mathrm{heat}}}(\rho)) - N(\rho)] \geq 2\pi e.$$

Therefore, taking the limit $t \to 0$, we obtain the desired bound. □

To conclude this section, we remark that the isoperimetric inequality for entropies (50) is tight in the one-mode case, $d = 1$: For a Gaussian thermal state $\omega_{\mathbf{n}}$ (40) with mean photon number $\mathbf{n}$ we obtain

$$J(\omega_{\mathbf{n}}) N(\omega_{\mathbf{n}}) = 4\pi \left(\frac{\mathbf{n}+1}{\mathbf{n}}\right)^{\mathbf{n}} \log \left(\frac{\mathbf{n}+1}{\mathbf{n}}\right)^{\mathbf{n}+1} \to 4\pi e, \qquad \text{for } \mathbf{n} \to \infty.$$

Detailed calculations are provided in Appendix C.

V. GAUSSIAN OPTIMALITY FOR ENERGY-CONSTRAINED ENTROPY RATES

In this section, we show that Gaussian thermal states minimize the entropy rate for the one-mode attenuator semigroup among states with bounded mean photon number. The semigroup is defined by its generator

$$\mathcal{L}_{-}(\rho) = a\rho a^{\dagger} - \frac{1}{2}\{a^{\dagger}a, \rho\}. \tag{51}$$

We prove the following theorem:

Theorem 5. *For any* $\mathbf{n} > 0$, *the infimum* $\inf_{\rho:\mathrm{tr}(\hat{n}\rho)\leq \mathbf{n}} \frac{d}{dt}\big|_{t=0} S(e^{t\mathcal{L}_{-}}(\rho))$ *over states* ρ *with mean photon number* $\mathbf{n}$ *is achieved by the Gaussian thermal state* $\omega_{\mathbf{n}}$ *defined in (40). In particular,*

$$\inf_{\rho:\mathrm{tr}(\hat{n}\rho)\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_{-}}(\rho)) = \begin{cases} -\mathbf{n}\log\left(1 + \dfrac{1}{\mathbf{n}}\right) & \text{if } \mathbf{n} > 0, \\ 0 & \text{if } \mathbf{n} = 0. \end{cases}$$

The proof proceeds by reduction to a recent result by De Palma, Trevisan, and Giovannetti,[24] where it is shown that Gaussian thermal states minimize the entropy rate of the quantum attenuator among all states with a given input entropy.[45] Our argument additionally uses the recently introduced concept of Fock majorization and associated results by Jabbour, Garcia-Patron, and Cerf,[31] as well as the (classical) Gaussian maximum entropy principle.

In more detail, we first show that the problem of minimizing the entropy rate reduces to the study of properties of a classical semigroup describing a pure-death process. This connection was used previously in Ref. 26 (see Section II B) and is also an essential first step in Ref. 24. More explicitly, in Section V A (Theorem 6) we prove the identity

$$\inf_{\rho:\mathrm{tr}(\hat{n}\rho)\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_{-}}(\rho)) = \inf_{p:\mathbb{E}_p[N]\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} H(e^{tC_{-}}(p)).$$

Here the infimum is over all probability distributions p on $\mathbb{N}_0$ with expectation value $\mathbb{E}_p[N]$ bounded by $\mathbf{n}$, the quantity $H(p)$ is the Shannon entropy of the distribution p, and C_{-} is the generator of a semigroup describing a classical pure-death process (see (52) for a precise definition).

Using the results of Ref. 24, we then show that the entropy rate for the classical process is optimized by a geometric distribution: In Section V B (Theorem 6) we prove that for $\mathbf{n} > 0$

$$\inf_{p:\mathbb{E}_p[N]\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} H(e^{tC_-}(p)) = -\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right).$$

Finally, in Section V C, we calculate the entropy rate for the Gaussian thermal state $\omega_{\mathbf{n}}$ with mean photon number at most $\mathbf{n}$ and find that

$$\frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_-}(\omega_{\mathbf{n}})) = -\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right).$$

A. Connection to a classical pure-death process

For an initial state $\rho = \sum_n p_n |n\rangle\langle n|$, which is diagonal in the number state basis $\{|n\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}}|0\rangle\}_{n\in\mathbb{N}_0}$ (where $\hat{n}\,|0\rangle = 0$), the time-evolved state has the same form, i.e., $\rho(t) = e^{t\mathcal{L}_-}(\rho) = \sum_n p_n(t)|n\rangle\langle n|$. Thus the attenuator semigroup $\{e^{t\mathcal{L}_-}\}_{t\geq 0}$ gives rise to a semigroup $\{e^{tC_-}\}_{t\geq 0}$ on classical probability distributions by $p(t) = e^{tC_-}(p)$. Its generator C_- describes the dynamics of a classical pure-death process. It can be obtained from (51) by inserting a number state $|n\rangle$: it is straightforward to check that

$$\mathcal{L}_-(|n\rangle\langle n|) = \begin{cases} n(|n-1\rangle\langle n-1| - |n\rangle\langle n|) & \text{for } n > 0, \\ 0 & \text{for } n = 0. \end{cases}$$

In particular, the coefficients $\{p_n(t)\}_{n\in\mathbb{N}_0}$ satisfy the system of differential equations

$$\dot{p}_n(t) = -np_n(t) + (n+1)p_{n+1}(t), \qquad \text{for all } n \in \mathbb{N}_0, \tag{52}$$

with initial condition $p_n(0) = p_n$ for $n \in \mathbb{N}_0$. The expression on the right-hand side of (52) defines the generator C_-, that is, we have

$$(C_-(p))_n = -np_n + (n+1)p_{n+1}, \quad \text{for all } n \in \mathbb{N}_0. \tag{53}$$

The following theorem reduces the problem of minimizing the entropy rate for the quantum attenuator semigroup to the classical problem of minimizing the entropy rate for this pure-death process.

Theorem 6 (Correspondence to classical problem). *We have the identity*

$$\inf_{\rho:\mathrm{tr}(\hat{n}\rho)\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_-}(\rho)) = \inf_{p:\mathbb{E}_p[N]\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} H(e^{tC_-}(p)),$$

where $\mathbb{E}_p[N] = \sum_{n=0}^\infty np_n$ *and* $H(p) = -\sum_{n=0}^\infty p_n\log p_n$ *are the expectation value and entropy of the distribution* p*, respectively.*

The proof of Theorem 6 relies on results obtained in Refs. 32 and 31. Here we review the necessary definitions and results and specialize them to our situation.

Definition 1 (Majorization). Let p *and* q *be the decreasing summable sequences of positive numbers. Then* p *weakly sub-majorizes* q, $q \prec_w p$, *if and only if*

$$\sum_{i=0}^n q_i \leq \sum_{i=0}^n p_i, \qquad \textit{for all } n \in \mathbb{N}.$$

Definition 2. Let P *and* Q *be the positive trace-class operators with eigenvalues* $\{p_n\}_{n\in\mathbb{N}}$ *and* $\{q_n\}_{n\in\mathbb{N}}$*, arranged in decreasing order. Then* P *weakly majorizes* Q*, i.e.,* $Q \prec_w P$*, if and only if* $q \prec_w p$*. Also,* P *majorizes* Q*, i.e.,* $Q \prec P$*, if the traces of* P *and* Q *are identical and* $Q \prec_w P$.

Definition 3 (Fock rearrangement). Let X be a positive trace-class operator with eigenvalues $\{x_n\}_{n\in\mathbb{N}_0}$ in decreasing order. The Fock rearrangement (or passive rearrangement) is defined as

$$X^{\downarrow} := \sum_{n=0}^{\infty} x_n \, |n\rangle \, \langle n| \, .$$

Our restriction on the mean photon number forces us to consider an additional majorization relation. In contrast, in Ref. 24, where the authors restrict the input entropy instead, Fock majorization does not need to be considered.

Definition 4 (Fock majorization). The Fock majorization relation, denoted $\prec_F$, was introduced in Ref. 31 as follows:

$$\sigma \prec_F \rho \qquad \Leftrightarrow \qquad \mathrm{tr}(\Pi_n \sigma) \leq \mathrm{tr}(\Pi_n \rho), \quad \textit{for all } n \in \mathbb{N}_0, \tag{54}$$

where

$$\Pi_n = \sum_{j=0}^{n} |j\rangle\langle j|. \tag{55}$$

Theorem 7. *(Ref. 32) For any state ρ and all $t \geq 0$ we have*

$$e^{t\mathcal{L}_-}(\rho) \prec e^{t\mathcal{L}_-}(\rho^{\downarrow}). \tag{56}$$

This implies that

$$S\left(e^{t\mathcal{L}_-}(\rho)\right) \geq S\left(e^{t\mathcal{L}_-}(\rho^{\downarrow})\right).$$

Proof. The first statement is shown in Ref. 32, Eq. (VI.10), and the second statement then follows from Ref. 32, Theorem III.3. □

Lemma 6 (Ref. 32, Lemma IV.9). Suppose X, Y, Z are positive trace-class operators with

$$Y \prec_w Z \qquad \textit{and} \qquad Z^{\downarrow} = Z.$$

Then

$$\mathrm{tr}(XY) \leq \mathrm{tr}(X^{\downarrow} Z).$$

We begin proving Theorem 6 by investigating the change of the mean photon number under the Fock rearrangement procedure.

Lemma 7. For any state ρ, the Fock rearrangement does not increase the mean photon number

$$\mathrm{tr}(\hat{n}\rho^{\downarrow}) \leq \mathrm{tr}(\hat{n}\rho),$$

where $\hat{n} = a^{\dagger}a$.

Proof. Let $n \in \mathbb{N}_0$ be arbitrary, and set $X = \Pi_n$ with Π_n defined in (55). Then clearly $X^{\downarrow} = X = \Pi_n$. Setting $Y = \rho$, $Z = \rho^{\downarrow}$, we have $Z^{\downarrow} = Z$, and $Y \prec Z$, according to (56) for $t = 0$, i.e.,

$$\rho \prec \rho^{\downarrow}.$$

Thus, Lemma 6 leads to $\mathrm{tr}(\Pi_n \rho) \leq \mathrm{tr}(\Pi_n \rho^{\downarrow})$. According to Definition (54), this is equivalent to

$$\rho \prec_F \rho^{\downarrow}.$$

It was shown in Ref. 31, Eq. (5) that

$$\rho \prec_F \sigma \qquad \Rightarrow \qquad \mathrm{tr}(\hat{n}\sigma) \leq \mathrm{tr}(\hat{n}\rho).$$

The claim follows by taking $\sigma = \rho^{\downarrow}$ in the last statement. □

Proof of Theorem 6. Since $S(\rho^{\downarrow}) = S(\rho)$, Theorem 7 implies

$$\frac{d}{dt}\Big|_{t=0} S\left(e^{t\mathcal{L}_-}(\rho)\right) \geq \frac{d}{dt}\Big|_{t=0} S\left(e^{t\mathcal{L}_-}(\rho^{\downarrow})\right).$$

With Lemma 7, we therefore obtain

$$\inf_{\rho:\mathrm{tr}(\hat{n}\rho)\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} S\left(e^{t\mathcal{L}_-}(\rho)\right) = \inf_{\rho=\rho^{\downarrow}:\mathrm{tr}(\hat{n}\rho^{\downarrow})\leq \mathbf{n}} \frac{d}{dt}\Big|_{t=0} S\left(e^{t\mathcal{L}_-}(\rho^{\downarrow})\right).$$

Since $\rho^{\downarrow}$ is a passive, Fock-rearranged state, the right-hand side is a classical problem related to the pure-death process (52), and, therefore, can be replaced with the infimum of the entropy rate of a probability distribution evolving under a pure-death process. Thus the claim follows.

B. Geometric distributions optimize entropy rates of the classical death process under energy constraint

For a probability distribution p on $\mathbb{N}_0$, let

$$J_-(p) := 2\frac{d}{dt}\Big|_{t=0} H(e^{t\mathcal{C}_-}(p))$$

denote the entropy rate when p evolves under the classical death-process $\mathcal{C}_-$ (cf. (53)). We are interested in distributions p with a fixed expectation value $\mathbb{E}_p[N] = \sum_{n=0}^{\infty} np_n$. The main result we use here is Ref. 24, Theorem 24: it states that for any probability distribution p on $\mathbb{N}_0$, the quantity $J_-(p)$ is bounded by

$$\inf_{p:H(p)\leq H} J_-(p) \geq 2 \inf_{p:H(p)\leq g(\mathbf{n})} f(H(p)) \tag{57}$$

where $f(H) = -g^{-1}(H)g'(g^{-1}(H))$ and where $g(\mathbf{n}) = (\mathbf{n}+1)\log(\mathbf{n}+1) - \mathbf{n}\log\mathbf{n}$ is the entropy of a geometric distribution with expectation value $\mathbf{n}$ (or equivalently the entropy of a Gaussian state with mean photon number $\mathbf{n}$). We use this to show the following:

Theorem 8. *The infimum* $\inf_{p:\mathbb{E}[N]\leq\mathbf{n}} J_-(p)$ *is achieved by the geometric distribution* $p_k^{\mathrm{geo},\mathbf{n}} = (1-r)r^k$ *with* $r = \frac{\mathbf{n}}{\mathbf{n}+1}$. *In particular, for* $\mathbf{n} > 0$,

$$\inf_{p:\mathbb{E}[N]\leq\mathbf{n}} J_-(p) = -2\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right).$$

Proof. We show that

$$\inf_{p:\mathbb{E}[N]=\mathbf{n}} J_-(p) = J(p^{\mathrm{geo},\mathbf{n}}) = -2\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right). \tag{58}$$

Since the right-hand side is monotonically decreasing with $\mathbf{n}$, Eq. (58) implies the claim of the theorem.

The geometric distribution $p^{\mathrm{geo},\mathbf{n}} = (1-r)r^k$ with $r = \frac{\mathbf{n}}{\mathbf{n}+1}$ has expectation value $\mathbb{E}_{p^{\mathrm{geo},\mathbf{n}}}[N] = \mathbf{n}$ and entropy $H(p^{\mathrm{geo},\mathbf{n}}) = g(\mathbf{n})$. By the maximum entropy principle (Ref. 40, Chapter 12), we know that geometric distributions are the distributions with maximal entropy among all distributions with a fixed expectation value $\mathbb{E}_p[N]$. Therefore we have

$$\mathbb{E}_p[N] = \mathbf{n} \qquad \Rightarrow \qquad H(p) \leq g(\mathbf{n}).$$

Combining this with (57) we obtain

$$\inf_{p:\mathbb{E}_p[N]=\mathbf{n}} J_-(p) \geq \inf_{p:H(p)\leq g(\mathbf{n})} J_-(p) \geq 2 \inf_{p:H(p)\leq g(\mathbf{n})} f(H(p)),$$

Since f is decreasing by Ref. 24, Lemma 5, it follows that

$$\inf_{p:\mathbb{E}[N]=\mathbf{n}} J_-(p) \geq 2f(g(\mathbf{n})) = -2\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right). \tag{59}$$

However, since $\mathbb{E}_{p^{\mathrm{geo},\mathbf{n}}}[N] = \mathbf{n}$ and $J_-(p^{\mathrm{geo},\mathbf{n}}) = -2\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right)$, we have equality in (59). □

C. Gaussian optimality of entropy rates for the quantum attenuator semigroup

Proof of Theorem 5. From Theorems 6 and 8 we have

$$\inf_{\rho:\mathrm{tr}(\hat{n}\rho)\leq\mathbf{n}} \frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_-}(\rho)) = -\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right).$$

Let $\omega_{\mathbf{n}}$ be the Gaussian thermal state with mean photon number $\mathbf{n}$ as defined in (40). We have $S(\omega_{\mathbf{n}}) = g(\mathbf{n}) = (\mathbf{n}+1)\log(\mathbf{n}+1) - \mathbf{n}\log\mathbf{n}$ for $\mathbf{n} > 0$. Under the map $e^{t\mathcal{L}_-}$, the state $\omega_{\mathbf{n}}$ evolves into the thermal state $\omega_{\mathbf{n}_t}$ according to

$$e^{t\mathcal{L}_-}(\omega_{\mathbf{n}}) = \omega_{\mathbf{n}_t}, \qquad \text{where} \qquad \mathbf{n}_t = e^{-t}\mathbf{n}.$$

In particular,

$$\frac{d}{dt}\Big|_{t=0} S(e^{t\mathcal{L}_-}(\omega_{\mathbf{n}})) = g'(\mathbf{n})\mathbf{n}'_t\Big|_{t=0} = -\mathbf{n}\log\left(1+\frac{1}{\mathbf{n}}\right).$$

Therefore the considered infimum is achieved by the Gaussian thermal state $\omega_{\mathbf{n}}$. □

VI. APPLICATION TO FAST CONVERGENCE OF THE ORNSTEIN-UHLENBECK SEMIGROUP

In this section, we consider a one-parameter group of CPTP maps $\{e^{\mathcal{L}_{\mu,\lambda}}\}_{t\geq 0}$ generated by the linear combination

$$\mathcal{L}_{\mu,\lambda} = \mu^2\mathcal{L}_- + \lambda^2\mathcal{L}_+, \qquad \text{for } \mu > \lambda > 0.$$

where $\mathcal{L}_-$ is defined by (51), and $\mathcal{L}_+$ is defined by

$$\mathcal{L}_+(\rho) = a^\dagger\rho a - \frac{1}{2}\{aa^\dagger, \rho\}.$$

In the following, we use the entropy production rates

$$J_\pm(\rho) := 2\frac{d}{dt}S(e^{t\mathcal{L}_\pm}(\rho)).$$

The factor 2 here is for convenience to match de Bruijn identity (23) for $J(\rho)$. These quantities are related to the Fisher information $J(\rho)$ by

$$J(\rho) = 2\pi(J_-(\rho) + J_+(\rho)), \tag{60}$$

because of de Bruijn's identity and the fact that $\mathcal{L}_{\mathsf{heat}} = 2\pi\mathcal{L}_- + 2\pi\mathcal{L}_+$.

Lemma 8. Let $\mu > \lambda > 0$. Then

$$-\zeta D(\rho\|\sigma_{\mu,\lambda}) - \frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) = \frac{\mu^2}{2}J_-(\rho) + \frac{\lambda^2}{2}J_+(\rho) + \zeta S(\rho) + \lambda^2\log\nu + \zeta\log(1-\nu), \tag{61}$$

for any state ρ, where $\nu = \frac{\lambda^2}{\mu^2}$, $\zeta = \mu^2 - \lambda^2$, and $\sigma_{\mu,\lambda}$ is the fixed point of $\mathcal{L}_{\mu,\lambda}$.

Proof. As mentioned in Section II B, the unique fixed point of the semigroup $\{e^{t\mathcal{L}_{\mu,\lambda}}\}_{t\geq 0}$ is the state

$$\sigma_{\mu,\lambda} = (1-\nu)\sum_{n=0}^{\infty}\nu^n|n\rangle\langle n| = (1-\nu)\nu^{\hat{n}},$$

where $\nu = \lambda^2/\mu^2$. In particular, this implies that for any state ρ,

$$D(\rho\|\sigma_{\mu,\lambda}) = -S(\rho) - \mathrm{tr}(\rho\log\sigma_{\mu,\lambda}) = -S(\rho) - (\log\nu)\mathrm{tr}(\rho\hat{n}) - \log(1-\nu). \tag{62}$$

Straightforward calculations show that the mean photon number of ρ converges to the mean photon number of the fixed point $\sigma_{\mu,\lambda}$ with an exponential rate

$$\mathbf{n}_t = \mathrm{tr}(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\hat{n}) = \mathrm{tr}(\rho e^{t\mathcal{L}^{\dagger}_{\mu,\lambda}}(\hat{n})) = e^{-(\mu^2-\lambda^2)t}\mathrm{tr}(\rho\hat{n}) + (1-e^{-(\mu^2-\lambda^2)t})\mathbf{n}_\infty,$$

where $\mathbf{n}_\infty = \mathrm{tr}(\sigma_{\mu,\lambda}\hat{n}) = \frac{\lambda^2}{\mu^2-\lambda^2} = \frac{\nu}{1-\nu}$. Therefore

$$\frac{d}{dt}\Big|_{t=0}\mathrm{tr}(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\hat{n}) = -(\mu^2-\lambda^2)\left(\mathrm{tr}(\rho\hat{n}) - \mathbf{n}_\infty\right).$$

Combining the last equality with (62), we find that

$$-\frac{d}{dt}\Big|_{t=0}D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) = \frac{\mu^2}{2}J_-(\rho) + \frac{\lambda^2}{2}J_+(\rho) - (\log\nu)(\mu^2-\lambda^2)\mathrm{tr}(\rho\hat{n}) + \lambda^2\log\nu. \tag{63}$$

With (62) and (63), and setting $\zeta = \mu^2-\lambda^2$, we obtain the desired equality. □

The choice of $\zeta = \mu^2-\lambda^2$ in the lemma is motivated by Gaussian states: in Appendix D we show that for any Gaussian state ρ, the following inequality holds:

$$\frac{d}{dt}\Big|_{t=0}D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \le -\zeta D(\rho\|\sigma_{\mu,\lambda}) \qquad \text{with} \qquad \zeta = \mu^2-\lambda^2 > 0.$$

Furthermore, for any $\epsilon > 0$ there exists a Gaussian state ρ such that

$$\frac{d}{dt}\Big|_{t=0}D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \ge -(\zeta+\epsilon)D(\rho\|\sigma_{\mu,\lambda}).$$

Let us now consider a specific example of a quantum Ornstein-Uhlenbeck process.

Example 1. Consider $\mu^2 = 2$, $\lambda^2 = 1$. *Then*

$$\frac{d}{dt}\Big|_{t=0}D(e^{t\mathcal{L}_{\sqrt{2},1}}(\rho)\|\sigma_{\sqrt{2},1}) \le -D(\rho\|\sigma_{\sqrt{2},1}),$$

for any state ρ *with* $S(\rho) \gtrsim 2.06$. *In comparison, the entropy of the fixed point is* $S(\sigma_{\sqrt{2},1}) = 2\log(2) \approx 1.39$.

In Lemma 8 we can bound $J_+(\rho) \ge 2$ because of Ref. 30, Eq. (43), and the linear combination of $J_-(\rho)$ and $S(\rho)$ can be bounded by the result of De Palma *et al.*[24] That is, for any state ρ with $S(\rho) \ge S_0$,

$$\frac{\mu^2}{2}J_-(\rho) + \zeta S(\rho) \ge \inf_{S\ge S_0}\left(\mu^2 f(S) + \zeta S\right),$$

where $f(S) = -g^{-1}(S)g'(g^{-1}(S))$ and $g(\mathbf{n}) = (\mathbf{n}+1)\log(\mathbf{n}+1) - \mathbf{n}\log(\mathbf{n})$. Substituting $S = g(\mathbf{n})$ gives

$$\frac{\mu^2}{2}J_-(\rho) + \zeta S(\rho) \ge \inf_{\mathbf{n}\ge g^{-1}(S_0)}\left(\mu^2(-\mathbf{n}g'(\mathbf{n})) + \zeta g(\mathbf{n})\right) =: F(S_0).$$

That is, for any state with $S(\rho) \ge S_0$, we have

$$-\zeta D(\rho\|\sigma_{\mu,\lambda}) - \frac{d}{dt}\Big|_{t=0}D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \ge F(S_0) + \lambda^2 + \lambda^2\log\nu + \zeta\log(1-\nu).$$

With the choice $\mu^2 = 2$, $\lambda^2 = 1$, the function $F(S)$ is monotonically increasing for $S \gtrsim 0.5$. For $S_0 \gtrsim 2.06$, the right-hand side of the last inequality is non-negative. □

The isoperimetric inequality for entropies (50) for one mode ($d = 1$) can be written as

$$-S(\rho) \le \log\left(\frac{1}{4\pi e}J(\rho)\right).$$

For any $A > 0$, using $\log x \leq x - 1$, we get

$$\begin{aligned}\log\Big(\frac{1}{4\pi e}J(\rho)\Big) &= \log\Big(\frac{AJ(\rho)}{4\pi eA}\Big)\\ &= \log\Big(\frac{1}{4\pi eA}\Big) + \log\Big(AJ(\rho)\Big)\\ &\leq AJ(\rho) - 2 - \log(4\pi A).\end{aligned}$$

Therefore

$$-S(\rho) \leq AJ(\rho) - (2 + \log(4\pi A)) \text{ for } A > 0. \tag{64}$$

Lemma 9 (Log-Sobolev inequality for the qOU semigroup). Let $\mu > \lambda > 0$ and $\zeta > 0$. Then

$$-\zeta D(\rho\|\sigma_{\mu,\lambda}) - \frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \geq \alpha_- J_-(\rho) + \alpha_+ J_+(\rho) + \gamma \mathrm{tr}(\rho\hat{n}) + \delta, \tag{65}$$

for all states ρ, where

$$\begin{aligned}\alpha_- &= \mu^2/2 - 2\pi A\zeta,\\ \alpha_+ &= \lambda^2/2 - 2\pi A\zeta,\\ \gamma &= (\log\nu)\left(\zeta - (\mu^2 - \lambda^2)\right),\\ \delta &= \zeta\,(\log(1-\nu) + 2 + \log(4\pi A)) + \lambda^2\log\nu,\end{aligned}$$

for any $A > 0$. Here $\nu = \frac{\lambda^2}{\mu^2}$, and $\sigma_{\mu,\lambda}$ is the fixed point of $\mathcal{L}_{\mu,\lambda}$. In particular, choosing $\zeta = \mu^2 - \lambda^2$ and $A = \frac{\lambda^2}{4\pi(\mu^2-\lambda^2)}$, we have

$$-\zeta D(\rho\|\sigma_{\mu,\lambda}) - \frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \geq -\zeta\mathbf{n}\log(1 + 1/\mathbf{n}) + \delta, \tag{66}$$

where $\mathbf{n} = \mathrm{tr}(\rho\hat{n})$.

Proof. Combining (60) and (64) leads to

$$S(\rho) \geq -2\pi AJ_-(\rho) - 2\pi AJ_+(\rho) + 2 + \log(4\pi A).$$

Using this inequality in Lemma 8, we obtain the result (65).

For the second part of the statement, Theorem 5 shows that

$$J_-(\rho) \geq J_-(\omega_{\mathbf{n}}) = -2\mathbf{n}\log(1 + 1/\mathbf{n}), \tag{67}$$

where $\omega_{\mathbf{n}}$ is the Gaussian thermal state with mean photon number $\mathbf{n}$. Eq. (66) then follows directly from the first part of the statement with the choice $A = \frac{\lambda^2}{4\pi(\mu^2-\lambda^2)}$, inequality (67), and the choice of $\zeta = \mu^2 - \lambda^2$. □

Considering the same specific qOU process as in Example 1, we obtain the same rate of convergence to its fixed point, but now only for states with low mean photon number instead of large initial entropy as in Example 1.

Example 2. Consider $\mu^2 = 2$ and $\lambda^2 = 1$. Then

$$\frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\sqrt{2},1}}(\rho)\|\sigma_{\sqrt{2},1}) \leq -D(\rho\|\sigma_{\sqrt{2},1}),$$

for any state ρ with $\mathrm{tr}(\rho\hat{n}) \lesssim 0.67$. In comparison, the mean photon number of the fixed point is $\mathrm{tr}(\sigma_{\sqrt{2},1}\hat{n}) = 1$.

Choosing $\mu^2 = 2$ and $\lambda^2 = 1$ in Lemma 9, we obtain

$$-\zeta D(\rho\|\sigma_{\mu,\lambda}) - \frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \geq -\mathbf{n}\log(1 + 1/\mathbf{n}) + 2 - 2\log(2).$$

The right-hand side is monotonically decreasing and for $\mathbf{n} \lesssim 0.67$, it is non-negative. □

VII. DISCUSSION

We have established new information-theoretic inequalities for bosonic quantum systems. Our inequalities are motivated by and directly generalize well-known existing results concerning the sum of two real-valued random variables. They also complement recent results concerning the "addition" of two bosonic quantum states by means of a beamsplitter: we consider a hybrid operation which amounts to a certain way of combining a classical probability distribution on phase space with a quantum state. Mathematically, our work thus makes progress towards a unified view of three types of convolution operations: the convolution of two classical probability density functions, of two quantum Wigner functions, and of a pair consisting of a classical probability density function and a quantum Wigner function. Operationally, our results extend entropic characterizations of classical additive noise and quantum additive noise to the so-called classical noise in bosonic systems. Indeed, the proofs of our main inequalities, which include hybrid versions of the Fisher information and entropy power inequalities, are formally very similar to existing proofs in the fully classical as well as fully quantum settings.

The consideration of the hybrid classical-quantum convolution operation (2) brings additional advantages, however, it allows for the study of infinitesimal Gaussian perturbations to a given quantum state. In contrast, existing fully quantum entropy power inequalities are not directly amenable to such arguments (at least not in an obvious way) since basic uncertainty relations prevent us from making sense of, e.g., a Gaussian state with infinitesimal variance. Mirroring the derivation of the isoperimetric inequality from the Brunn-Minkowski inequality (where a given set is perturbed by adding an infinitesimally small ball), we obtain a quantum isoperimetric inequality relating Fisher information and entropy power. A striking simple consequence of the latter is the statement that the entropy power is a concave function of time for the quantum heat diffusion semigroup: again, this provides a quantum generalization of a fundamental result about the classical heat equation.

Let us conclude by mentioning a few potential directions for future work, as well as some open problems. For concreteness and simplicity, we have considered the simplest non-trivial definition of a hybrid convolution operation defined on multiple modes, as studied in Ref. 11. One may generalize our convolution (2) and the associated results by considering additional (linear) operations along the lines of Ref. 14. From the point of view of information theory, it is also interesting to examine the implications of our results for the capacity of the classical noise channels similarly to Ref. 18.

On a more speculative side, one may wonder whether alternative quantum generalizations of the results considered here exist, especially related to the conjectured photon number inequality by Guha, Erkmen, and Shapiro.[41] These authors (and subsequent work such as Ref. 21) suggest replacing the quantum entropy power $e^{S(\rho)}$ by the arguably more natural expression $g^{-1}(S(\rho))$. This is the mean photon number of a Gaussian state with identical entropy as ρ. It appears that at least for our isoperimetric inequality, such a generalization would require more than a naïve substitution as there is no meaningful lower bound on the product $g^{-1}(S(\rho))J(\rho)$ even for Gaussian states. While this may be considered as additional mathematical justification for our formulation of these inequalities, we believe that progress in this direction could be helpful in resolving, e.g., our conjecture concerning the convergence rate to the fixed point of the quantum Ornstein-Uhlenbeck (qOU) semigroup. For the latter problem, apart from proving our conjecture, it would also be interesting to obtain multi-mode generalizations. This concerns, in particular, the entropy production rates for the qOU semigroup. Here, a resolution of the conjecture of Ref. 24 for the multi-mode attenuator would likely provide important insights.

Finally, we mention some challenging mathematical problems resulting from our work. For example, while our isoperimetric inequality is tight for Gaussian states, necessary and sufficient conditions for equality in most of our statements are currently unknown. Finally, a rigorous discussion of the family of states for which the de Bruijn identity (23) holds, possibly using the framework of Schwartz operators,[36] would be an interesting task for future work.

VIII. REMARK

After posting our paper to the arxiv, we were made aware of concurrent related work by Rouze, Datta, and Pautrat. Their paper has now been made available.[42]

ACKNOWLEDGMENTS

R.K. thanks the organizers of the workshop on Hypercontractivity and Log-Sobolev Inequalities in Banff and R. F. Werner for interesting discussions. R.K. and S.H. are supported by the Technische Universität München—Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under Grant Agreement No. 291763. They acknowledge additional support by DFG project no. K05430/1-1. A.V. is supported by the John Templeton Foundation Grant No. 48322, and by the Basque Government through the BERC 2014-2017 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323.

APPENDIX A: A LOG-SOBOLEV INEQUALITY AND THE CLASSICAL ORNSTEIN-UHLENBECK PROCESS

In this appendix, we discuss known classical results for the reader's convenience: we briefly review the relationship between the isoperimetric inequality for classical Fisher information, the Log-Sobolev inequality, and the rate of convergence to the fixed point for the classical Ornstein-Uhlenbeck semigroup. These arguments were given by Carlen[25] for a particular element of the two-parameter family of Ornstein-Uhlenbeck processes. Here we specialize to real-valued random variables, but allow arbitrary parameters in order to illustrate the parallels to the qOU semigroup. We also explicitly discuss the convergence to the fixed point, which is only implicit in Ref. 25 but appears to be folklore.

Let f_0 be a probability density on $\mathbb{R}$ of a real-valued random variable X. The classical Ornstein-Uhlenbeck (cOU) process for $\theta > 0$ and $\sigma > 0$ is given by the Fokker-Planck equation

$$\frac{\partial f}{\partial t} = \theta \frac{\partial}{\partial x}[xf] + \frac{\sigma^2}{2}\frac{\partial^2 f}{\partial x^2} = \mathcal{A}_{\theta,\sigma^2}(f).$$

(Carlen considers the case where $\theta = 1$ and $\sigma^2 = 1/\pi$.) The solution to this equation can be written (in terms of random variables) as

$$X_t = e^{-\theta t}X_0 + \frac{\sigma}{\sqrt{2\theta}}\sqrt{1 - e^{-2\theta t}}Z, \tag{A1}$$

where $Z \sim \mathcal{N}(0,1)$ is an independent centered Gaussian random variable with unit variance. The stationary solution f_∞ therefore is a centered Gaussian distribution with variance $\sigma^2/(2\theta)$. In particular, (A1) implies that the second moments satisfy

$$\mathbb{E}[X_t^2] = e^{-2\theta t}\mathbb{E}[X_0^2] + (1 - e^{-2\theta t})\frac{\sigma^2}{2\theta}. \tag{A2}$$

We use that the relative entropy between a random variable X and a centered normal variable $Z_{\sigma^2} \sim \mathcal{N}(0,\sigma^2)$ is given by $D(X\|Z_{\sigma^2}) = -H(X) + \frac{1}{2}\log 2\pi\sigma^2 + \frac{1}{2\sigma^2}\mathbb{E}[X^2]$ as can be verified easily. In particular, the relative entropy between the solution at time t and the fixed point $Z_{\sigma^2/(2\theta)}$ is given by

$$D(X_t\|Z_{\sigma^2/(2\theta)}) = -H(X_t) + \frac{1}{2}\log \pi\sigma^2/\theta + \frac{\theta}{\sigma^2}\mathbb{E}[X_t^2]. \tag{A3}$$

Furthermore, we have the following de Bruijn-type identity:

Lemma 10. Let $\{X_t\}_{t\geq 0}$ be of the form (A1), i.e., a solution to the cOU process. Let $J(X) = \int \frac{|\frac{\partial f(x)}{\partial x}|^2}{f(x)}dx$ denote the Fisher information of a random variable X with distribution function f. Then

$$\frac{d}{dt}\Big|_{t=0} H(X_t) = \frac{\sigma^2}{2}J(X_0) - \theta. \tag{A4}$$

Observe that the second summand essentially stems from the fact that entropies transform very simply under rescaling of random variables, namely

$$H(e^{-\theta t}X) = H(X) - \theta t. \tag{A5}$$

Eq. (A5) significantly simplifies the analysis. Such a property does not hold in the quantum case: as a consequence, we do not have a simple expression in terms of $J(X_0)$ only.

Proof. The derivative of the entropy along a semigroup with generator $\mathcal{A}$ is given by the expression $\frac{d}{dt}\Big|_{t=0} H(e^{t\mathcal{A}}(f)) = -\int \mathcal{A}(f)(x)\log f(x)dx$. Using this fact gives

$$\frac{d}{dt}\Big|_{t=0} H(X_t) = -\theta \int \left(\frac{\partial}{\partial x}[xf(x)]\right) \log f(x)dx - \frac{\sigma^2}{2}\int \frac{\partial^2 f(x)}{\partial x^2}\log f(x)dx.$$

Denoting $f'(x) = \frac{\partial}{\partial x}f(x)$, we obtain

$$\int (xf)'\log f dx = -\int (xf)f'/f dx = -\int xf'dx = \int f dx = 1,$$

where we have used partial integration and the fact that boundary terms vanish twice. Similarly, we have

$$\int f''\log f dx = -\int (f')^2/f dx,$$

by partial integration. Combining these statements gives the claim. □

Following Ref. 25, we can write the isoperimetric inequality $1/N(X) \leq \frac{J(X)}{2\pi e}$ as

$$\begin{aligned} -H(X) &\leq \frac{1}{2}\log\left(\frac{J(X)}{2\pi e}\right) \\ &= \frac{1}{2}\log\left(\frac{J(X)}{2\pi}\right) - \frac{1}{2}. \end{aligned}$$

In particular, using $\log x \leq x - 1$, we get

$$\begin{aligned} \log\left(\frac{J(X)}{2\pi e}\right) &= \log\left(\frac{J(X)\cdot A}{2\pi e A}\right) \\ &= \log\frac{1}{2\pi e A} + \log(AJ(X)) \\ &\leq AJ(X) - 2 - \log 2\pi A, \end{aligned} \tag{A6}$$

for any $A > 0$. Using inequality (A6), it is straightforward to show the following.

Theorem 9 (Fast convergence of the cOU semigroup[25]). *Let $\{X_t\}_{t\geq 0}$ be of the form (A1), i.e., a solution to the cOU process with parameters $\theta > 0, \sigma > 0$. Then*

$$\frac{d}{dt}\Big|_{t=0} D(X_t \| Z_{\sigma^2/(2\theta)}) \leq -2\theta D(X_0 \| Z_{\sigma^2/(2\theta)}).$$

Note that because we are considering a semigroup, this result immediately implies that

$$D(X_t \| Z_{\sigma^2/(2\theta)}) \leq e^{-2\theta t} D(X_0 \| Z_{\sigma^2/(2\theta)}), \qquad \text{for all } t \geq 0.$$

Also, this result is tight: if $X_0 \sim \mathcal{N}(0, \sigma^2_{X_0})$ is a centered Gaussian random variable with variance $\sigma^2_{X_0}$, then (A1) implies $X_t \sim \mathcal{N}(0, \sigma^2_t)$ where the variance of X_t is

$$\sigma^2_t = e^{-2\theta t}\sigma^2_{X_0} + \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}).$$

Inserting into (A3), using $H(X_t) = \frac{1}{2}(1 + \log(2\pi\sigma^2_t))$ yields

$$D(X_t \| Z_{\sigma^2/(2\theta)}) = -\frac{1}{2}(1 + \log(2\pi\sigma^2_t)) + \frac{1}{2}\log \pi\sigma^2/\theta + \frac{\theta}{\sigma^2}\sigma^2_t.$$

In particular, we obtain

$$\lim_{\sigma^2_{X_0}\to\infty}\left(\frac{d}{dt}\Big|_{t=0}D(X_t\|Z_{\sigma^2/(2\theta)})\right)/D(X_0\|Z_{\sigma^2/(2\theta)}) = -2\theta.$$

Proof. According to Eqs. (A2), (A3), and (A4), we have

$$\frac{d}{dt}\Big|_{t=0}D(X_t\|Z_{\sigma^2/(2\theta)}) = -\frac{\sigma^2}{2}J(X_0) - \frac{2\theta^2}{\sigma^2}\mathbb{E}[X^2] + 2\theta. \tag{A7}$$

Combining (A6) with (A3) yields

$$D(X_0\|Z_{\sigma^2/(2\theta)}) \le \frac{AJ(X)}{2} - (1 + \frac{1}{2}\log 2\pi A) + \frac{1}{2}\log \pi\sigma^2/\theta + \theta/\sigma^2\mathbb{E}[X_0^2]. \tag{A8}$$

Combining (A8) with (A7) yields

$$\begin{aligned}-2\theta D(X_0\|Z_{\sigma^2/(2\theta)}) - \frac{d}{dt}\Big|_{t=0}D(X_t\|Z_{\sigma^2/(2\theta)}) \ge \left(-\theta A + \frac{\sigma^2}{2}\right)J(X_0)\\ + 2\theta\left[\left(1 + \frac{1}{2}\log 2\pi A - \frac{1}{2}\log\pi\sigma^2/\theta\right) - 1\right].\end{aligned}$$

The claim then follows by choosing $A = \frac{\sigma^2}{2\theta}$. □

APPENDIX B: TIGHTNESS OF THE QUANTUM FISHER INFORMATION ISOPERIMETRIC INEQUALITY

Consider a one-mode Gaussian thermal state $\omega_{\mathbf{n}}$ with mean photon number $\mathbf{n} > 0$. Its entropy is

$$S(\omega_{\mathbf{n}}) = g(\mathbf{n}) = (\mathbf{n}+1)\log(\mathbf{n}+1) - \mathbf{n}\log\mathbf{n}. \tag{B1}$$

Under the diffusion semigroup, the state $\omega_{\mathbf{n}}$ evolves as

$$e^{t\mathcal{L}_{\text{heat}}}(\omega_{\mathbf{n}}) = \omega_{\mathbf{n}_t}, \qquad \text{where} \qquad \mathbf{n}_t = \mathbf{n} + 2\pi t.$$

In particular, by the de Bruijn identity

$$J(\omega_{\mathbf{n}}) = 2\frac{d}{dt}S(e^{t\mathcal{L}_{\text{heat}}}(\omega_{\mathbf{n}}))\Big|_{t=0} = 2g'(\mathbf{n})\mathbf{n}_t'\Big|_{t=0} = 4\pi\log\left(\frac{\mathbf{n}+1}{\mathbf{n}}\right). \tag{B2}$$

Also,

$$J(e^{t\mathcal{L}_{\text{heat}}}(\omega_{\mathbf{n}})) = 4\pi\log\left(1 + \frac{1}{\mathbf{n}+2\pi t}\right).$$

Calculating the right-hand side of the quantum Fisher information isoperimetric inequality (39), we obtain

$$\frac{d}{dt}\Big|_{t=0}\left[\frac{1}{2}J(e^{t\mathcal{L}_{\text{heat}}}(\omega_{\mathbf{n}}))\right]^{-1} = \frac{1}{\mathbf{n}(\mathbf{n}+1)}\log^{-2}\left(1+\frac{1}{\mathbf{n}}\right) \to 1, \qquad \text{as } \mathbf{n}\to\infty.$$

APPENDIX C: TIGHTNESS OF THE ISOPERIMETRIC INEQUALITY

Consider a one-mode Gaussian thermal state $\omega_{\mathbf{n}}$ with mean photon number $\mathbf{n}$. From (B2) we know that

$$J(\omega_{\mathbf{n}}) = 4\pi\log\left(\frac{\mathbf{n}+1}{\mathbf{n}}\right),$$

and that the entropy power of $\omega_{\mathbf{n}}$ is given by (cf. (B1))

$$N(\omega_{\mathbf{n}}) = \exp(S(\omega_{\mathbf{n}})/1) = \frac{(\mathbf{n}+1)^{\mathbf{n}+1}}{\mathbf{n}^{\mathbf{n}}}.$$

Combining these two expressions, we see that the left-hand side of (50) is

$$J(\omega_{\mathbf{n}})N(\omega_{\mathbf{n}}) = 4\pi\left(\frac{\mathbf{n}+1}{\mathbf{n}}\right)^{\mathbf{n}} \log\left(\frac{\mathbf{n}+1}{\mathbf{n}}\right)^{\mathbf{n}+1} \to 4\pi e, \qquad \text{for } \mathbf{n}\to\infty.$$

APPENDIX D: ON THE CONVERGENCE RATE FOR GAUSSIAN INITIAL STATES

In this section, we show that for a one-mode Gaussian state ρ the following inequality holds:

$$\frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \leq -\zeta D(\rho\|\sigma_{\mu,\lambda}) \qquad \text{with } \zeta = \mu^2 - \lambda^2 > 0. \tag{D1}$$

Furthermore, this statement is optimal: for any $\epsilon > 0$ there exists a Gaussian state ρ such that

$$\frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \geq -(\zeta+\epsilon) D(\rho\|\sigma_{\mu,\lambda}).$$

First, we note that the entropy of a Gaussian state does not depend on its first moments. Hence it follows that for a Gaussian state the right-hand side of (61) does not depend on its first moments, and it is suffices to consider centered Gaussian states.

Calculating the right-hand side of (61), we focus on calculating $J_\pm(\rho)$. Recall that the covariance matrix M of a centered state ρ is defined as $M_{jk} = \mathrm{tr}(\rho\{R_j, R_k\})$.

Lemma 11. Let ρ be a one-mode centered Gaussian state with mean-photon number $\mathbf{n}$ and covariance matrix given by $M = \kappa S^T S$, where $\kappa = 2\mathbf{n}+1$ and $S = O_1 \left(\begin{smallmatrix} z & 0 \\ 0 & 1/z \end{smallmatrix}\right) O_2^T$ with $O_i \in Sp(2) \cap O(2)$ and $z \geq 1$. Denote

$$J_\pm(\rho) = 2\frac{d}{dt} S\left(e^{t\mathcal{L}_\pm}(\rho)\right)\Big|_{t=0}.$$

Then we have

$$J_\pm(\rho) = \left(\frac{1}{2}\left(1/z^2 + z^2\right) \pm \kappa\right)\left(\log\left(\frac{\kappa+1}{2}\right) - \log\left(\frac{\kappa-1}{2}\right)\right). \tag{D2}$$

Proof. The covariance matrix of the time-evolved state $e^{t\mathcal{L}_\pm}(\rho)$ is

$$M_\pm(t) = c_1^\pm(t) M(0) + c_2^\pm(t)\mathrm{id},$$

where

$$c_1^-(t) = e^{-t} \qquad \text{and} \qquad c_2^-(t) = 1 - e^{-t}, \tag{D3}$$

$$c_1^+(t) = e^{t} \qquad \text{and} \qquad c_2^+(t) = e^{t} - 1. \tag{D4}$$

Therefore writing $S[M]$ for the entropy of a Gaussian state with covariance matrix M, we have

$$\begin{aligned} S\left[M_\pm(t)\right] &= S\left[c_1^\pm(t)\kappa S^T S + c_2^\pm(t)\mathrm{id}\right] = S\left[c_1^\pm(t)\kappa O_2^T K^T O_1^T O_1 K O_2 + c_2^\pm(t) O_2^T O_2\right] \\ &= S\left[c_1^\pm(t)\kappa K^T O_1^T O_1 K + c_2^\pm(t)\mathrm{id}\right] \\ &= S\left[\begin{pmatrix} c_1^\pm(t)\kappa z^2 + c_2^\pm(t) & 0 \\ 0 & c_1^\pm(t)\kappa/z^2 + c_2^\pm(t)\end{pmatrix}\right]. \end{aligned} \tag{D5}$$

The symplectic eigenvalue of the matrix argument in (D5) is the square root of its determinant:

$$\kappa_\pm(t) = \sqrt{\left(c_1^\pm(t)\kappa + c_2^\pm(t) z^2\right)\left(c_1^\pm(t)\kappa + c_2^\pm(t)/z^2\right)}. \tag{D6}$$

The entropy of the time evolved state is

$$S(e^{t\mathcal{L}_\pm}(\rho)) = g(\mathbf{n}(\kappa_\pm(t))),$$

where $g(x) = (x+1)\log(x+1) - x\log x$ and $\mathbf{n}(\kappa) = \frac{1}{2}(\kappa - 1)$. By the chain rule, we have that

$$\frac{d}{dt} S\left(e^{t\mathcal{L}_\pm}(\rho)\right) = \frac{1}{2} g'(\mathbf{n}(\kappa_\pm(t)))\,\kappa_\pm{}'(t).$$

Since $g'(x) = \log(x+1) - \log x$, we only need to find $\kappa_\pm{}'(t)$.

Combining (D6) and (D3), we obtain

$$\begin{aligned}\kappa_-(t) &= \sqrt{\left(e^{-t}\kappa + (1-e^{-t})\,z^2\right)\left(e^{-t}\kappa + (1-e^{-t})\,/z^2\right)}\\ &= \kappa + t\left(\frac{1}{2}\left(z^2 + 1/z^2\right) - \kappa\right) + O\left(t^2\right).\end{aligned}$$

Therefore $\kappa_-{}'(t)|_{t=0} = \frac{1}{2}\left(1/z^2 + z^2\right) - \kappa$ and finally

$$J_-(\rho) = \left(\frac{1}{2}\left(1/z^2 + z^2\right) - \kappa\right)\log\frac{\kappa+1}{\kappa-1}.$$

Similarly, using (D6) together with (D4), we obtain

$$\begin{aligned}\kappa_+(t) &= \sqrt{\left(e^{t}\kappa + (e^{t}-1)\,z^2\right)\left(e^{t}\kappa + (e^{t}-1)\,/z^2\right)}\\ &= \kappa + t\left(\frac{1}{2}\left(z^2 + 1/z^2\right) + \kappa\right) + O\left(t^2\right).\end{aligned}$$

Thus $\kappa_+{}'(t)|_{t=0} = \frac{1}{2}\left(1/z^2 + z^2\right) + \kappa$ and

$$J_+(\rho) = \left(\frac{1}{2}\left(1/z^2 + z^2\right) + \kappa\right)\log\frac{\kappa+1}{\kappa-1}.$$

□

From Lemma 8 it is clear that we are interested in minimizing $J_\pm(\rho)$. Both expressions in (D2) have a minimum at $z = 1$. Therefore, with $\zeta = \mu^2 - \lambda^2$ and $\mathbf{n} = \frac{\kappa-1}{2}$, from Lemma 8 we obtain

$$\begin{aligned}-\zeta D(\rho\|\sigma_{\mu,\lambda}) - \frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \geq{}& \mu^2\log(\mathbf{n}+1) - \lambda^2\log\mathbf{n} + \lambda^2\log\lambda^2\\ &- \mu^2\log\mu^2 + (\mu^2-\lambda^2)\log(\mu^2-\lambda^2)\\ =:{}& h_{\mu,\lambda}(\mathbf{n}).\end{aligned}$$

For fixed $\mu^2 > \lambda^2$, the function $h_{\mu,\lambda}$ satisfies $\lim_{\mathbf{n}\to 0} h_{\mu,\lambda}(\mathbf{n}) = \lim_{\mathbf{n}\to\infty} h_{\mu,\lambda}(\mathbf{n}) = \infty$. Since $h_{\mu,\lambda}$ is differentiable (in fact smooth) for $\mathbf{n} > 0$, we can find the global minimum by finding the zeros of the derivative

$$\frac{d}{d\mathbf{n}} h_{\mu,\lambda}(\mathbf{n}) = \frac{\mu^2}{\mathbf{n}+1} - \frac{\lambda^2}{\mathbf{n}} = 0.$$

The only solution is $\mathbf{n} = \frac{\lambda^2}{\mu^2-\lambda^2}$, and since

$$\frac{d^2}{d\mathbf{n}^2}\Big|_{\mathbf{n}=\frac{\lambda^2}{\mu^2-\lambda^2}} h_{\mu,\lambda}(\mathbf{n}) = (\mu^2-\lambda^2)\left(\frac{1}{\lambda^2} - \frac{1}{\mu^2}\right) > 0,$$

it is the minimum. The value of the minimum is $h_{\mu,\lambda}\left(\frac{\lambda^2}{\mu^2-\lambda^2}\right) = 0$, hence it follows that $h_{\mu,\lambda}(\mathbf{n}) \geq 0$ for all $\mathbf{n} > 0$ and we have

$$\frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\rho)\|\sigma_{\mu,\lambda}) \leq -\zeta D(\rho\|\sigma_{\mu,\lambda}).$$

Moreover, the last inequality becomes equality for $\rho = \sigma_{\mu,\lambda} = \omega_{\mathbf{n}_\infty}$, with $\mathbf{n}_\infty = \frac{\lambda^2}{\mu^2-\lambda^2}$.

It remains to show that $\zeta = \mu^2 - \lambda^2$ is optimal. Let $\epsilon > 0$ and $\zeta' = \zeta + \epsilon$. Then for the Gaussian thermal state $\omega_{\mathbf{n}}$ we have

$$\begin{aligned}-\zeta' D(\omega_{\mathbf{n}}\|\sigma_{\mu,\lambda}) - \frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\omega_{\mathbf{n}})\|\sigma_{\mu,\lambda}) &= h_{\mu,\lambda}(\mathbf{n}) + \epsilon(\mathbf{n}+1)\log(\mathbf{n}+1) - \epsilon\,\mathbf{n}\log\mathbf{n} \\ &\quad + \epsilon\,\mathbf{n}\log\left(\frac{\lambda^2}{\mu^2}\right) + \epsilon\log\left(1-\frac{\lambda^2}{\mu^2}\right) \\ &= \log\left(\left(\frac{\mathbf{n}+1}{\mathbf{n}}\right)^{\mu^2+\epsilon(\mathbf{n}+1)} \mathbf{n}^{\mu^2-\lambda^2+\epsilon}\left(\frac{\lambda^2}{\mu^2}\right)^{\epsilon\mathbf{n}}\right) \\ &\quad + \mathrm{c}(\mu,\lambda) \;\to -\infty \qquad \text{for } \mathbf{n}\to\infty,\end{aligned}$$

where $\mathrm{c}(\mu,\lambda) = \log\left((\mu^2)^{-\mu^2-\epsilon}(\lambda^2)^{\lambda^2}(\mu^2-\lambda^2)^{\mu^2-\lambda^2+\epsilon}\right)$. Therefore, for any $\epsilon > 0$, there exists $\mathbf{n}$ such that

$$\frac{d}{dt}\Big|_{t=0} D(e^{t\mathcal{L}_{\mu,\lambda}}(\omega_{\mathbf{n}})\|\sigma_{\mu,\lambda}) > -(\zeta+\epsilon)D(\omega_{\mathbf{n}}\|\sigma_{\mu,\lambda}).$$

This shows that the constant $\zeta = \mu^2 - \lambda^2$ is optimal in the inequality (D1).

[1] W. Beckner, "Inequalities in Fourier analysis," Ann. Math. **102**(1), 159–182 (1975).

[2] H. Brascamp and E. Lieb, "Best constants in Young's inequality, its converse and its generalization to more than three functions," Adv. Math. **20**, 151–172 (1976).

[3] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," Inf. Control **2**(2), 101–112 (1959).

[4] C. E. Shannon, "A mathematical theory of communication," Bell Syst. Tech. J. **27**, 623–656, (1948).

[5] N. Blachman, "The convolution inequality for entropy powers," IEEE Trans. Inf. Theory **11**, 267–271 (1965).

[6] A. R. Barron, "Entropy and the central limit theorem," Ann. Probab. **14**(1), 336–342 (1986).

[7] M. H. M. Costa and T. M. Cover, "On the similarity of the entropy power inequality and the Brunn–Minkowski inequality," IEEE Trans. Inf. Theory **30**(6), 837–839 (1984).

[8] S. J. Szarek and D. Voiculescu, *Shannon's Entropy Power Inequality Via Restricted Minkowski Sums*, Lecture Notes in Mathematics Vol. 1745 (Springer Berlin Heidelberg, 2000), pp. 257–262.

[9] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," IEEE Trans. Inf. Theory **37**, 1501–1518 (1991).

[10] F. Barthe, "The Brunn–Minkowski theorem and related geometric and functional inequalities," in *Proceedings of the International Congress of Mathematicians* (European Mathematical Society, Madrid, 2006), pp. 1529–1546.

[11] R. Werner, "Quantum harmonic analysis on phase space," J. Math. Phys. **25**(5), 1404–1411 (1984).

[12] E. A. Carlen and E. H. Lieb, "Brascamp-Lieb inequalities for non-commutative integration," Doc. Math. **13**, 553–584 (2008).

[13] R. König and G. Smith, "The entropy power inequality for quantum systems," IEEE Trans. Inf. Theory **60**, 1536–1548 (2014).

[14] G. De Palma, A. Mari, S. Lloyd, and V. Giovannetti, "Multimode quantum entropy power inequality," Phys. Rev. A **91**, 032320 (2015).

[15] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," Commun. Math. Phys. **62**(1), 35–41 (1978).

[16] S. Verdu and D. Guo, "A simple proof of the entropy-power inequality," IEEE Trans. Inf. Theory **52**, 2165–2166 (2006).

[17] R. König, "The conditional entropy power inequality for Gaussian quantum states," J. Math. Phys. **56**, 022201 (2015).

[18] R. König and G. Smith, "Limits on classical communication from quantum entropy power inequalities," Nat. Photonics **7**, 142–146 (2012).

[19] K. Audenaert, N. Datta, and M. Ozols, "Entropy power inequalities for qudits," J. Math. Phys. **57**, 052202 (2016).

[20] E. A. Carlen, E. H. Lieb, and M. Loss, "On a quantum entropy power inequality of Audenaert, Datta and Ozols," J. Math. Phys. **57**, 062203 (2016).

[21] S. Guha, J. H. Shapiro, and R. G.-P. Sanchez, "Thinning, photonic beam splitting, and a general discrete entropy power inequality," in *2016 IEEE International Symposium on Information Theory (ISIT) (IEEE, Barcelona, 2016)*, pp. 705–709.

[22] M. Costa, "A new entropy power inequality," IEEE Trans. Inf. Theory **31**, 751–760 (1985).

[23] C. Villani, "A short proof of the concavity of entropy power," IEEE Trans. Inf. Theory **46**, 1695–1696 (2000).

[24] G. De Palma, D. Trevisan, and V. Giovannetti, "Gaussian states minimize the output entropy of the one-mode quantum attenuator," IEEE Trans. Inf. Theory **63**(1), 728–737 (2017).

[25] E. A. Carlen, "Superadditivity of Fisher's information and logarithmic Sobolev inequalities," J. Funct. Anal. **101**(1), 194–211 (1991).

[26] R. Carbone and E. Sasso, "Hypercontractivity for a quantum Ornstein–Uhlenbeck semigroup," Probab. Theory Relat. Fields **140**(3-4), 505–522 (2008).

[27] R. Olkiewicz and B. Zegarlinski, "Hypercontractivity in noncommutative lpspaces," J. Funct. Anal. **161**(1), 246–285 (1999).

[28] M. J. Kastoryano and K. Temme, "Quantum logarithmic Sobolev inequalities and rapid mixing," J. Math. Phys. **54**(5), 052202 (2013).

[29] A. Müller-Hermes, D. S. França, and M. M. Wolf, "Relative entropy convergence for depolarizing channels," J. Math. Phys. **57**(2), 022202 (2016).

[30] F. Buscemi, S. Das, and M. M. Wilde, "Approximate reversibility in the context of entropy gain, information gain, and complete positivity," Phys. Rev. A **93**(6), 062314 (2016).

[31] M. Jabbour, R. Garcia-Patron, and N. Cerf, "Majorization preservation of gaussian bosonic channels," New J. Phys. **18**, 073047 (2016).

[32] G. De Palma, D. Trevisan, and V. Giovannetti, "Passive states optimize the output of bosonic Gaussian quantum channels," IEEE Trans. Inf. Theory **62**(5), 2895–2906 (2016).

[33] R. Zamir, "A proof of the Fisher information inequality via a data processing argument," IEEE Trans. Inf. Theory **44**, 1246–1250 (1998).

[34] A. Dembo, "Simple proof of the concavity of the entropy power with respect to added Gaussian noise," IEEE Trans. Inf. Theory **35**(4), 887–888 (1989).

[35] D. Petz and C. Ghinea, "Introduction to quantum Fisher information," QP-PQ: Quantum Probab. White Noise Anal. **27**, 261–281 (2011).

[36] M. Keyl, J. Kiukas, and R. F. Werner, "Schwartz operators," Rev. Math. Phys. **28**, 1630001 (2016).

[37] J. Eisert and M. M. Wolf, "Gaussian quantum channels," in *Quantum Information with Continuous Variables of Atoms and Light* (Imperial College Press, London, 2007), pp. 23–42.

[38] E. H. Lieb, "Convex trace functions and the Wigner-Yanase-Dyson conjecture," Adv. Math. **11**(3), 267–288 (1973).

[39] R. König and G. Smith, "Erratum: 'The entropy power inequality for quantum systems' [IEEE Trans. Inf. Theory **60**, 1536–1548 (2014)]," IEEE Trans. Inf. Theory **62**(7), 4358–4359 (2016).

[40] T. M. Cover and J. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications and Signal Processing (Wiley-Interscience, 2006).

[41] S. Guha, B. Erkmen, and J. Shapiro, "The entropy photon-number inequality and its consequences," in *Information Theory and Applications Workshop* (IEEE, 2008), pp. 128–130.

[42] C. Rouze, N. Datta, and Y. Pautrat, "Contractivity properties of a quantum diffusion semigroup," J. Math. Phys. **58**, 012205 (2017).

[43] Here we follow the notation of Refs. 27–29. In contrast, this is denoted α_1 in Ref. 26.

[44] In more general terms, (14) is the trace of the Fisher information matrix

$$\left(\frac{\partial^2}{\partial\theta_i\partial\theta_j}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} D\left(f^{(\boldsymbol{\theta}_0)}\middle\| f^{(\boldsymbol{\theta})}\right)\right)_{i,j=1}^{m}$$

corresponding to (15). The Fisher information matrix provides a lower bound on the variance of an unbiased estimate of the unknown parameter $\boldsymbol{\theta}$ according to the Cramer-Rao inequality. For our purposes, it is sufficient to consider the family (15).

[45] More precisely, Ref. 24, Theorem 6 states that the entropy rate of a state with finite support is lower bounded by that of a Gaussian state with the same entropy. Since we are interested in a statement applicable to arbitrary states, we use their Theorem 24 instead: the latter gives an analogous statement for the corresponding classical process but without a finiteness assumption.

A.2 Coherent state coding approaches the capacity of non-Gaussian bosonic channels

Coherent state coding approaches the capacity of non-Gaussian bosonic channels

Stefan Huber and Robert König

The additivity problem asks if the use of entanglement can boost the information-carrying capacity of a given channel beyond what is achievable by coding with simple product states only. This has recently been shown not to be the case for phase-insensitive one-mode Gaussian channels, but remains unresolved in general. Here we consider two general classes of bosonic noise channels, which include phase-insensitive Gaussian channels as special cases: these are beamsplitters with general, potentially non-Gaussian environment states and classical noise channels with general probabilistic noise. We show that additivity violations, if existent, are rather minor for all these channels: the maximal gain in classical capacity is bounded by a constant independent of the input energy. Our proof shows that coding by simple classical modulation of coherent states is close to optimal.

A.2.1 Main Results

We consider the family of channels defined by beamsplitters of transmissivity $\lambda \in [0,1]$, defined in Eqs. (4.9) and (4.10). Let us denote the channel defined by interaction of a one-mode system with an environment in a fixed state σ_E via a beamsplitter of transmissivity λ by

$$\mathcal{E}_{\lambda,\sigma_E}(\rho) := \operatorname{tr}_E\left(U_\lambda(\rho\otimes\sigma_E)U_\lambda^\dagger\right) \ .$$

Furthermore, we consider general classical noise channels as defined in Eq. (4.11) and denote this family by

$$\mathcal{F}_{t,f}(\rho) := \int f(\xi)D(\sqrt{2\pi t}\xi)\rho D(\sqrt{2\pi t}\xi)^\dagger \mathrm{d}^2\xi \ ,$$

for a probability density function $f : \mathbb{R}^2 \to \mathbb{R}$ of the noise. We then have the following main theorem:

Theorem A.2.1 (Main Theorem)**.** *The maximal degree of non-additivity of the classical capacity of $\mathcal{E}_{\lambda,\sigma_E}$ and $\mathcal{F}_{t,f}$ is bounded as*

$$\begin{aligned} C_N(\mathcal{E}_{\lambda,\sigma_\mathrm{E}}) - \chi_N(\mathcal{E}_{\lambda,\sigma_\mathrm{E}}) &\le 2g\left((1-\lambda)N_\mathrm{E}\right) - g\left((1-\lambda)N_\mathrm{E}^{\mathrm{ep}}\right) - \log\left(\lambda + (1-\lambda)e^{S(\sigma_\mathrm{E})}\right) \ , \\ C_N(\mathcal{F}_{t,f}) - \chi_N(\mathcal{F}_{t,f}) &\le 2g(\pi t\mathrm{E}(f)) - \log\left(1 + te^{H(f)}\right) \ , \end{aligned}$$

independently of the input energy N, where $N_E = \operatorname{tr}(a^\dagger a\sigma_E)$ is the mean photon number of σ_E, $N_\mathrm{E}^{\mathrm{ep}} = g^{-1}\left(S(\sigma_\mathrm{E})\right)$ is the mean photon number of a thermal state with the same entropy as σ_E, and $E(f) = \sum_{i=1}^2 \int \xi_i^2 f(\xi)\mathrm{d}^2\xi$ is the energy of f. The function $g(N) := (N+1)\log(N+1) - N\log(N)$ is the entropy of the thermal state with mean photon number N. For both channels, coherent states modulated with a Gaussian distribution achieve a rate which differs from the capacity by at most the rhs of these bounds.

This theorem gives an upper bound on the maximal additivity violation of the classical capacity of these channels. This upper bound is independent of the input energy N and hence additivity violations are at most constant, whereas the classical capacity goes to infinity as N increases. It is important to note that the families of channels defined by $\mathcal{E}_{\lambda,\sigma_E}$ and $\mathcal{F}_{t,f}$ include a wide variety of non-Gaussian channels (in the cases where the environment state σ_E respectively the noise function f is not Gaussian). Non-Gaussian channels are rarely considered in literature and our result is the first statement about additivity violations of non-Gaussian bosonic channels.

The proof of Theorem A.2.1 relies on the entropy power inequalities from Theorems .5.4.1 and 5.5.1. The results can be slightly improved by assuming that the Entropy Photon-Number Inequality conjecture (Conjecture 5.4.2) is true, leading to the following theorem.

Theorem A.2.2. *Assuming the EPNI conjecture 5.4.2 holds, we have*

$$C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) - \chi_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leq 2\Big[g\left((1-\lambda)N_{\mathrm{E}}\right) - g\left((1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\right)\Big] \,.$$

$$C_N(\mathcal{F}_{t,f}) - \chi_N(\mathcal{F}_{t,f}) \leq 2\left[g(\pi t \mathrm{E}(f)) - g\left(\frac{t}{e}e^{H(f)}\right)\right] \,.$$

Furthermore, coherent state modulation with a Gaussian distribution achieves a rate which differs from the capacity by at most the rhs of these bounds.

It is important to note that the statement of Theorem A.2.2, while making the bounds aesthetically more pleasing, is not very different from the statement of Theorem A.2.1: In both cases, the corresponding upper and lower bounds on the capacity are separated by a constant independent of the input energy. However, assuming the EPNI, in the case where σ_E is a thermal state or where f is a unit-variance centered normal distribution, the additivity violation disappears and we recover the landmark result on classical capacities of phase-insensitive Gaussian bosonic channels [8].

Theorem A.2.1 is directly implied by the following Lemmas, which give explicit upper and lower bounds on the classical capacity of our families of channels.

Lemma A.2.3. *The classical capacity of the single-mode attenuation channel $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}$ satisfies*

$$C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \geq g\left(\lambda N + (1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\right) - g\left((1-\lambda)N_{\mathrm{E}}\right) \,, \tag{A.2}$$

$$C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leq g\left(\lambda N + (1-\lambda)N_{\mathrm{E}}\right) - \log\left(\lambda + (1-\lambda)e^{S(\sigma_{\mathrm{E}})}\right) \,.$$

The lower bound (A.2) *is achievable with a coherent state ensemble. The difference between this upper and lower bound is bounded by*

$$\Delta(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leq 2g\left((1-\lambda)N_{\mathrm{E}}\right) - g\left((1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\right) - \log\left(\lambda + (1-\lambda)e^{S(\sigma_{\mathrm{E}})}\right) \,,$$

independently of the input photon number N.

Lemma A.2.4. *For the classical capacity of the single-mode classical noise channel $\mathcal{F}_{t,f}$, we have the bounds*

$$C_N(\mathcal{F}_{t,f}) \geq \log\left(e^{g(N)} + te^{H(f)}\right) - g\left(\pi t \mathrm{E}(f)\right) \,, \tag{A.3}$$

$$C_N(\mathcal{F}_{t,f}) \leq g\left(N + \pi t \mathrm{E}(f)\right) - \log\left(1 + te^{H(f)}\right) \,.$$

The lower bound (A.3) *is achievable with a coherent state ensemble. The difference between this upper and lower bound is bounded by*

$$\Delta(\mathcal{F}_{t,f}) \leq 2g\left(\pi t \mathrm{E}(f)\right) - \log\left(1 + te^{H(f)}\right) \,,$$

independently of the input photon number N.

Going through the proofs of above Lemmas and replacing the applications of the EPI by the conjectured EPNI, it is easy to formulate analogous statements assuming the validity of the Entropy Photon-Number Inequality. This implies Theorem A.2.2.

A.2.2 Individual Contribution

I am the principal author of this article. The idea for this work came into being after the publication of Article III, during a discussion about possible related projects with Robert König. This project was inspired by previous work by König and Smith which applied the entropy power inequality to the classical capacity of thermal noise channels [11, 79]. I proved all the results of the paper, and with the exception of the Introduction and the first half of Section 2, I wrote all sections.

Permission to include:

Stefan Huber and Robert König
Coherent state coding approaches the capacity of non-Gaussian bosonic channels.
J. Phys. A: Math. Theor. 51, 184001 (2018).

Assignment of copyright and publication agreement

IOP Publishing Limited (“IOP”) agrees to publish:

Manuscript Title: Coherent state coding approaches the capacity of non-Gaussian bosonic channels (the “Article”) written by

Names of all authors: Huber, Stefan; König, Robert (“the Named Authors”) in the following journal Journal of Physics A: Mathematical and Theoretical (“the Journal”)

Name of copyright owner(s) (if not the Named Author(s) – see Important Information above):
(“the Copyright Owner”)

IOP Ref: JPhysA-109093.R1

Part 1 - Publication on a Subscription basis

1.1 In consideration for acceptance and publication of the Article, the Named Authors of the Article and/or the Copyright Owner hereby assign, where necessary by present assignment of future copyright, to IOP with full title guarantee the entire copyright in all original material published as part of the Article (which expression includes but is not limited to the text, abstract, tables, figures and graphs, related corrigenda or “comments” and multimedia content but excludes any other item referred to as supplementary material and/or any video abstract) throughout the world for the full term of copyright (including any extensions or renewals thereof) for all media and formats, whether known or unknown. Such assignment shall be effective only if the Article (or any resubmission of the Article) is accepted for publication. For the avoidance of doubt, copyright does not subsist in any fundamental data underlying the Article and nothing in this agreement is intended to limit access to or use of such data.

1.2 If the Article, or any part of it, is protected by Crown copyright, in consideration for acceptance and publication of the Article, the relevant Named Authors and the relevant originating department or agency hereby grant IOP a non-exclusive royalty-free worldwide freely-transferrable licence for the full term of copyright (including any extensions or renewals thereof) for all media and formats, whether known or unknown, to do in relation to the Article all acts restricted by copyright worldwide including, but not limited to, the right of action under section 101A of the Copyright Designs and Patents Act 1988. Such licence shall be effective only if the Article is accepted for publication.

1.3 If all the Named Authors are employees of the US Government, they represent and warrant to IOP that the Article was prepared as part of their official duties. In such circumstances, or where the Article was created as part of a work for hire, none of the original content within the Article is subject to copyright protection as it is in the public domain.

1.4 In consideration for acceptance and publication of the Article, the Named Authors and/or the Copyright Owner hereby grant IOP a royalty-free non-exclusive worldwide freely transferrable licence for the full term of copyright (including any extensions or renewals thereof) to do in relation to any supplementary material not deemed to be part of the Article and/or any video abstract all acts restricted by copyright worldwide. This shall include, but not be limited to, making the material available under any licence that IOP deems appropriate for purposes including, but not limited to, the maximisation of visibility and the long term preservation of the content.

1.5 Each of the Named Authors consents to all publication and processing of their personal data by IOP, as that data is displayed on the Article, including, but not limited to, the names and email addresses of the Named Authors. Accordingly, the Named Authors shall not object on data protection grounds to the use of their personal data on the Article wherever IOP chooses to display it, whether itself or via a third party.

Representations and warranties

2.1 The Copyright Owner and/or the Submitting Author on behalf of the Named Authors (as appropriate) represent and warrant that:

2.1.1 the Article is the original work of the Named Authors;

2.1.2 the Article has not been published previously in any form, other than in accordance with our Preprint pre-publication policy;

2.1.3 each of the Named Authors has made a material contribution to the conception and/or writing of the Article, has received the final version of the Article, has agreed to its submission on the terms contained herein and takes responsibility for it and submission has been approved as necessary by the authorities at the establishment where the research was carried out;

2.1.4 the Submitting Author completes and returns this agreement as authorised agent for and on behalf of all the Named Authors and the Copyright Owner (as applicable) and has the full power to enter into this agreement and to make the grants and assignments it contains;

2.1.5 the Article has not been and shall not be submitted to another publisher prior to withdrawal or rejection by IOP;

2.1.6 the Article does not infringe any third party rights, it contains nothing libellous or unlawful, all factual statements are to the best of the Named Authors' knowledge true or based on valid research conducted according to accepted norms and all required permissions have been obtained in writing;

2.1.7 the Article expressly acknowledges any third party funding and/or potential conflicts of interest; and

2.1.8 any supplementary material or video abstract is the original work of the Named Authors, or is the property of the Copyright Owner, or permission has been obtained from its owner(s) for its publication by IOP and permission has been obtained for the inclusion of any third party content.
2.2 The Named Authors and/or the Copyright Owner (as appropriate) indemnify and will keep indemnified IOP against all costs and expenses suffered or incurred by IOP as a result of and/or arising out of any breach of the representations and/or warranties in this section 2.

The Named Authors' rights

3.1 IOP grants the Named Authors the rights specified in paragraphs 3.2 and 3.3. All such rights must be exercised solely for non-commercial purposes. Where possible, any use should display citation information and IOP's copyright notice, and, for electronic use, best efforts must be made to include a link to the online abstract in the Journal.

Exercise of the rights in paragraph 3.2 may use the peer reviewed, edited, formatted and typeset version of the Article including any tagging, indexing and other enhancements published by IOP and/or its licensors ("Final Published Version").

Exercise of the rights referred to in paragraph 3.3 must not use the Final Published Version and extend only to the version of the Article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP of a header, an article ID, a cover sheet and/or an 'Accepted Manuscript' watermark, but excluding any other editing, typesetting or other changes made by IOP and/or its licensors (the "Accepted Manuscript") and must be accompanied by the following statement of provenance:

'This is the Accepted Manuscript version of an article accepted for publication in Journal of Physics A: Mathematical and Theoretical.
IOP Publishing Ltd is not responsible for any errors or omissions in this version of the manuscript or any

version derived from it. The Version of Record is available online at [insert DOI].'

3.2 The rights are:

3.2.1 To make copies of the Final Published Version of the Article (all or part) for teaching purposes;

3.2.2 To include the Final Published Version of the Article (all or part) in a research thesis or dissertation provided it is not then published commercially;

3.2.3 To make oral presentation of the Final Published Version of the Article (all or part) and to include a summary and/or highlights of it in papers distributed at such presentations or in conference proceedings; and

3.2.4 To use original figures and text from the Final Published Version of the Article falling within the quota outlined in and subject to the STM Permissions Guidelines (http://www.stm-assoc.org/permissions-guidelines/) at the relevant time in force.
For the avoidance of doubt, the Named Authors retain all proprietary rights in the Article other than copyright.

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **51** (2018) 184001 (20pp) https://doi.org/10.1088/1751-8121/aab7ff

Coherent state coding approaches the capacity of non-Gaussian bosonic channels

Stefan Huber[1] and Robert König

Institute for Advanced Study & Zentrum Mathematik, Technical University of Munich, 85748 Garching, Germany

E-mail: stefan.huber@ma.tum.de

Received 2 November 2017, revised 13 March 2018
Accepted for publication 20 March 2018
Published 9 April 2018

Abstract

The additivity problem asks if the use of entanglement can boost the information-carrying capacity of a given channel beyond what is achievable by coding with simple product states only. This has recently been shown not to be the case for phase-insensitive one-mode Gaussian channels, but remains unresolved in general. Here we consider two general classes of bosonic noise channels, which include phase-insensitive Gaussian channels as special cases: these are attenuators with general, potentially non-Gaussian environment states and classical noise channels with general probabilistic noise. We show that additivity violations, if existent, are rather minor for all these channels: the maximal gain in classical capacity is bounded by a constant independent of the input energy. Our proof shows that coding by simple classical modulation of coherent states is close to optimal.

Keywords: quantum channels, classical capacity, quantum information theory, quantum optics, non-Gaussian channels, entropy power inequality

(Some figures may appear in colour only in the online journal)

1. Introduction

Communication—be it over time (as storage in a memory) or over space (as transmission from a sender to a receiver)—is one of the central primitives studied in information theory. A channel represents a general model for communication. With respect to communication, its arguably most fundamental characteristic is its classical capacity: the maximal number of bits or—more precisely—the maximal rate at which bits may be transmitted through the channel

[1] Author to whom any correspondence should be addressed.

asymptotically in the limit of many uses. This quantity is ubiquitous in information theory since it has both practical meaning and is interesting from a purely mathematical point of view.

The classical capacity of quantum channels has been studied for decades [1–3], but is not understood in general. Not only is it very hard to find the optimal encoding for one use of the quantum channel, one also has to take into account the possibility of input states which are entangled across several channel uses. The use of such entangled states can potentially boost the capacity when the channel is used multiple times in parallel as opposed to the use of simple product or separable states. The question of whether such an increase in capacity occurs is commonly referred to as the additivity question. If the use of entanglement can increase the capacity in comparison to product states, we speak of an additivity violation. If, however, such an increase cannot occur, the classical capacity is said to be additive. In a recent breakthrough, the classical capacity for the so-called phase-insensitive bosonic Gaussian channels has been found, and it was shown that the capacity for these channels is achieved by using products of coherent states for the encoding [4]. Furthermore, for more general Gaussian channels which are phase-sensitive, the capacity has been found above a certain energy threshold [4]. This was achieved by proving the optimality of Gaussian inputs above the energy threshold, while the capacity restricted to Gaussian inputs above the energy threshold had been calculated in earlier work [5, 6]. Despite these landmark results, the classical capacity of a general Gaussian channel is not known for all energies, and very little is known about the capacity of non-Gaussian bosonic channels, which are rarely considered in literature. However, recent developments in the area of entropy power inequalities [7–9] and the minimum output entropies of bosonic channels [10–13] make it possible to give bounds on the output entropy of non-Gaussian bosonic channels. These inequalities promise to help at least partially resolve the question of classical capacity for these channels. Here we apply entropy power inequalities to give upper and lower bounds on the classical capacity of a wide class of bosonic noise channels, which includes non-Gaussian channels.

The channels we consider are attenuator channels with general environment states which are possibly non-Gaussian and classical noise channels with general additive noise. While we cannot show that the classical capacity of these channels is additive, we can show that additivity violations, if existent, are minor. By this we mean that the maximal difference between the full capacity and the one-shot capacity is a constant independent of the input energy. Furthermore, the lower bounds we obtain are achievable by simple coherent state coding. As mentioned, this coding achieves the capacity for phase-insensitive Gaussian channels. It is, however, known not to be optimal in general even for Gaussian channels [14]. Nonetheless, as our results show, it still is a suitable choice of encoding for the large class of bosonic channels we consider (which includes the Gaussian channels for which the capacity is known as special cases).

Our paper is structured as follows: in section 2, we review the notion of capacity and related formulas and give a technical introduction to the problem we consider. In section 3, we state the relevant entropy power and photon number inequalities and give an overview of the techniques we apply. In the following section 4, we state our bounds on the classical capacity of the channels. We present the proofs in section 5. We close with a discussion of our work and related problems as well as directions for future research.

2. Capacity of channels

In the classical setting, where a channel is simply a conditional distribution $P_{Y|X}$ describing the channel's probabilistic output Y on a given input X, the classical capacity has been studied by Shannon in his landmark paper [15]. It is remarkable that although the definition of the capacity involves an arbitrarily large number of channel uses, in the classical setting it can be expressed in terms of an optimization problem defined in terms of a single channel use only. It is given by

$$C\left(P_{Y|X}\right)=\sup_{P_X} I(X:Y), \tag{1}$$

where $I(X:Y)=H(X)+H(Y)-H(XY)$ is the mutual information between the input and output of the channel, defined in terms of the Shannon entropy $H(X)=-\sum_x P_X(x)\log P_X(x)$. In the continuous-variable case of interest here, X and Y are random variables on $\mathbb{R}$, and an energy constraint needs to be imposed on the distributions P_X in (1): the capacity with input 'energy' E is then defined as in (1), but with a constraint $\mathbb{E}[X^2]\leqslant E$ on the second moment. This constraint amounts to the demand that in the operational coding problem defining the capacity, only codewords, i.e. sequences $m_x=(m_1,\ldots,m_n)\in\mathbb{R}^n$ of elements having a mean energy $\frac{1}{n}\sum_{i=1}^n m_i^2$ bounded by E are allowed. We will denote the corresponding capacity by $C_{\mathrm{E}}(P_{Y|X})$.

In the quantum setting, a quantum channel is a completely positive trace-preserving (CPTP) map $\mathcal{E}:\mathcal{B}(\mathcal{H})\to\mathcal{B}(\mathcal{H})$ on the set $\mathcal{B}(\mathcal{H})$ of bounded operators on $\mathcal{H}$. For the bosonic systems considered here, $\mathcal{H}\cong L^2(\mathbb{R})$ is an infinite-dimensional separable Hilbert space. We are concerned with the energy-constrained classical capacity of such channels: it is operationally defined as the maximal achievable rate R at which classical bits can be sent by (i) encoding a message $x\in\{0,1\}^{\lfloor nR\rfloor}$ into a state ϱ_x on $\mathcal{H}^{\otimes n}$, and (ii) decoding the received state $\mathcal{E}(\varrho_x)$ using a suitable POVM $\{F_x\}_{x\in\{0,1\}^{\lfloor nR\rfloor}}$ on $\mathcal{H}^{\otimes n}$, in such a way that the average decoding error probability vanishes in the limit $n\to\infty$. In this operational problem, the energy constraint amounts to imposing the physical restriction that the mean photon number $\frac{1}{n}\mathrm{tr}(\sum_{j=1}^n a_j^\dagger a_j\overline{\varrho})$ of the average input state $\overline{\varrho}$ is bounded by some constant N. Here $a_j^\dagger, a_j$ are the creation and annihiliation operators of the j-th mode of a system of n harmonic oscillators, satisfying the canonical commutation relations $[a_j,a_k^\dagger]=\delta_{jk}\mathbb{1}$, $[a_j,a_k]=0$.

An expression for the classical capacity of a quantum channel similar to (1) is known: the Holevo–Schumacher–Westmoreland theorem (HSW theorem) [1, 2] states that the classical capacity of a quantum channel $\mathcal{E}$ subject to the energy constraint N can be obtained by evaluating the limit

$$C_N(\mathcal{E})=\lim_{n\to\infty}\frac{1}{n}\chi_{nN}\left(\mathcal{E}^{\otimes n}\right). \tag{2}$$

In this expression, $\chi_{nN}(\mathcal{E}^{\otimes n})$ is the Holevo quantity evaluated for the channel $\mathcal{E}^{\otimes n}:\mathcal{B}(\mathcal{H}^{\otimes n})\to\mathcal{B}(\mathcal{H}^{\otimes n})$ with average energy constraint N. The latter is defined as

$$\chi_{nN}(\mathcal{E}^{\otimes n})=\sup_{\{p_x,\varrho_x\}_x} S\big(\mathcal{E}^{\otimes n}(\overline{\varrho})\big)-\sum_x p_x S\big(\mathcal{E}^{\otimes n}(\varrho_x)\big), \tag{3}$$

where $S(\varrho)=-\mathrm{tr}\varrho\log\varrho$ denotes the von Neumann entropy, and where the optimization is over all ensembles $\{p_x,\varrho_x\}_x$ of states on $\mathcal{H}^{\otimes n}$ with average signal state $\overline{\varrho}=\sum_x p_x\varrho_x$ satisfying the average energy constraint $\mathrm{tr}\left(\sum_{j=1}^n a_j^\dagger a_j\overline{\varrho}\right)\leqslant nN$. In general, one may also consider continuous ensembles of the form $\{p(x)\mathrm{d}x,\varrho_x\}_x$ where $\mathrm{d}x$ is e.g. the Lebesgue measure on $\mathbb{R}^n$, and p is a probability density function. In this case, sums need to be replaced by integrals.

Expression (2) for the classical capacity generalizes formula (1) to quantum channels. Unfortunately, though, it is generally intractable both numerically and analytically: it requires optimization over an arbitrarily large number n of copies of the channel. The regularization (i.e. the process of taking the limit $n\to\infty$ in (2)) is necessary to allow for input (code) states which are entangled across several channel uses: such states may potentially improve upon the capacity compared to product states. It is worth noting that the use of separable states as input states, which are neither entangled nor product states, does not improve the capacity in comparison to the use of product states only. This follows from the fact that it is enough to consider pure input states [1] and the fact that pure separable states are product states. The capacity when one restricts to codes using only unentangled signal states, that is, the *one-shot capacity*, is given by $\chi_N(\mathcal{E})$. It gives the lower bound

$$C_N(\mathcal{E}) \geqslant \chi_N(\mathcal{E}) \tag{4}$$

on the classical capacity.

The additivity problem consists in the question of whether or not the inequality (4) is strict (in which case we speak of an additivity violation), or simply an equality. Its name derives from the fact that if one has

$$\chi_{nN}(\mathcal{E}^{\otimes n}) = n\chi_N(\mathcal{E}) \qquad \text{for all } n \in \mathbb{N}, \tag{5}$$

then one immediately obtains equality in (4), implying that entangled signal states offer no operational advantage.

While it is still unknown whether or not equality holds in (4) in general in the continuous-variable case, the simpler additivity property (5) for the Holevo quantity has been shown not to hold in general by Hastings [16]. He showed that there exists a channel $\mathcal{T} : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$ for which

$$\chi(\mathcal{T}^{\otimes 2}) > 2\chi(\mathcal{T})\ .$$

(There is no energy constraint here because only finite-dimensional Hilbert spaces are involved.) In principle, this leaves room for an improvement upon the classical capacity via the use of entangled signal states, i.e. the inequality (4) may still be strict for certain channels. Understanding when this may or may not be the case is one of the central challenges of quantum information theory, and fits into the larger theme of investigating the impact of quantum effects such as entanglement on the power of information-processing primitives.

2.1. Bosonic noise channels

Here we explore the potential of additivity violations in channels associated with bosonic systems. The quantum channels discussed here are attenuation channels (i.e. beamsplitters coupling the system to an environment in a general state) as well as channels mixing the system with a classical random variable. These are natural generalizations of the corresponding Gaussian channels: the thermal noise channel and the classical noise channel. These Gaussian channels have been the subject of various earlier analyses [4, 17–20], and, as discussed below, have been shown not to violate additivity in a recent breakthrough development. In contrast, our emphasis here is on general, potentially non-Gaussian bosonic channels, for which no additivity statements have been known previously.

In more detail, we consider an input system of d bosonic modes (with Hilbert space $\mathcal{H}^{\otimes d}$ where $\mathcal{H} \cong L^2(\mathbb{R})$) with the vector of mode operators $R = (Q_1, P_1, \ldots, Q_d, P_d)$. The action of a beamsplitter with transmissivity $0 \leqslant \lambda \leqslant 1$ which couples the system with an environment of d bosonic modes with mode operators $(Q_1^{(\mathrm{E})}, P_1^{(\mathrm{E})}, \ldots, Q_d^{(\mathrm{E})}, P_d^{(\mathrm{E})})$ is given by a Gaussian unitary U_λ on $\mathcal{H}^{\otimes d} \otimes \mathcal{H}^{\otimes d}$. Its action on the $2d$ modes is defined (in the Heisenberg picture) by the symplectic matrix

$$S_\lambda = \begin{pmatrix} \sqrt{\lambda}\mathbb{1}_{2d} & \sqrt{1-\lambda}\mathbb{1}_{2d} \\ \sqrt{1-\lambda}\mathbb{1}_{2d} & -\sqrt{\lambda}\mathbb{1}_{2d} \end{pmatrix}$$

with respect to the ordering $(Q_1, P_1, \ldots, Q_d, P_d, Q_1^{(\mathrm{E})}, P_1^{(\mathrm{E})}, \ldots, Q_d^{(\mathrm{E})}, P_d^{(\mathrm{E})})$ of modes, i.e. $U_\lambda^\dagger R_j U_\lambda = \sum_k (S_\lambda)_{j,k} R_k$. If we assume that the environment is in some state σ_E (decoupled from the system), and consider only the action of this unitary on the system, we obtain the quantum channel

$$\mathcal{E}_{\lambda,\sigma_\mathrm{E}}(\varrho) := \mathrm{tr}_\mathrm{E}\left(U_\lambda(\varrho \otimes \sigma_\mathrm{E})U_\lambda^\dagger\right) . \tag{6}$$

We are interested in the classical capacity of channels $\mathcal{E}_{\lambda,\sigma_\mathrm{E}}$ of the form (6), which we call attenuation channels. Note that this set of channels includes Gaussian channels (for Gaussian states σ_E) such as the thermal noise channels (when $\sigma_\mathrm{E} = \mathrm{e}^{-\beta(\sum_{j=1}^d Q_j^2 + P_j^2)}/Z$ is a thermal state, i.e. the Gibbs state of a certain quadratic Hamiltonian) or the pure loss channel (when σ_E is the vacuum state). It also includes non-Gaussian channels (for σ_E a non-Gaussian state): typical examples include, e.g. the case where σ_E slightly deviates from a thermal state, or is, e.g. some finite superposition of number states.

A second class of channels we consider here are channels which act by displacing the system according to some probability density function $f : \mathbb{R}^{2d} \to \mathbb{R}$ on phase space. We call these (general) classical noise channels. They act as

$$\mathcal{F}_{t,f}(\varrho) := \int f(\xi) W(\sqrt{t}\xi) \varrho W(\sqrt{t}\xi)^\dagger \mathrm{d}^{2d}\xi, \tag{7}$$

where $W(\xi) = \mathrm{e}^{\mathrm{i}\sqrt{2\pi}\xi\cdot(\sigma R)}$ for $\xi \in \mathbb{R}^{2d}$ are the Weyl displacement operators with the symplectic form $\sigma = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus d}$ and the mode operators $R = (Q_1, P_1, \ldots, Q_d, P_d)$. Here $t > 0$ is some parameter analogous to the transmissivity. Again, this channel may be non-Gaussian (if f is not a Gaussian distribution). Channels of this type have been considered by Werner [21], who described them as a convolution operation between a probability distribution and a state. They satisfy a number of convenient properties with respect to displacements in phase space as well as a data processing inequality. For more background we refer to [9, 21, 22].

For a specific kind of quantum channels, the so-called single-mode phase-insensitive[2] Gaussian channels, the classical capacity has recently been found by Giovannetti *et al* [4, 25].

The channels (6) and (7) fall into this class if the environment state σ_E is a thermal state or if the probability density function f is a Gaussian distribution whose covariance matrix is proportional to the identity (this special case has commonly been referred to as the classical noise channel, see [17, 26]). In particular, if σ_E is a single-mode Gaussian thermal state, the capacity of the single-mode ($d = 1$) channel $\mathcal{E}_{\lambda,\sigma_\mathrm{E}}$ is given by

$$C_N(\mathcal{E}_{\lambda,\sigma_\mathrm{E}}) = g\big(\lambda N + (1-\lambda)N_\mathrm{E}\big) - g\big((1-\lambda)N_\mathrm{E}\big), \tag{8}$$

with the mean photon number $N_\mathrm{E} = \mathrm{tr}(a^\dagger a \sigma_\mathrm{E})$ of the environment. Furthermore, if f is a centered Gaussian distribution of unit variance, then the single-mode channel $\mathcal{F}_{t,f}$ has capacity [4]

$$C_N(\mathcal{F}_{t,f}) = g(N + 2\pi t) - g(2\pi t) . \tag{9}$$

In both cases, the quantities on the rhs of equations (8) and (9) have been shown to be equal to the single-shot Holevo information ($\chi_N(\mathcal{E}_{\lambda,\sigma_\mathrm{E}})$ and $\chi_N(\mathcal{F}_{t,f})$, respectively). In particular, there is no violation of additivity in these channels and thus no operational gain in using entangled

[2] A phase-insensitive channel is a channel Φ which has one of the following properties under phase shift operations $\mathrm{e}^{\mathrm{i}\varphi a^\dagger a}$ [23, 24]: $\Phi(\mathrm{e}^{\mathrm{i}\varphi a^\dagger a}\varrho \mathrm{e}^{-\mathrm{i}\varphi a^\dagger a}) = \mathrm{e}^{\mathrm{i}\varphi a^\dagger a}\Phi(\varrho)\mathrm{e}^{-\mathrm{i}\varphi a^\dagger a}$ for all ϱ for gauge-covariant channels and with reversed order of operators on the rhs for gauge-contravariant channels.

signal states. Similar capacity formulas have been found for single-mode phase-sensitive Gaussian channels, where the environment is in a Gaussian state σ_{E} which might be squeezed (and respectively, if the function f is a Gaussian whose covariance matrix is not proportional to the identity). In this case the capacity is only known if the energy constraint allows for input energies larger than a certain threshold value [4–6].

This fundamental result is striking, but leaves open the question of whether additivity violation is possible in more general channels. This is one motivation for considering the more general families $\{\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}\}$ and $\{\mathcal{F}_{t,f}\}$ of single-mode channels, which also include non-Gaussian examples.

We find that additivity violations, if at all existent, must be limited: the difference $C_N(\Phi) - \chi_N(\Phi)$ between the two sides of (4) is upper bounded by a constant independent of the input photon number N, for any channel Φ in the class of attenuators and classical noise channels. In other words, we show that the maximal potential gain achievable by entangled coding strategies is limited. This means that for these channels, the use of entanglement cannot improve the classical capacity achieved by classical modulation of coherent states by much: with growing input energies, the maximal gain by coding strategies using entanglement becomes negligible compared to the value of the capacity.

Our work follows similar reasoning as that of König and Smith [18], who addressed the question whether entangled coding strategies can substantially increase the classical capacity of thermal noise channels: we also employ entropy power inequalities to obtain upper bounds on the capacity. However, in contrast to [18], we also need to establish new achievability (lower) bounds on the capacity: here we again use entropy power inequalities, as well as Gaussian extremality—this reasoning follows pioneering work by Shannon [15]. While the results of [18] for the thermal noise channel have by now been superseded by the explicit capacity formulas of [4], it appears unlikely that similar explicit formulas can be established with present-day analytical methods for the non-Gaussian channels discussed here.

3. Analytical tools for bosonic systems

Evaluating the classical capacity amounts to solving a highly non-trivial optimization problem. Even for the one-shot capacity, which does not involve an infinite limit over parallel uses of the channel, explicit capacity formulas are generally not known. In special cases, e.g. when the channel exhibits certain symmetries (in particular, gauge-covariance or contravariance in the bosonic context), this difficulty can be overcome [4, 27, 28]. However, the focus of our work is on more general, possibly non-Gaussian channels. In this context, only few analytical tools are known: these include entropy power inequalities (EPI), the Gaussian maximum entropy principle, as well as certain more recent Gaussian optimizer results. In this section, we briefly review these results, and also discuss related conjectures. In section 4, we then discuss the implications of these statements to classical capacities: we will see that they imply various bounds on the possible degree of non-additivity.

3.1. Gaussian extremality

A key tool in dealing with non-Gaussian distributions, states and optimization are Gaussian extremality results. The main result we use here is Gaussian extremality for the von Neumann entropy (but see [29] for more general statements and applications): this states that among all states with fixed first and second moments, the Gaussian state has maximal entropy. Succinctly, this can be expressed by the inequality

$$S(\varrho) \leqslant S([\varrho]), \tag{10}$$

where $[\varrho]$ is the Gaussian state with the same first and second moments as ϱ. As a corollary, among all one-mode states ϱ with mean photon number $\mathrm{tr}(a^\dagger a\varrho)$ smaller or equal to N, the Gaussian thermal state $\varrho_{\mathrm{th},N} = \frac{1}{N+1}\sum_{n=0}^{\infty}\left(\frac{N}{N+1}\right)^n |n\rangle\langle n|$ has maximal entropy $g(N) := (N+1)\log(N+1) - N\log(N)$. A simple proof of this corollary can be found, for instance, in [30, lemma 9], while the entropy of Gaussian states has been calculated in [31].

Gaussian states also turn out to be optimal for various entropy or entropy-related quantities defined in terms of Gaussian operations. For example, Gaussian states have recently been shown to be the optimizers of the defining problem of $\|\Phi\|_{p\to p}$-norms for a Gaussian channel Φ, see [32, 33]. In a similar context, in a series of recent works by De Palma, Trevisan, and Giovannetti [10–13], it was shown that the output entropy of any gauge-covariant one-mode Gaussian channel for a fixed input entropy is minimized by taking as input state the thermal state of this fixed input entropy. For the beamsplitter with environment in the thermal state $\varrho_{\mathrm{th},N}$, this can be stated as [13, theorem 4]

$$S(\mathcal{E}_{\lambda,\varrho_{\mathrm{th}}}(\sigma)) \geqslant g\big(\lambda g^{-1}[S(\sigma)] + (1-\lambda)N\big)\ . \tag{11}$$

As explained below, equation (11) is a special case of the currently unproven entropy photon number inequality (EPNI) conjecture, which is of relevance to this work.

3.2. Entropy power inequalities

In classical information theory, the Shannon entropy of an $\mathbb{R}^n$-valued random variable X with probability density $f:\mathbb{R}^n \to \mathbb{R}$ is given by $H(X) = -\int f(x)\log f(x)\mathrm{d}^n x$. In order to estimate the capacity of additive noise channels, Shannon [34] proposed the entropy power inequality (EPI)

$$\mathrm{e}^{2H(X+Y)/n} \geqslant \mathrm{e}^{2H(X)/n} + \mathrm{e}^{2H(Y)/n}, \tag{12}$$

where the lhs is the entropy power of the sum of two independent random variables X and Y. A rigorous proof of (12) was established by Stam [35, 36] under the assumption that X and Y are of finite variance. Blachman in [36] gave a detailed account of Stam's proof, and Lieb in [37] subsequently found a different proof of the entropy power inequality using Young's inequality for convolutions.

In the context of quantum information theory, Shannon's entropy power inequality has been generalized. In [7], the inequality

$$\mathrm{e}^{S\left(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho)\right)/n} \geqslant \lambda \mathrm{e}^{S(\varrho)/n} + (1-\lambda)\mathrm{e}^{S(\sigma_{\mathrm{E}})/n} \tag{13}$$

was shown for $\lambda = 1/2$. This was then generalized by De Palma *et al* [8] to all $\lambda \in [0, 1]$. The lhs involves the von Neumann entropy of the output $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho)$ under a beamsplitter of transmissivity λ, defined in (6), and can be considered as a quantum convolution operation of two n-mode states ϱ and σ_{E}, analogous to the convolution $X+Y$ of two independent random variables X and Y. More recently, a conditional version of the entropy power inequality has been proven by De Palma and Trevisan [30], generalizing a statement previously established for Gaussian states only [38]. This result can be stated as

$$\exp\frac{S(C|E)_{\varrho_{CE}}}{n} \geqslant \lambda\exp\frac{S(A|E)_{\varrho_{AE}}}{n} + (1-\lambda)\exp\frac{S(B|E)_{\varrho_{BE}}}{n} \tag{14}$$

for the input system A, the environment system B, and the output system C of the beamsplitter. This concerns a tripartite state ϱ_{ABE} consisting of two n-mode systems A and B which are conditionally independent given E, as well as the result $\varrho_{CE} = \mathrm{tr}_B\left((U_\lambda \otimes I_E)\varrho_{ABE}(U_\lambda \otimes I_E)^\dagger\right)$ of applying a transmissivity-λ beamsplitter U_λ to AB. Remarkably, the proof presented in [30] circumvents all regularity assumptions required in earlier proofs: it is valid for any state ϱ_{ABM} with finite mean photon number (second moments) in AB and satisfying $S(\varrho_E) < \infty$. This also implies the validity of (13) for all states ϱ and σ_E with finite mean photon number. The conditional entropy power inequality (14) has application to establishing upper bounds on the entanglement-assisted classical capacity of bosonic quantum channels, as proposed in [38].

Another generalization of (12) has been established in [9] and can be stated as

$$\mathrm{e}^{S\left(\mathcal{F}_{t,f}(\varrho)\right)/n} \geqslant \mathrm{e}^{S(\varrho)/n} + t\mathrm{e}^{H(f)/n}, \tag{15}$$

for a probability density function $f : \mathbb{R}^{2n} \to \mathbb{R}$ and an n-mode state ϱ. The lhs of this inequality involves the entropy power of the output $\mathcal{F}_{t,f}(\varrho)$ of the classical noise channel, defined in (7). As before, the expression $\mathcal{F}_{t,f}(\varrho)$ can be considered as a convolution operation, in this case between a probability density function f and a quantum state ϱ. The proof presented in [9] follows earlier heat-flow arguments and requires certain regularity assumptions. It appears straightforward, however, to adapt the proof of [30] to this setting—this would show that (15) holds for all probability density functions f with finite second moments and all states ϱ with finite mean photon number.

3.3. *Conjectured entropy photon number inequalities*

In [39] and [40], an alternative to the entropy power inequality (13) has been proposed, replacing the entropy power of ϱ by the mean photon number of a Gaussian thermal state with the same entropy. This inequality is called the Entropy Photon-Number Inequality (EPNI) and can be stated as

$$g^{-1}\left(\frac{S(\mathcal{E}_{\lambda,\sigma_\mathrm{E}}(\varrho))}{n}\right) \geqslant \lambda g^{-1}\left(\frac{S(\varrho)}{n}\right) + (1-\lambda)g^{-1}\left(\frac{S(\sigma_\mathrm{E})}{n}\right), \tag{16}$$

where the channel $\mathcal{E}_{\lambda,\sigma_\mathrm{E}}$ is used in parallel on n modes. This statement can be seen as a generalization of (11) to multiple modes and the case when the environment is not a thermal state. Despite progress in certain special cases [41], and the special case of (11), the EPNI remains unproven. However, its implications on capacities have been studied in a number of works [39, 40, 42].

It is natural to ask whether an EPNI also holds in the case of the channel $\mathcal{F}_{t,f}$. We conjecture that this is the case and that the 'classical-quantum' EPNI reads

$$g^{-1}\left(\frac{S(\mathcal{F}_{t,f}(\varrho))}{n}\right) \geqslant g^{-1}\left(\frac{S(\varrho)}{n}\right) + \frac{t}{e}\mathrm{e}^{\frac{H(f)}{n}}. \tag{17}$$

A weaker statement which would still be useful for establishing lower bounds on the classical capacity is the specialization of this EPNI to the case of one mode and a thermal input state:

$$g^{-1}\left[S\left(\mathcal{F}_{t,f}(\varrho_{\mathrm{th},N})\right)\right] \geqslant N + \frac{t}{e}\mathrm{e}^{H(f)}, \tag{18}$$

where $\varrho_{\mathrm{th},N}$ is the Gaussian thermal state with mean photon number N as introduced above. This inequality can be seen as a classical noise channel analog of equation (11) for the attenuator. A proof of equation (18) might be easier than the full proof of equation (17) and still have desirable implications: we show that assuming the validity of equations (18) and (17) leads to better bounds than using the entropy power inequality, again without altering the spirit of the theorems.

Having reviewed the key tools required for the discussion, we continue with the statement of our results in the next section.

4. Limited non-additivity for non-Gaussian bosonic channels

Here we show that the recent generalizations of the EPI as well as the related results discussed in section 3 have direct implications for the classical capacities of non-Gaussian bosonic channels. In particular, they imply that the degree of potential non-additivity is limited for attenuators and classical noise channels. A similar analysis has been carried out in [18] for special cases of the channels considered here, namely Gaussian thermal noise channels. In contrast to that work, we show that the degree of non-additivity can also be bounded for non-Gaussian channels. This extends our understanding of classical capacities to previously untreatable cases.

One of the key observation is that EPIs can be used to obtain not only upper (converse) bounds on the classical capacity of non-Gaussian channels, but also achievability bounds. Remarkably, these achievability bounds concern simple product state codes consisting of coherent states. This coding strategy has recently been shown to be optimal for phase-insensitive Gaussian channels. It is known to not be optimal [4–6] for certain phase-sensitive Gaussian channels, where coding with squeezed coherent states achieves a higher rate. Nonetheless, coherent state coding gives a good lower bound also in these cases, as we will see below.

Our work implies that the same strategy of using coherent states is also essentially optimal for the more general non-Gaussian channels considered here. Thus although this coding strategy is in general not optimal even for Gaussian channels, we show that it is a suitable choice of coding strategy for non-Gaussian channels.

In spirit, our results can be seen as quantum generalizations of Shannon's work, in which he applied the entropy power inequality to the capacity of the additive noise channel [34]: He found that the capacity C_P of a classical additivity channel $Y = X + Z$, where Z is noise independent of the input X (but otherwise arbitrary), is bounded by

$$\log \frac{P+N_1}{N_1} \leqslant C_P \leqslant \log \frac{P+N}{N_1}, \tag{19}$$

where P is the average transmitter power, N is the average noise power, and $N_1 = \mathrm{e}^{2H(Z)}/(2\pi e)$ is the entropy power of the noise Z. In the special case where Z is distributed according to a standard normal distribution, the upper and lower bounds in equation (19) coincide and reduce to Shannon's capacity formula for the additive white Gaussian noise channel.

Our main result concerns the single-mode attenuation channel $\mathcal{E}_{\lambda,\sigma_\mathrm{E}}$ introduced in (6), and the classical noise channel $\mathcal{F}_{t,f}$ introduced in (7). We shall denote the mean photon number $\mathrm{tr}(a_\mathrm{E}^\dagger a_\mathrm{E}\sigma_\mathrm{E})$ of σ_E by N_E, and its von Neumann entropy by $S(\sigma_\mathrm{E})$. Similarly, we write

$$\mathrm{E}(f) = \sum_{i=1}^{2} \int \xi_i^2 f(\xi)\mathrm{d}^2\xi$$

for the energy (i.e. the sum of second moments) of the distribution f and

$$H(f) = -\int f(\xi)\log f(\xi)\mathrm{d}^2\xi$$

for the Shannon entropy of the associated random variable. Throughout, we will assume that these quantities are finite. We then have the following main result:

Theorem 1. *The maximal degree of non-additivity of the classical capacity of $\mathcal{E}_{\lambda,\sigma_{\rm E}}$ and $\mathcal{F}_{t,f}$ is bounded as*

$$C_N(\mathcal{E}_{\lambda,\sigma_{\rm E}}) - \chi_N(\mathcal{E}_{\lambda,\sigma_{\rm E}}) \leqslant 2g\big((1-\lambda)N_{\rm E}\big) - g\big((1-\lambda)N_{\rm E}^{\rm ep}\big) - \log\Big(\lambda + (1-\lambda)\mathrm{e}^{S(\sigma_{\rm E})}\Big),$$

$$C_N(\mathcal{F}_{t,f}) - \chi_N(\mathcal{F}_{t,f}) \leqslant 2g(\pi t\mathrm{E}(f)) - \log\Big(1 + t\mathrm{e}^{H(f)}\Big),$$

independently of the input energy N, where $N_{\rm E}^{\rm ep} = g^{-1}\left[S(\sigma_{\rm E})\right]$ is the mean photon number of a thermal state with the same entropy as $\sigma_{\rm E}$. For both channels, coherent states modulated with a Gaussian distribution achieve a rate which differs from the capacity by at most the rhs of these bounds.

The fact that the rhs of these bounds is independent of the input energy N implies that the degree of violation is at most constant. In particular, the potential violation is negligible compared to the actual value of the capacity for large N. In other words, there is no significant advantage in using entangled states for coding. This result is indeed not surprising: For very large values of N, the associated quantum channels are 'almost classical' and therefore it is natural to expect quantum effects such as non-additivity to become small in this regime.

The maximal degree of violation depends on the structure of the environment (the state $\sigma_{\rm E}$ respectively the distribution f) and is not simply a universal constant as in [18]. This is not surprising because [18] only considered attenuation channels with Gaussian thermal states in the environment (also called thermal noise channels).

Unlike Shannon's result (19), the bounds in theorem 1 do not specialize to the known capacity results for Gaussian channels in the case where $\sigma_{\rm E}$ is a thermal state or f is a unit-variance centered normal distribution. We show that stronger bounds with this property can be derived assuming the validity of the EPNI conjecture:

Theorem 2. *Assuming the EPNI conjecture (16) holds, we have*

$$C_N(\mathcal{E}_{\lambda,\sigma_{\rm E}}) - \chi_N(\mathcal{E}_{\lambda,\sigma_{\rm E}}) \leqslant 2\Big[g\big((1-\lambda)N_{\rm E}\big) - g\big((1-\lambda)N_{\rm E}^{\rm ep}\big)\Big].$$

Similarly, assuming that the EPNI conjecture (17) holds, we have

$$C_N(\mathcal{F}_{t,f}) - \chi_N(\mathcal{F}_{t,f}) \leqslant 2\left[g(\pi t\mathrm{E}(f)) - g\left(\frac{t}{e}\mathrm{e}^{H(f)}\right)\right].$$

Furthermore, coherent state modulation with a Gaussian distribution achieves a rate which differs from the capacity by at most the rhs of these bounds.

We stress that these bounds hold independently of whether or not $\sigma_{\rm E}$ or f (and thus the channels) are Gaussian. In the special case where $\sigma_{\rm E}$ is a thermal state, we have $N_{\rm E}^{\rm ep} = N_{\rm E}$: here theorem 2 implies that there is no additivity violation, and we recover the Gaussian capacity result for the thermal noise channel (8) (see below). Similarly, if f is a unit-variance centered normal distribution, theorem 2 reduces to the capacity result (9) for the Gaussian classical noise channel. In this respect, theorem 2 behaves similarly as Shannon's bounds (19) and is compatible with the capacity formulas of [4].

Furthermore, if σ_{E} is a squeezed thermal state, coherent state coding is known not to be optimal [5]. The classical capacity for input energies higher than a certain threshold value in this case is achieved by input states which are squeezed coherent states. In this case the rhs in theorem 2 becomes a bound on the difference in rate between squeezed coherent state coding and coherent state coding. The same holds true in the case when f is a Gaussian whose covariance matrix is not proportional to the identity [6]. In these particular cases (above the threshold energy), the gap between the lower and upper bounds we obtain is not due to non-additivity, but simply due to the fact that coherent state coding is not optimal in the one-shot case for these channels. We stress that this does not weaken the statement of our theorems in the case of non-Gaussian channels, which are our main focus and for which our bounds are new.

We point out, however, that theorem 1 (which does not require the EPNI conjectures) essentially gives the same qualitative conclusions for non-Gaussian channels, and theorem 2 does not provide additional information. Indeed, consider for example the case of the attenuation channel, with $\sigma_{\mathrm{E}} = |N_{\mathrm{E}}\rangle\langle N_{\mathrm{E}}|$ equal to one of the number states $|N_{\mathrm{E}}\rangle$. Then both theorems 1 and 2 specialize to

$$C_N(\mathcal{E}_{\lambda,|N_{\mathrm{E}}\rangle\langle N_{\mathrm{E}}|}) - \chi_N(\mathcal{E}_{\lambda,|N_{\mathrm{E}}\rangle\langle N_{\mathrm{E}}|}) \leqslant 2g\big((1-\lambda)N_{\mathrm{E}}\big) .$$

Observe also that for $N_{\mathrm{E}} = 0$, the rhs vanishes, showing that the so-called pure loss channel (with $\sigma_{\mathrm{E}} = |0\rangle\langle 0|$ equal to the vacuum state) does not violate additivity. In particular, this provides a new rederivation of the capacity result of [43] based on the EPI only.

We present the derivation of these results in section 5.

5. Derivation of non-additivity bounds

In this section, we present the proof of theorems 1 and 2. Let us first sketch the basic idea: The Holevo quantity (3) consists of a difference between two entropic quantities. The maximum entropy principle (10) allows us to restrict to Gaussian states when we require upper bounds on entropies (subject to fixed second moments), whereas the entropy power inequalities (13) and (15) (respectively the entropy photon-number inequalities (16) and (17)), as well as equation (11) are suitable to obtain lower bounds on entropies. The bounds on the capacity are then expressions which depend on output entropies restricted to Gaussian input states and Gaussian environments. These quantities can then be bounded with elementary calculations. The derivation of upper bounds on the capacity thus proceeds similarly as in [18]; in addition, we obtain lower bounds on the capacity in a similar fashion.

5.1. Derivation of capacity bounds from EPIs

We first consider the attenuation channel $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}$ introduced in equation (6). The corresponding inequality in theorem 1 follows immediately from the following lemma. The corresponding statement for the classical noise channel $\mathcal{F}_{t,f}$ is shown in lemma 1B below.

5.2. Derivation of bounds from EPIs

Lemma 1A. *The classical capacity of the single-mode attenuation channel $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}$ satisfies*

$$C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \geqslant g\big(\lambda N + (1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\big) - g((1-\lambda)N_{\mathrm{E}}), \tag{20}$$

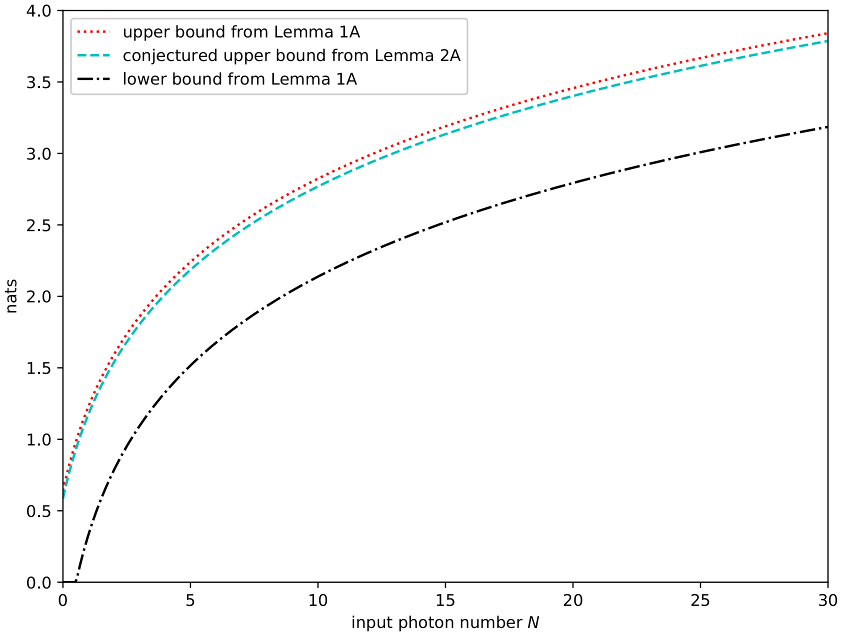


Figure 1. Bounds on the capacity $C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}})$ with $\lambda = \frac{3}{4}$ for any environment state σ_{E} with mean photon number $N_{\mathrm{E}} = 2$ and entropy $S(\sigma_{\mathrm{E}}) \simeq 0.91$ nats $< g(N_{\mathrm{E}}) \simeq 1.91$ nats.

$$C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leqslant g\big(\lambda N + (1-\lambda)N_{\mathrm{E}}\big) - \log\left(\lambda + (1-\lambda)\mathrm{e}^{S(\sigma_{\mathrm{E}})}\right) . \tag{21}$$

The lower bound (20) is achievable with a coherent state ensemble. The difference between this upper and lower bound is bounded by

$$\Delta(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leqslant 2g\big((1-\lambda)N_{\mathrm{E}}\big) - g\big((1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\big) - \log\left(\lambda + (1-\lambda)\mathrm{e}^{S(\sigma_{\mathrm{E}})}\right), \tag{22}$$

independently of the input photon number N.

The corresponding upper and lower bounds on the capacity are visualized in figure 1.

Proof of lemma 1A. *The upper bound.* We prove the upper bound (21) in a similar fashion as [18]. By bounding the Holevo quantity in the Holevo–Schumacher–Westmoreland formula (3) for the classical capacity we have

$$C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leqslant S_N^{\max}(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) - \lim_{n\to\infty}\frac{1}{n}S^{\min}(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}^{\otimes n}), \tag{23}$$

where

$$S_N^{\max}(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) = \sup_{\mathrm{tr}(a^\dagger a\varrho)\leqslant N} S(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho))$$

and $S^{\min}(\mathcal{E}) = \inf_\varrho S(\mathcal{E}(\varrho))$ is the minimum output entropy. For any n-mode state ϱ_n, by the entropy power inequality we have that

$$\begin{aligned}\frac{1}{n}S\left(\mathcal{E}^{\otimes n}_{\lambda,\sigma_{\mathrm{E}}}(\varrho_n)\right) &\geqslant \log\left(\lambda \mathrm{e}^{S(\varrho_n)/n} + (1-\lambda)\mathrm{e}^{S\left(\sigma_{\mathrm{E}}^{\otimes n}\right)/n}\right)\\ &\geqslant \log\left(\lambda + (1-\lambda)\mathrm{e}^{S(\sigma_{\mathrm{E}})}\right)\end{aligned} \tag{24}$$

by the entropy power inequality (13) since $S(\varrho) \geqslant 0$ for any state ϱ. This is a useful bound on the second term in equation (23). In order to find a bound on the first term, let us consider the mean photon number of the output state $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho)$, where the input ϱ has mean photon number bounded by N. It can be bounded as

$$\begin{aligned}\mathrm{tr}\left(a^\dagger a \mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho)\right) &= \lambda \mathrm{tr}\left(a^\dagger a \varrho\right) + (1-\lambda)\mathrm{tr}\left(a_{\mathrm{E}}^\dagger a_{\mathrm{E}} \sigma_{\mathrm{E}}\right)\\ &\leqslant \lambda N + (1-\lambda)N_{\mathrm{E}}\ .\end{aligned} \tag{25}$$

Thus the output entropy is bounded as

$$S_N^{\max}\left(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho)\right) \leqslant g\big(\lambda N + (1-\lambda)N_{\mathrm{E}}\big) \tag{26}$$

by the maximum entropy principle. Combining equations (23), (24) and (26), the upper bound (21) follows.

The lower bound. To show (20), recall that taking the one-shot expression of the Holevo quantity and plugging in a specific ensemble of signal states $\{p_x, \varrho_x\}_x$ gives a lower bound on the capacity. We pick a Gaussian ensemble of coherent states, $\{\frac{1}{2\pi N}\mathrm{e}^{-\frac{|\xi|^2}{2N}}\mathrm{d}^2\xi, |\xi\rangle\langle\xi|\}_\xi$. Note that the ensemble average $\overline{\varrho} = \frac{1}{2\pi N}\int \mathrm{e}^{-\frac{|\xi|^2}{2N}}|\xi\rangle\langle\xi|\mathrm{d}^2\xi$ is the Gaussian thermal state $\varrho_{\mathrm{th},N}$ with mean photon number N. Therefore $S(\overline{\varrho}) = g(N)$. A lower bound on the classical capacity is thus given by

$$\begin{aligned}C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) &\geqslant \chi_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}})\\ &\geqslant S\big(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho_{\mathrm{th},N})\big) - \frac{1}{2\pi N}\int \mathrm{e}^{-\frac{|\xi|^2}{2N}} S\big(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(|\xi\rangle\langle\xi|)\big)\mathrm{d}^2\xi\ .\end{aligned} \tag{27}$$

We can lower bound the first term as follows: we have

$$S\big(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho_{\mathrm{th},N})\big) = S\big(\mathcal{E}_{1-\lambda,\varrho_{\mathrm{th},N}}(\sigma_{\mathrm{E}})\big) \geqslant g\big(\lambda N + (1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\big)\ . \tag{28}$$

The first equality follows because for general states ϱ, σ, we have $\mathcal{E}_{\lambda,\sigma}(\varrho) = \mathcal{E}_{1-\lambda,\varrho}(\sigma)$, as can be seen by considering the characteristic function of the output state under a beamsplitter [7]: it satisfies $\chi_{\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(\varrho)}(\xi) = \chi_\varrho(\sqrt{\lambda}\xi)\cdot\chi_{\sigma_{\mathrm{E}}}(\sqrt{1-\lambda}\xi) = \chi_{\mathcal{E}_{1-\lambda,\varrho}(\sigma_{\mathrm{E}})}(\xi)$ for all $\xi \in \mathbb{R}^2$. The inequality in equation (28) follows from the lower bound (11) on the output entropy of the phase-covariant Gaussian channel $\mathcal{E}_{1-\lambda,\varrho_{\mathrm{th},N}}$ for fixed input entropy $S(\sigma_{\mathrm{E}})$.

To bound the second term in (27), observe that by the maximum entropy principle, we have

$$\begin{aligned}S(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(|\xi\rangle\langle\xi|)) &\leqslant S([\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}(|\xi\rangle\langle\xi|)])\\ &= S\big(\mathcal{E}_{\lambda,[\sigma_{\mathrm{E}}]}(|\xi\rangle\langle\xi|)\big)\\ &= S\big(\mathcal{E}_{\lambda,[\sigma_{\mathrm{E}}]}(|0\rangle\langle 0|)\big)\\ &\leqslant \sup_{\substack{\sigma_{\mathrm{E}}\ \text{Gaussian}\\ \mathrm{tr}(a_{\mathrm{E}}^\dagger a_{\mathrm{E}}\sigma_{\mathrm{E}})\leqslant N_{\mathrm{E}}}} S\big(\mathcal{E}_{\lambda,\sigma_E}(|0\rangle\langle 0|)\big) =: A(\lambda, N_E)\ .\end{aligned}$$

The first identity holds because both expressions $[\mathcal{E}_{\lambda,\sigma_{\rm E}}(|\xi\rangle\langle\xi|)]$ and $\mathcal{E}_{\lambda,[\sigma_{\rm E}]}(|\xi\rangle\langle\xi|)$ define the same Gaussian state, as can be verified by computing the associated covariance matrices. The second identity is a consequence of the compatibility of the beamsplitter with displacements [7, lemma VI.1] and invariance of the von Neumann entropy under unitaries.

It remains to find an upper bound on $A(\lambda, N_{\rm E})$. In order to find such an upper bound, we consider the mean photon number at the output and apply equation (25), obtaining

$$\mathrm{tr}\left(a^\dagger a\mathcal{E}_{\lambda,\sigma_{\rm E}}(|0\rangle\langle 0|)\right) \leqslant (1-\lambda)N_{\rm E} \ .$$

Hence by the maximum entropy principle, we have that

$$A(\lambda, N_{\rm E}) \leqslant g\big((1-\lambda)N_{\rm E}\big) \ . \tag{29}$$

Combining equations (27)–(29), the lower bound (20) follows.

The difference between the upper and lower bound. The difference between the upper and lower bound is given by

$$\Delta(\mathcal{E}_{\lambda,\sigma_{\rm E}})(N) = \delta(N) + g\big((1-\lambda)N_{\rm E}\big) - \log\left(\lambda + (1-\lambda)\mathrm{e}^{S(\sigma_{\rm E})}\right) \tag{30}$$

where

$$\delta(N) \ := \ g\big(\lambda N + (1-\lambda)N_{\rm E}\big) - g\big(\lambda N + (1-\lambda)N_{\rm E}^{\rm ep}\big)$$

collects the terms depending on N. Since $g'(N) = \log(N+1) - \log(N)$ is strictly decreasing, $N_{\rm E}^{\rm ep} \leqslant N_{\rm E}$ by the maximum entropy principle, and the monotonicity of g, we have

$$\delta'(N) = \lambda\left[g'\big(\lambda N + (1-\lambda)N_{\rm E}\big) - g'\big(\lambda N + (1-\lambda)N_{\rm E}^{\rm ep}\big)\right] \leqslant 0$$

and δ is decreasing as well. Therefore, we have

$$\delta(N) \leqslant \delta(0) = g\big((1-\lambda)N_{\rm E}\big) - g\big((1-\lambda)N_{\rm E}^{\rm ep}\big) \ .$$

Inserting this into (30), we finally obtain the bound (22), as claimed. □

We now turn to the classical noise channel $\mathcal{F}_{t,f}$ introduced in equation (7). The following lemma immediately implies the second inequality in theorem 1. The bounds given in this lemma are illustrated in figure 2.

Lemma 1B. *For the classical capacity of the single-mode classical noise channel $\mathcal{F}_{t,f}$, we have the bounds*

$$C_N(\mathcal{F}_{t,f}) \geqslant \log\left(\mathrm{e}^{g(N)} + t\mathrm{e}^{H(f)}\right) - g(\pi t\mathrm{E}(f)), \tag{31}$$

$$C_N(\mathcal{F}_{t,f}) \leqslant g\big(N + \pi t\mathrm{E}(f)\big) - \log\left(1 + t\mathrm{e}^{H(f)}\right) \ . \tag{32}$$

The lower bound (31) is achievable with a coherent state ensemble. The difference between this upper and lower bound is bounded by

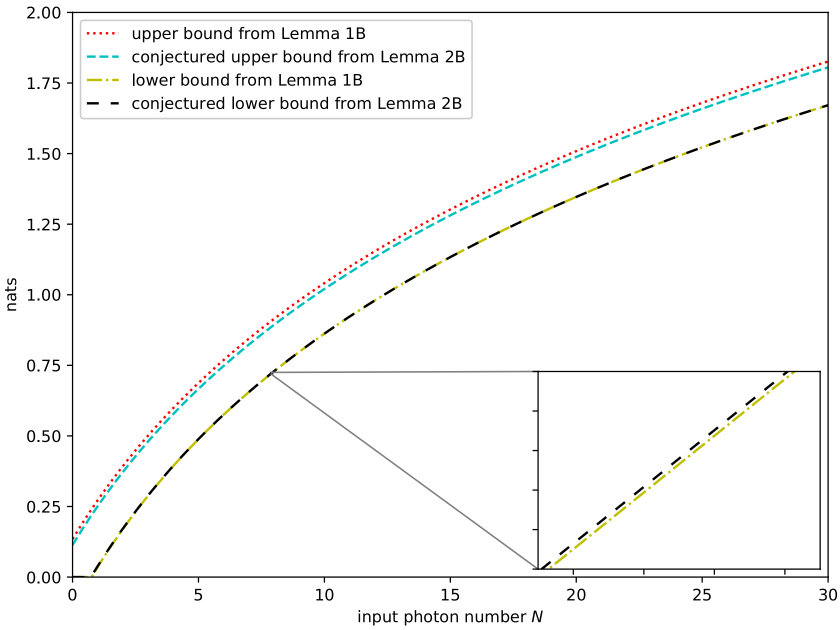


Figure 2. Bounds on the capacity $C_N(\mathcal{F}_{t,f})$ with $t = 1$ for any distribution f satisfying $\mathrm{E}(f) = 2$ and $\mathrm{e}^{H(f)} \simeq 15.1 < 2\pi e$.

$$\Delta(\mathcal{F}_{t,f}) \leqslant 2g\left(\pi t\mathrm{E}(f)\right) - \log\left(1 + t\mathrm{e}^{H(f)}\right),$$

independently of the input photon number N.

Proof of lemma 1B. *The upper bound*. To obtain the upper bound (32), we again bound the Holevo quantity by

$$C_N(\mathcal{F}_{t,f}) \leqslant S_N^{\max}(\mathcal{F}_{t,f}) - \lim_{n\to\infty} \frac{1}{n} S^{\min}(\mathcal{F}_{t,f}^{\otimes n}) . \tag{33}$$

By the entropy power inequality (15) we have

$$\frac{1}{n} S^{\min}(\mathcal{F}_{t,f}^{\otimes n}) \geqslant \log\left(1 + t\mathrm{e}^{H(f)}\right) .$$

This is because $\mathrm{e}^{S(\varrho)/n} \geqslant 1$ for all n-mode states ϱ, and because

$$\mathcal{F}_{t,f}^{\otimes n}(\varrho_n) = \int \tilde{f}^{(n)}(\xi) W(\sqrt{t}\xi) \varrho_n W(\sqrt{t}\xi)^\dagger \mathrm{d}^{2n}\xi$$

for the probability density function

$$\tilde{f}^{(n)}(\xi) = \Pi_{i=1}^n f(\xi_{2i-1}, \xi_{2i}) = \Pi_{i=1}^n f(q_i, p_i) \qquad \text{for } \xi = (\xi_1, \ldots, \xi_{2n}) \in \mathbb{R}^n$$

on $\mathbb{R}^{2n}$, which has Shannon entropy $H(\tilde{f}^{(n)}) = nH(f)$.

To bound the first term in (33), we again use the maximum entropy principle to obtain

$$\begin{aligned} S_N^{\max}(\mathcal{F}_{t,f}) &= \sup_{\operatorname{tr}(a^\dagger a\varrho)\leqslant N} S(\mathcal{F}_{t,f}(\varrho)) \\ &\leqslant \sup_{\operatorname{tr}(a^\dagger a\varrho)\leqslant N} S([\mathcal{F}_{t,f}(\varrho)]) \\ &= \sup_{\operatorname{tr}(a^\dagger a\varrho)\leqslant N} S\big(\mathcal{F}_{t,[f]}([\varrho])\big) \\ &= g(N+\pi t\mathrm{E}(f)), \end{aligned}$$

giving the claimed upper bound. The last step follows because for Gaussian f and ϱ, it is easy to directly evaluate the behavior of the mean photon number under the channel and conclude that we can replace $\sup_{\operatorname{tr}(a^\dagger a\varrho)\leqslant N} S([\mathcal{F}_{t,f}(\varrho)])$ with $\sup_{\operatorname{tr}(a^\dagger a\varrho)\leqslant N+\pi t\mathrm{E}(f)} S([\varrho])$.

The lower bound. For the lower bound (31) we use a Gaussian ensemble $\{g(\xi)\mathrm{d}^2\xi, |\xi\rangle\langle\xi|\}_\xi$ (where $g(\xi)=\frac{1}{2\pi N}\mathrm{e}^{-\frac{|\xi|^2}{2N}}$) of displaced coherent states with mean photon number (of the ensemble) N. Then the ensemble average is $\varrho_{\mathrm{th},N}=\frac{1}{2\pi N}\int\mathrm{e}^{-\frac{|\xi|^2}{2N}}|\xi\rangle\langle\xi|d\xi$ and we obtain

$$\begin{aligned} C_N(\mathcal{F}_{t,f}) \geqslant \chi_N(\mathcal{F}_{t,f}) &\geqslant S\big(\mathcal{F}_{t,f}(\varrho_{\mathrm{th},N})\big) - \int g(\xi)S\big(\mathcal{F}_{t,f}(|\xi\rangle\langle\xi|)\big)\mathrm{d}^2\xi \\ &\geqslant \log\left(\mathrm{e}^{S(\varrho_{\mathrm{th},N})}+t\mathrm{e}^{H(f)}\right) - S\big(\mathcal{F}_{t,f}(|0\rangle\langle 0|)\big) \\ &\geqslant \log\left(\mathrm{e}^{g(N)}+t\mathrm{e}^{H(f)}\right) - g\big(\pi t\mathrm{E}(f)\big), \end{aligned}$$

where we have used the entropy power inequality (15) and invariance of the von Neumann entropy of a state under unitary conjugation, as well as compatibility of the classical noise channel with displacements [9, lemma 2] in the first step and the fact that $\varrho_{\mathrm{th},N}$ is a Gaussian thermal state with mean photon number N in the second step.

The difference between the upper and lower bounds. We again write the difference between the upper and lower bound as

$$\Delta(\mathcal{F}_{t,f})(N) = \delta(N) - \log\left(1+t\mathrm{e}^{H(f)}\right) + g\big(\pi t\mathrm{E}(f)\big),$$

where

$$\delta(N) = g\big(N+\pi t\mathrm{E}(f)\big) - \log\left(\mathrm{e}^{g(N)}+t\mathrm{e}^{H(f)}\right) .$$

We have

$$\begin{aligned} \log\left(\mathrm{e}^{g(N)}+t\mathrm{e}^{H(f)}\right) &= \log\left[\mathrm{e}^{g(N)}\left(1+t\mathrm{e}^{H(f)-g(N)}\right)\right] \\ &= g(N) + \log\left(1+t\mathrm{e}^{H(f)-g(N)}\right) \geqslant g(N), \end{aligned}$$

and thus

$$\delta(N) \leqslant g\big(N+\pi t\mathrm{E}(f)\big) - g(N) .$$

Now an adaptation of the argument given in the proof of lemma 1A proves the claimed statement: the rhs is monotonically decreasing in N, and thus maximal for $N = 0$: this yields $\delta(N) \leqslant g\big(\pi t\mathrm{E}(f)\big)$, which implies the claim. □

5.3. Derivation of strengthened capacity bounds from conjectured EPNIs

We now sketch how to obtain theorem 2, focusing only on the differences in the derivation between theorems 1 and 2. Again, we first treat the case of the attenuation channel $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}$. The classical noise channel $\mathcal{F}_{t,f}$ is then discussed in lemma 2B below.

Lemma 2A. *Assuming the EPNI conjecture (16) holds, the classical capacity of $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}$ satisfies*

$$C_N(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leqslant g\big(\lambda N + (1-\lambda)N_{\mathrm{E}}\big) - g\big((1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\big)\ .$$

The difference between this upper bound and the lower bound (20) is bounded by

$$\Delta(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}) \leqslant 2\Big[g\big((1-\lambda)N_{\mathrm{E}}\big) - g\big((1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\big)\Big],$$

independently of the input photon number N.

We note that the EPNI does not improve upon the lower bound on the classical capacity of $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}$ given in lemma 1A. This is because the lower bound only requires a bound on the minimal output entropy of $\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}$ for thermal input states, a case where the statement of the EPNI is already covered by the established equation (11).

Proof. Lemma 2A follows when we use the EPNI (16) instead of the EPI (12) in the proof. In particular, we obtain from the EPNI that

$$\begin{aligned}\frac{1}{n}S\big(\mathcal{E}_{\lambda,\sigma_{\mathrm{E}}}^{\otimes n}(\varrho_n)\big) &\geqslant g\Big(\lambda g^{-1}\big[S(\varrho_n)/n\big] + (1-\lambda)g^{-1}\big[S(\sigma_{\mathrm{E}}^{\otimes n})/n\big]\Big)\\ &\geqslant g\big((1-\lambda)N_{\mathrm{E}}^{\mathrm{ep}}\big),\end{aligned}$$

since $S(\varrho) \geqslant 0$ for any state ϱ. This expression replaces the second term in the upper bound of lemma 1A. The bound on the difference between the upper and the lower bound in lemma 2A follows analogously to before. □

Employing the EPNI (17) instead of the EPI (15) in a similar fashion shows the second part of theorem 2 for the classical noise channel $\mathcal{F}_{t,f}$:

Lemma 2B. *Assuming equation (18) holds, the classical capacity of $\mathcal{F}_{t,f}$ satisfies*

$$C_N(\mathcal{F}_{t,f}) \geqslant g\left(N + \frac{t}{e}\mathrm{e}^{H(f)}\right) - g\big(\pi t\mathrm{E}(f)\big)\ .$$

Assuming in addition that the EPNI conjecture (17) holds, we further have

$$C_N(\mathcal{F}_{t,f}) \leqslant g\big(N + \pi t\mathrm{E}(f)\big) - g\left(\frac{t}{e}\mathrm{e}^{H(f)}\right)\ .$$

The difference between this upper and lower bound is bounded by

$$\Delta(\mathcal{F}_{t,f}) \leqslant 2\left[g\big(\pi t\mathrm{E}(f)\big) - g\left(\frac{t}{e}\mathrm{e}^{H(f)}\right)\right],$$

independently of the input photon number N.

Proof. Here we obtain

$$\frac{1}{n}S^{\min}(\mathcal{F}_{t,f}^{\otimes n}) \geqslant g\left(\frac{t}{e}\mathrm{e}^{H(f)}\right),$$

replacing the second term in the upper bound of lemma 1B. The first term in the lower bound is replaced by

$$S\big(\mathcal{F}_{t,f}(\varrho)\big) \geqslant g\left(N + \frac{t}{e}\mathrm{e}^{H(f)}\right),$$

replacing the first term in the lower bound of the lemma. The maximal difference between the upper and the lower bound follows again analogously to the reasoning from before. □

6. Discussion

We have derived bounds on the classical capacity of a general class of bosonic channels which includes both Gaussian as well as non-Gaussian examples. We have shown that for these channels, the rates achievable by modulation of coherent states are not too far away from the maximal achievable rate (using arbitrary, possibly entangled code states). The maximal gain resulting from the use of general coding strategies is bounded by a channel-dependent constant independently of the average energy of the signal states. In particular, this means that these channels can only exhibit a limited amount of additivity violation: the product-state and general classical capacities essentially coincide.

There are a few paths one could follow for future work: First, a proof of equation (18) would be desirable as a first step towards a proof of the classical-quantum EPNI (17), implying better bounds on the classical capacity of the channels considered here. One may also wonder about similar statements for the multi-mode versions of the considered channels. The two main ingredients of our proofs, the maximum entropy principle and the entropy power inequality, both hold in the multi-mode case. Furthermore, minimal output entropy results analogous to (11) have been shown in the multi-mode setting [24], essentially providing us with all tools required in the proof of our bounds. Therefore, similar bounds should hold. Another interesting question concerns the implications of the EPI and EPNI to other capacities than the classical capacity, such as the entanglement-assisted classical capacity. The recently proven conditional version of the EPI has been shown to imply an upper bound on the entanglement-assisted capacity [30, 38], but a corresponding lower bound for general non-Gaussian channels is missing so far.

Acknowledgments

We thank the anonymous referees for their comments. This work is supported by the Technische Universität München—Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763. The authors acknowledge additional support by DFG project no. K05430/1-1.

ORCID iDs

Stefan Huber https://orcid.org/0000-0002-0564-5436

References

[1] Schumacher B and Westmoreland M D 1997 Sending classical information via noisy quantum channels *Phys. Rev.* A **56** 131–8

[2] Holevo A S 1998 The capacity of the quantum channel with general signal states *IEEE Trans. Inf. Theory* **44** 269–73

[3] Caves C M and Drummond P D 1994 Quantum limits on bosonic communication rates *Rev. Mod. Phys.* **66** 481–537

[4] Giovannetti V, García-Patrón R, Cerf N J and Holevo A S 2014 Ultimate classical communication rates of quantum optical channels *Nat. Photon.* **8** 796–800

[5] Pilyavets O V, Lupo C and Mancini S 2012 Methods for estimating capacities and rates of gaussian quantum channels *IEEE Trans. Inf. Theory* **58** 6126–64

[6] Schäfer J, Karpov E and Cerf N J 2011 Gaussian capacity of the quantum bosonic memory channel with additive correlated gaussian noise *Phys. Rev.* A **84** 032318

[7] König R and Smith G 2014 The entropy power inequality for quantum systems *IEEE Trans. Inf. Theory* **60** 1536–48

[8] Palma G D, Mari A, Lloyd S and Giovannetti V 2015 Multimode quantum entropy power inequality *Phys. Rev.* A **91** 032320

[9] Huber S, König R and Vershynina A 2017 Geometric inequalities from phase space translations *J. Math. Phys.* **58** 012206

[10] Palma G D, Trevisan D and Giovannetti V 2016 Passive states optimize the output of bosonic gaussian quantum channels *IEEE Trans. Inf. Theory* **62** 2895–906

[11] Palma G D, Trevisan D and Giovannetti V 2016 One-mode quantum-limited Gaussian channels have Gaussian maximizers (arXiv:1610.09967)

[12] Palma G D, Trevisan D and Giovannetti V 2017 Gaussian states minimize the output entropy of the one-mode quantum attenuator *IEEE Trans. Inf. Theory* **63** 728–37

[13] Palma G D, Trevisan D and Giovannetti V 2017 Gaussian states minimize the output entropy of one-mode quantum gaussian channels *Phys. Rev. Lett.* **118** 160503

[14] Lupo C, Pilyavets O V and Mancini S 2009 Capacities of lossy bosonic channel with correlated noise *New J. Phys.* **11** 063023

[15] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423

[16] Hastings M B 2009 Superadditivity of communication capacity using entangled inputs *Nat. Phys.* **5** 255–7

[17] Holevo A S and Werner R F 2001 Evaluating capacities of bosonic Gaussian channels *Phys. Rev.* A **63** 032312

[18] König R and Smith G 2013 Limits on classical communication from quantum entropy power inequalities *Nat. Photon.* **7** 142–6

[19] König R and Smith G 2013 Classical capacity of quantum thermal noise channels to within 1.45 bits *Phys. Rev. Lett.* **110** 040501

[20] Giovannetti V, Lloyd S, Maccone L and Shapiro J H 2013 Electromagnetic channel capacity for practical purposes *Nat. Photon.* **7** 834–8

[21] Werner R 1984 Quantum harmonic analysis on phase space *J. Math. Phys.* **25** 1404–11

[22] Datta N, Pautrat Y and Rouzé C 2017 Contractivity properties of a quantum diffusion semigroup *J. Math. Phys.* **58** 012205

[23] Giovannetti V, Guha S, Lloyd S, Maccone L and Shapiro J H 2004 Minimum output entropy of bosonic channels: a conjecture *Phys. Rev.* A **70** 032315

[24] Holevo A S 2015 Gaussian optimizers and the additivity problem in quantum information theory *Russ. Math. Surv.* **70** 331–67

[25] Giovannetti V, Holevo A S and García-Patrón R 2014 A solution of Gaussian optimizer conjecture for quantum channels *Commun. Math. Phys.* **334** 1553–71

[26] Eisert J and Wolf M M 2007 *Gaussian Quantum Channels* (*Quantum Information with Continuous Variables of Atoms and Light*) (London: Imperial College Press) pp 23–42

[27] García-Patrón R, Navarrete-Benlloch C, Lloyd S, Shapiro J H and Cerf N J 2014 The holy grail of quantum optical communication *AIP Conf. Proc.* **1633** 109

[28] Holevo A S 2016 On the constrained classical capacity of infinite-dimensional covariant quantum channels *J. Math. Phys.* **57** 015203

[29] Wolf M M, Giedke G and Cirac J I 2006 Extremality of Gaussian quantum states *Phys. Rev. Lett.* **96** 080502

[30] Palma G D and Trevisan D 2018 The Conditional entropy power inequality for bosonic quantum systems *Commun. Math. Phys.* (https://doi.org/10.1007/s00220-017-3082-8)

[31] Agarwal G S 1971 Entropy, the Wigner distribution function, and the approach to equilibrium of a system of coupled Harmonic oscillators *Phys. Rev.* A **3** 828–31

[32] Frank R L and Lieb E H 2017 Norms of quantum gaussian multi-mode channels *J. Math. Phys.* **58** 062204

[33] Holevo A S 2017 On quantum Gaussian optimizers conjecture in the case $p = q$ *Russ. Math. Surv.* **72** 1177

[34] Shannon C E 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 623–56

[35] Stam A J 1959 Some inequalities satisfied by the quantities of information of Fisher and Shannon *Inf. Control* **2** 101–12

[36] Blachman N 1965 The convolution inequality for entropy powers *IEEE Trans. Inf. Theor.* **11** 267–71

[37] Lieb E H 1978 Proof of an entropy conjecture of Wehrl *Commun. Math. Phys.* **62** 35–41

[38] König R 2015 The conditional entropy power inequality for Gaussian quantum states *J. Math. Phys.* **56** 022201

[39] Guha S, Erkmen B and Shapiro J 2008 The entropy photon-number inequality and its consequences *Information Theory and Applications Workshop* pp 128–30

[40] Guha S, Shapiro J and Erkmen B 2008 Capacity of the bosonic wiretap channel and the entropy photon-number inequality *IEEE Int. Symp. on Information Theory* pp 91–5

[41] Das S, Sharma N and Muthukrishnan S 2013 On some special cases of the entropy photon-number inequality *Theory of Quantum Computation, Communication, and Cryptography* (Berlin: Springer) pp 116–27

[42] Guha S, Shapiro J H and Erkmen B I 2007 Classical capacity of bosonic broadcast communication and a minimum output entropy conjecture *Phys. Rev.* A **76** 032303

[43] Giovannetti V, Guha S, Lloyd S, Maccone L, Shapiro J H and Yuen H P 2004 Classical capacity of the lossy bosonic channel: the exact solution *Phys. Rev. Lett.* **92** 027902

A.3 The conditional entropy power inequality for quantum additive noise channels

The conditional entropy power inequality for quantum additive noise channels

Giacomo De Palma and Stefan Huber

We prove the quantum conditional entropy power inequality for quantum additive noise channels. This inequality lower bounds the quantum conditional entropy of the output of an additive noise channel in terms of the quantum conditional entropies of the input state and the noise when they are conditionally independent given the side information. We also show that this conditional entropy power inequality is optimal in the sense that we can achieve equality asymptotically by choosing a suitable sequence of Gaussian input states. We apply the conditional entropy power inequality to find an array of information-theoretic inequalities for conditional entropies which are the analogues of inequalities which have already been established in the setting without side information. Furthermore, we give a simple proof of the convergence rate of the quantum Ornstein-Uhlenbeck semigroup based on entropy power inequalities.

A.3.1 Main Results

We consider the "classical-quantum" convolution operation from Eq. (4.11) as a model for quantum additive noise channels. Our main result is the following conditional entropy power inequality for this class of channels:

Theorem A.3.1 (Conditional entropy power inequality for the convolution (4.11))**.** *Let A be an n-mode quantum system, R a classical system and M a generic quantum system. Let ρ_{ARM} be a quantum state on ARM such that its marginal on R has a probability density function $\rho_R : \mathbb{R}^{2n} \to \mathbb{R}$. Let ρ_{ARM} further satisfy*

$$\mathrm{tr}\left(\sum_{k=1}^{n} a_k^\dagger a \rho_A\right) < \infty, \qquad \mathrm{E}(\rho_R) < \infty, \qquad S(\rho_M) < \infty \ .$$

Let us suppose that A and R are conditionally independent given M, i.e.,

$$I(A:R|M)_{\rho_{ARM}} = 0 \ ,$$

and let

$$\rho_{CM} := (\mathcal{E} \otimes \mathbb{1}_M)(\rho_{ARM}) = \int_{\mathbb{R}^{2n}} D(\xi)\rho_{AM|R=\xi} D(\xi)^\dagger \rho_R(\xi) \frac{\mathrm{d}^{2n}\xi}{(2\pi)^n} \ .$$

Then, for any $0 \leq \lambda \leq 1$ the linear conditional entropy power inequality holds:

$$\frac{S(C|M)}{n} \geq \lambda \frac{S(A|M)}{n} + (1-\lambda)\frac{S(R|M)}{n} - \lambda \log \lambda - (1-\lambda)\log(1-\lambda) \ .$$

In particular, we obtain the conditional entropy power inequality for the convolution (4.11)*:*

$$\exp \frac{S(C|M)}{n} \geq \exp \frac{S(A|M)}{n} + \exp \frac{S(R|M)}{n} \ .$$

If the classical system R is uncorrelated with the system M, we have the inequality

$$\exp \frac{S(C|M)}{n} \geq \exp \frac{S(A|M)}{n} + \exp \frac{S(\rho_R)}{n} \ .$$

Remarkably, this conditional entropy power inequality is optimal in the following sense:

Theorem A.3.2 (Optimality of the conditional entropy power inequality)**.** *For any $a, b \in \mathbb{R}$ there exists a sequence of states $\left\{\rho_{AM}^{(k)}\right\}_{k\in\mathbb{N}}$ and a probability density function $f : \mathbb{R}^2 \to \mathbb{R}$ such that the classical system R is uncorrelated with M and*

$$\lim_{k\to\infty} S(A|M)_{\rho_{AM}^{(k)}} = a, \qquad S(R|M)_f = b \ ,$$

as well as

$$\lim_{k\to\infty} \exp S(C|M)_{\rho_{CM}^{(k)}} = \exp a + \exp b \ .$$

where $\rho_{CM}^{(k)} = (\mathcal{E}_f \otimes \mathbb{1}_M)(\rho_{AM})$ with $\mathcal{E}_f(\rho_A) = f \star \rho_A$.

A.3.2 Applications

In this section, let again A be an n-mode quantum system, M a generic quantum system and let ρ_{AM} be such that $\operatorname{tr}\left(\sum_{k=1}^n a_k^\dagger a \rho_A\right) < \infty$ and $S(\rho_M) < \infty$.

Lemma A.3.3 (Quantum conditional Fisher information isoperimetric inequality)**.**

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{n} J(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}\right]^{-1}\Bigg|_{t=0} \geq 1 \ ,$$

where $\mathcal{N}(t)(\rho_A) = f_Z \star_t \rho_A$ is the diffusion semigroup.

Theorem A.3.4 (Isoperimetric inequality for quantum conditional entropies)**.**

$$\frac{1}{n} J(A|M)_{\rho_{AM}} \exp\frac{S(A|M)_{\rho_{AM}}}{n} \geq e \ .$$

Theorem A.3.5 (Concavity of the quantum conditional entropy power along the heat flow)**.**

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \exp\frac{S(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}}{n}\Bigg|_{t=0} \leq 0 \ .$$

Versions of these inequalities in the setting without side information were previously obtained in Core Article I [1]. The proofs therein exhibited some regularity issues regarding the definition of the Fisher information. These issues are lifted by the proofs in this article, and choosing M to be trivial, we recover these inequalities for the setting of no side information.

A.3.2.1 Upper bound on the entanglement-assisted capacity of classical noise channels

Let A and C be n-mode bosonic quantum systems, $f : \mathbb{R}^{2n} \to \mathbb{R}$ be a (possibly non-Gaussian) probability density function with finite second moments, and consider again the classical noise channel

$$\begin{aligned} \mathcal{E}_f :\ & \mathcal{S}(\mathcal{H}_A) \to \mathcal{S}(\mathcal{H}_C) \ , \\ & \rho_A \mapsto \mathcal{E}_f(\rho_A) = f \star \rho_A \ . \end{aligned}$$

Write $E_0 := \frac{E(f)}{2n}, S_0 := \frac{H(f)}{n}$ for the average energy and entropy per mode of f. Then the *energy-constrained entanglement-assisted classical capacity* of the channel $\mathcal{E}_f$ is defined as [32–34]

$$C_{\text{ea}}(\mathcal{E}_f) = \sup \left\{ I(C:M)_{(\mathcal{E}_f \otimes \mathbb{1}_M)(\rho_{AM})} : \rho_{AM} \text{ pure, } \operatorname{tr}\left(\sum_{k=1}^{n} a_k^\dagger a_k \rho_A \right) \le nE \right\} ,$$

where the energy constraint $\operatorname{tr}\left(\sum_{k=1}^{n} a_k^\dagger a_k \rho_A\right) \le nE$ physically means that the sender can only use states of a finite average energy E per mode. Then we prove that

$$C_{\text{ea}}(\mathcal{E}) \le ng(E + E_0) - n \log\left(e^{-g(E)} + e^{S_0}\right) .$$

A.3.2.2 Fast convergence of the quantum Ornstein-Uhlenbeck semigroup

Consider a one-mode bosonic quantum system A and the quantum Ornstein-Uhlenbeck semigroup $\{\mathcal{P}^{(\mu,\lambda)}(t) = e^{t\mathcal{L}_{\mu,\lambda}}\}_{t\ge 0}$ defined in (A.1). Conjecture A.1.7 was proven shortly after the publication of Core Article I [1] in the work [7], using as tools newly developed methods in gradient flow. We give a short proof of the generalization of Conjecture A.1.7 to the setting with side information using the entropy power inequality, which directly implies Conjecture A.1.7 and provides an arguably simpler proof for it than the one which was previously known.

Theorem A.3.6. *We have for any quantum state ρ_{AM}*

$$D\left(\left(\mathcal{P}^{\mu,\lambda}(t) \otimes \mathbb{1}_M\right)(\rho_{AM}) \middle\| \rho_{\text{th},\frac{\lambda^2}{\mu^2-\lambda^2}} \otimes \rho_M\right) \le e^{-(\mu^2-\lambda^2)t} D\left(\rho_{AM} \middle\| \rho_{\text{th},\frac{\lambda^2}{\mu^2-\lambda^2}} \otimes \rho_M\right) ,$$

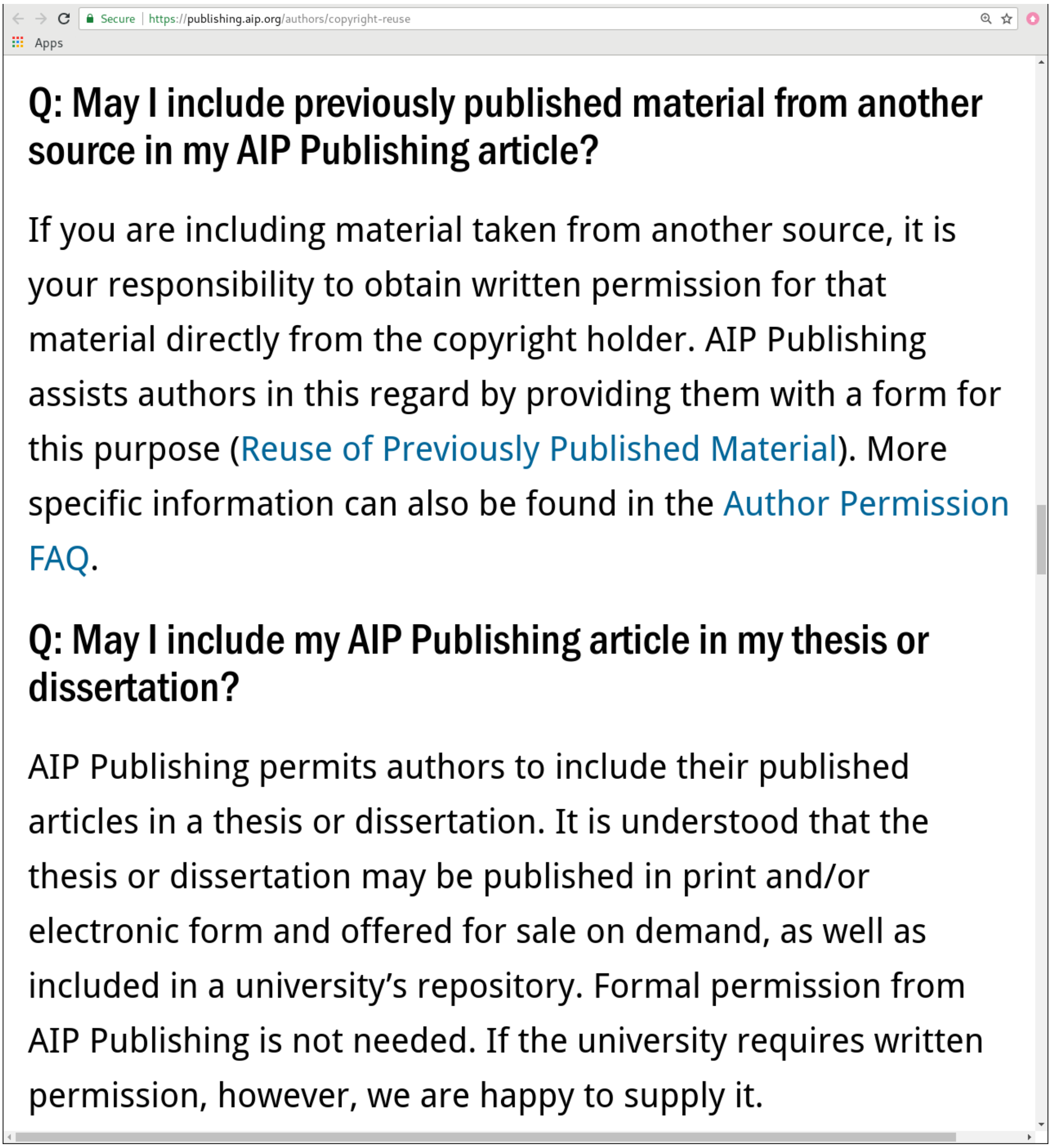
where $\rho_M = \operatorname{tr}_A(\rho_{AM})$ is the marginal state of ρ_{AM} on the system M.

A.3.3 Individual Contribution

I am the principal author of this article. The start of this project happened during a visit of Giacomo De Palma in Munich during a discussion of possible new entropic inequalities similar in spirit with recent results [13]. We worked out a sketch of proof of Theorem A.3.1 and the rough structure of auxiliary Lemmas together during this visit. I was then responsible for writing all sections of the article and completing the proofs. As an exception to this, Giacomo De Palma was responsible for proving Lemma 3, and the main idea for Theorem 9 is also due to him.

Permission to include:

Giacomo De Palma and Stefan Huber
The conditional entropy power inequality for quantum additive noise channels.
J. Math. Phys. 59, 122201 (2018).



Q: May I include previously published material from another source in my AIP Publishing article?

If you are including material taken from another source, it is your responsibility to obtain written permission for that material directly from the copyright holder. AIP Publishing assists authors in this regard by providing them with a form for this purpose (Reuse of Previously Published Material). More specific information can also be found in the Author Permission FAQ.

Q: May I include my AIP Publishing article in my thesis or dissertation?

AIP Publishing permits authors to include their published articles in a thesis or dissertation. It is understood that the thesis or dissertation may be published in print and/or electronic form and offered for sale on demand, as well as included in a university's repository. Formal permission from AIP Publishing is not needed. If the university requires written permission, however, we are happy to supply it.

The conditional entropy power inequality for quantum additive noise channels

Giacomo De Palma[1] and Stefan Huber[2]
[1] *QMATH, Department of Mathematical Sciences, University of Copenhagen, 2100 Copenhagen, Denmark*
[2] *Institute for Advanced Study and Zentrum Mathematik, Technical University of Munich, 85748 Garching, Germany*

(Received 2 March 2018; accepted 10 November 2018; published online 3 December 2018)

We prove the quantum conditional entropy power inequality for quantum additive noise channels. This inequality lower bounds the quantum conditional entropy of the output of an additive noise channel in terms of the quantum conditional entropies of the input state and the noise when they are conditionally independent given the memory. We also show that this conditional entropy power inequality is optimal in the sense that we can achieve equality asymptotically by choosing a suitable sequence of Gaussian input states. We apply the conditional entropy power inequality to find an array of information-theoretic inequalities for conditional entropies which are the analogs of inequalities which have already been established in the unconditioned setting. Furthermore, we give a simple proof of the convergence rate of the quantum Ornstein-Uhlenbeck semigroup based on entropy power inequalities. *Published by AIP Publishing.* https://doi.org/10.1063/1.5027495

I. INTRODUCTION

Additive noise channels are central objects of interest in information theory. A general class of such channels can be modeled by the well-known convolution operation: If X and Y are two independent random variables with values in $\mathbb{R}^k$, the convolution operation $(X, Y) \mapsto X + Y$ combines X and Y into a new random variable $X + Y$, the probability density function of which is given by

$$f_{X+Y}(z) := \int_{\mathbb{R}^k} f_X(z-x) f_Y(x)\, \mathrm{d}^k x. \tag{1}$$

The convolution is a well-studied operation, and it plays a role in many inequalities from functional analysis, such as Young's inequality and its sharp version[1,2] as well as the entropy power inequality.[3–6] These inequalities have important applications in classical information theory, as they can be used to bound communication capacities, which was originally carried out by Shannon.[3] An extensive overview of the many related inequalities in this area is given in Ref. 6.

Central to the work presented here is the entropy power inequality. It deals with the entropy of a linear combination of two independent random variables X and Y with values in $\mathbb{R}^k$,

$$Z := \sqrt{\lambda} X + \sqrt{|1-\lambda|}\, Y, \qquad \lambda \geq 0.$$

The statement of the entropy power inequality[3–6] is

$$\exp \frac{2S(Z)}{k} \geq \lambda \exp \frac{2S(X)}{k} + |1-\lambda| \exp \frac{2S(Y)}{k}, \tag{2}$$

where $S(X)$ is the Shannon differential entropy of the random variable X. A conditional version of (2) can easily be derived: If X and Y are conditionally independent given the random variable M (sometimes interpreted as a memory), then

$$\exp \frac{2S(Z|M)}{k} \geq \lambda \exp \frac{2S(X|M)}{k} + |1-\lambda| \exp \frac{2S(Y|M)}{k}.$$

In quantum information theory, an analogous operation to the convolution (1) is given by the action of a beam splitter U_λ with transmissivity $0 \leq \lambda \leq 1$ on a quantum state (i.e., a linear positive operator with a unit trace) ρ_{AB} which is bipartite on two n-mode Gaussian quantum systems A, B. This action has the form

$$\rho_{AB} \mapsto \rho_C = \mathrm{tr}_2\Big(U_\lambda \rho_{AB} U_\lambda^\dagger\Big), \tag{3}$$

where C is again an n-mode quantum system and tr_2 denotes the partial trace over the second system. The mathematical motivation of the study of this operation is that in the special case of a product state, that is, $\rho_{AB} = \rho_A \otimes \rho_B$, it is formally similar to the convolution described in (1) on the level of Wigner functions. For the beam splitter (3), several important inequalities in the same spirit as in classical information theory have been established.[7–11] For instance, the quantum entropy power inequality reads

$$\exp\frac{S(C)}{n} \geq \lambda \exp\frac{S(A)}{n} + (1-\lambda)\exp\frac{S(B)}{n}, \tag{4}$$

with $S(A) = S(\rho_A) = -\mathrm{tr}[\rho_A \log \rho_A]$ being the von Neumann entropy of a quantum state. Unlike in the classical setting, a conditional entropy power inequality for the operation (3) does not trivially follow from the unconditioned inequality (4). However, it was recently established in Ref. 11 that such an inequality holds nonetheless: For a joint quantum state ρ_{ABM} such that A and B are conditionally independent given the memory system M, we have

$$\exp\frac{S(C|M)}{n} \geq \lambda \exp\frac{S(A|M)}{n} + (1-\lambda)\exp\frac{S(B|M)}{n},$$

where $S(X|M) := S(XM) - S(M)$ is the quantum conditional entropy. The conditional independence of A and B given M is expressed with the condition that the quantum conditional mutual information equals zero,

$$I(A:B|M) := S(A|M) + S(B|M) - S(AB|M) = 0.$$

Our work concerns yet another convolution operation, which mixes a probability density function $f:\mathbb{R}^{2n}\to\mathbb{R}$ on phase space with an n-mode quantum state ρ,

$$(f,\rho)\mapsto f\star\rho, \qquad \text{where} \qquad f\star\rho = \int_{\mathbb{R}^{2n}} f(\xi)D(\xi)\rho D(\xi)^\dagger \frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}, \tag{5}$$

where $D(\xi)$ are the Weyl displacement operators in phase space. This operation was first introduced by Werner.[12] Werner established a number of results regarding (5), most notably a Young-type inequality. In Ref. 13, more inequalities involving this operation were shown, most prominently the entropy power inequality

$$\exp\frac{S(f\star\rho)}{n} \geq \exp\frac{S(f)}{n} + \exp\frac{S(\rho)}{n}. \tag{6}$$

In the context of mixing times of semigroups, the authors of Ref. 14 have used this convolution extensively and proved various properties which are related to the discussion of the entropy power inequality.

A. Our contribution

Similar to the work carried out in Ref. 11 for the beam splitter, we prove the conditional version of the entropy power inequality for the convolution given by (5). Let us consider an n-mode Gaussian quantum system A, a generic quantum system M, and a classical system R which "stores" a classical probability density function $\rho_R:\mathbb{R}^{2n}\to\mathbb{R}$. Let us further consider the map $\mathcal{E}:AR\to C$, $(\rho_A\otimes\rho_R)\mapsto\rho_R\star\rho_A$, linearly extended to generic states ρ_{AR} as

$$\rho_C = \mathcal{E}(\rho_{AR}) = \int_{\mathbb{R}^{2n}} D(\xi)\,\rho_{A|R=\xi}\,D(\xi)^\dagger\,\rho_R(\xi)\,\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}.$$

We show in Theorem 5 that the conditional entropy of the output of $\mathcal{E}\otimes\mathbb{1}_M : ARM\to CM$ is lower bounded as

$$\exp\frac{S(C|M)}{n} \geq \exp\frac{S(A|M)}{n} + \exp\frac{S(R|M)}{n},$$

if $I(A\colon R|M) = 0$, i.e., the systems A and R are conditionally independent given the system M. As a special case, this inequality implies useful inequalities about the convolution (5) in the case when R is uncorrelated with M,

$$\exp\frac{S(C|M)}{n} \geq \exp\frac{S(A|M)}{n} + \exp\frac{S(\rho_R)}{n}.$$

In the particular case when R is a Gaussian random variable with probability density function $f_{Z,t} = \exp\left(-\frac{\|\xi\|^2}{2t}\right)/t^n$, the inequality becomes

$$\exp\frac{S(C|M)}{n} \geq \exp\frac{S(A|M)}{n} + et.$$

The special cases mentioned above are important in various applications, as we will show later.

This conditional entropy power inequality is tight in the sense that it is saturated for any couple of values of $S(A|M)$ and $S(R|M)$ by an appropriate sequence of Gaussian input states, which we show in Theorem 6. This behavior is similar to the case of the beam splitter. On the way to this inequality, several intermediate results are proven which make up a set of information-theoretic inequalities regarding conditional Fisher information and conditional entropies. To complete the picture of information-theoretic inequalities involving quantum conditional entropies, we apply our results to prove a number of additional inequalities in a spirit similar to the classical case. Among them, there are the concavity of the quantum conditional entropy along the heat flow (Theorem 8) and an isoperimetric inequality for quantum conditional entropies (Lemma 7). Furthermore, we show in Sec. VIII C how, similar to the case of the beam splitter, the conditional entropy power inequality implies a converse bound on the entanglement-assisted classical capacity of a non-Gaussian quantum channel, the classical noise channel defined in (5).

Another part of our work regards the quantum Ornstein-Uhlenbeck (qOU) semigroup. It is the one-parameter semigroup of completely positive and trace-preserving (CPTP) maps $\left\{\mathcal{P}^{(\mu,\lambda)}(t) = e^{t\mathcal{L}_{\mu,\lambda}}\right\}_{t\geq 0}$ on the one-mode Gaussian quantum system A generated by the Liouvillian

$$\mathcal{L}_{\mu,\lambda} = \mu^2\mathcal{L}_- + \lambda^2\mathcal{L}_+ \qquad \text{for } \mu > \lambda > 0,$$

where

$$\mathcal{L}_+(\rho) = a^\dagger\rho a - \frac{1}{2}\{aa^\dagger, \rho\} \qquad \text{and} \qquad \mathcal{L}_-(\rho) = a\rho a^\dagger - \frac{1}{2}\{a^\dagger a, \rho\},$$

where a is the ladder operator of A. This quantum dynamical semigroup has a unique fixed point given by

$$\omega^{\mu,\lambda} := \frac{\mu^2 - \lambda^2}{\mu^2}\sum_{k=0}^{\infty}\left(\frac{\lambda^2}{\mu^2}\right)^k |k\rangle\langle k|,$$

where $\{|k\rangle\}_{k\in\mathbb{N}}$ is the Fock basis of A. It has been shown in Ref. 15 using methods of gradient flow that the quantum Ornstein-Uhlenbeck semigroup converges in relative entropy to the fixed point at an exponential rate given by the exponent $\mu^2 - \lambda^2$,

$$D\left(\mathcal{P}^{(\mu,\lambda)}(t)(\rho)\|\omega^{(\mu,\lambda)}\right) \leq e^{-(\mu^2-\lambda^2)t} D\left(\rho\|\omega^{(\mu,\lambda)}\right) \qquad \text{for all } t \geq 0, \tag{7}$$

where $D(\rho\|\sigma) = \operatorname{tr}[\rho(\log\rho - \log\sigma)]$ is the quantum relative entropy.[16]

We show that a simple application of the linear version of the entropy power inequality (4) for the beam splitter is sufficient to prove this convergence rate. We also show a simple derivation of an analogous result for the case of a bipartite quantum system AM, where the system A undergoes a qOU evolution, using the linear conditional entropy power inequality for the beam splitter recently proven in Ref. 11. Specifically, we are going to show in Theorem 9 that

$$D\left((\mathcal{P}^{(\mu,\lambda)} \otimes \mathbb{1}_M)(\rho_{AM})\|\omega_A^{(\mu,\lambda)} \otimes \rho_M\right) \leq e^{-(\mu^2-\lambda^2)t} D\left(\rho_{AM}\|\omega_A^{(\mu,\lambda)} \otimes \rho_M\right), \tag{8}$$

which directly implies the statement (7). Finite-dimensional versions of the statement (8) for general semigroups have recently been studied by Bardet.[17] Our argument shows that entropy power inequalities are a useful tool to study the convergence rate of semigroups.

The proof of the unconditioned entropy power inequality (6) given in Ref. 13 exhibits certain regularity issues regarding the Fisher information: the Fisher information was defined as the Hessian of a relative entropy, without a proof of well definedness. Various proofs of the entropy power inequality for the beam splitter had similar issues.[7–9] They were settled in Ref. 11 by the adoption of a proof technique which starts with an integral version of the quantum Fisher information. We adopt a similar approach here. Since the conditional entropy power inequality reduces to the unconditioned inequality in the case where the system M is trivial, this also gives a more rigorous proof of the unconditioned entropy power inequality. As such, our work can be seen as both a completion of the work carried out in Ref. 13 and a generalization thereof.

We now sketch the basic structure of the proof of our main result. The main ingredients in proving entropy power inequalities[5,7,9,11,13] are similar in all proofs, which all use the evolution under the heat semigroup. These ingredients are the Fisher information, de Bruijn's identity, the Stam inequality, and a result on the asymptotic scaling of the entropy under the heat flow. First we define a "classical-quantum" integral conditional Fisher information, by which we mean a Fisher information of a classical system which is conditioned on a quantum system. We show in Theorem 1 that this quantity satisfies a de Bruijn identity, which links it to the change of the conditional entropy under the heat flow. We show the regularity of the integral conditional Fisher information in Theorem 2 and then prove the conditional Stam inequality in Theorem 3. In the next part, we show in Theorem 4 that the quantum conditional entropy of a classical system undergoing the classical heat flow evolution conditioned on a quantum system satisfies the same universal scaling which was shown for the quantum conditional entropy of a quantum system undergoing the quantum heat flow evolution conditioned on a quantum system. It is crucial for the proof of our conditional entropy power inequality that these two scalings are not only both universal but also the same. This scaling then implies that asymptotically, the inequality we want to prove becomes an equality. Then it is left to show that it is enough to consider the inequality in the asymptotic limit, i.e., the difference of the two sides of the inequality behaves under the heat flow in a way which only makes the inequality "worse."

The paper is structured as follows: In Sec. II, we present bosonic quantum systems and the relevant quantities required for our discussion. In Sec. III, the integral version of the quantum conditional Fisher information is adapted to the convolution (5). Sections IV and V are dedicated to the proof of various inequalities that are central to the proof of entropy power inequalities, such as the Stam inequality and an asymptotic scaling of the conditional entropy. Section VI then proves the conditional entropy power inequality for the convolution (5) as our main result. Optimality of the conditional entropy power inequality is shown in Sec. VII. This is followed by the derivation of various related information-theoretic inequalities involving the quantum conditional entropy in Sec. VIII. Before concluding, we apply the conditional entropy power inequality to bound the convergence rate of bipartite systems where one system undergoes a quantum Ornstein-Uhlenbeck semigroup evolution in Sec. IX.

II. PRELIMINARIES

Let us consider an n-mode bosonic system[16,18] with "position" and "momentum" operators (Q_k, P_k), $k = 1, \ldots, n$, for each mode which satisfy the canonical commutation relations $[Q_j, P_k] = i\delta_{j,k}\mathbb{1}$. If we denote the vector of position and momentum operators by $R = (Q_1, P_1, \ldots, Q_n, P_n)$, the canonical commutation relations become

$$[R_j, R_k] = i\Delta^{jk}\mathbb{1}, \qquad i, j = 1, \ldots, 2n,$$

where $\Delta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{\oplus n}$ is the symplectic form.

The Weyl displacement operators are defined by

$$D(\xi) := \exp\Big(i\xi \cdot (\Delta^{-1}R)\Big), \qquad \text{for } \xi \in \mathbb{R}^{2n}.$$

The displacement operators satisfy the commutation relations

$$D(\xi)D(\eta) = \exp\left(-\frac{i}{2}\xi \cdot (\Delta^{-1}\eta)\right)D(\xi + \eta), \qquad \text{for } \xi, \eta \in \mathbb{R}^{2n},$$

and the "displacement property" on the mode operators

$$D(\xi)^\dagger R_j D(\xi) = R_j + \xi_j \mathbb{1}.$$

Given an n-mode quantum state ρ, we define its first moments as

$$d_k(\rho) := \mathrm{tr}[R_k \rho], \qquad \text{for } k = 1, \ldots, 2n,$$

and its covariance matrix (for finite first moments) as

$$\Gamma_{kl}(\rho) := \frac{1}{2}\mathrm{tr}\left[\{R_k - d_k(\rho), R_l - d_l(\rho)\}\rho\right], \qquad k, l = 1, \ldots, 2n,$$

with the anticommutator $\{X, Y\} := XY - YX$.

The aforementioned concepts of displacements and first and second moments are the quantum analogs of the classical concepts. For a probability distribution function $f : \mathbb{R}^{2n} \to \mathbb{R}$, we define its displacement by a vector $\eta \in \mathbb{R}^{2n}$ as

$$f^{(\eta)}(\xi) = f(\xi - \eta).$$

Furthermore, we denote the energy of the function f by the sum of its second moments,

$$E(f) = \sum_{k=1}^{2n} \int_{\mathbb{R}^{2n}} \xi_k^2 f(\xi) \frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}.$$

The quantities $\mu_k = \int_{\mathbb{R}^{2n}} \xi_k f(\xi) \frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}$ are called the first moments of f, and

$$\gamma_{kl} = \int_{\mathbb{R}^{2n}} f(\xi)(\xi_k - \mu_k)(\xi_l - \mu_l) \frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}$$

is called the covariance matrix of f. We remark that we have rescaled the Lebesgue measure on $\mathbb{R}^{2n}$ in these definitions, which we have done purely for convenience.

Definition 1 (Quantum heat semigroup). *The quantum heat semigroup is the following time evolution for any quantum state ρ:*

$$\mathcal{N}(t)(\rho) := \int_{\mathbb{R}^{2n}} e^{-\frac{\|\xi\|^2}{2t}} \rho^{(\xi)} \frac{\mathrm{d}^{2n}\xi}{(2\pi t)^n} \qquad \text{for } t > 0,$$
$$\mathcal{N}(0) := \mathbb{1},$$

where $\rho^{(\xi)} = D(\xi)\rho D(\xi)^\dagger$ is a displacement of the state ρ by $\xi \in \mathbb{R}^{2n}$.

The quantum heat semigroup has a semigroup structure, that is, for any s, $t \geq 0$, we have

$$\mathcal{N}(s) \circ \mathcal{N}(t) = \mathcal{N}(s + t).$$

We note that if $f_{Z,t}(\xi) = \exp\left(-\frac{\|\xi\|^2}{2t}\right)/t^n$ is the probability distribution of a Gaussian random variable with covariance matrix $t\mathbb{1}_{2n}$, then we have

$$\mathcal{N}(t)(\rho) = f_{Z,t} \star \rho.$$

The quantum heat semigroup is the quantum analog of the classical heat semigroup, which we will repeat here. It can be written in an analogous way to the quantum heat semigroup:

Definition 2 (Classical heat semigroup). *The classical heat semigroup is the following time evolution defined on a function $f : \mathbb{R}^{2n} \to \mathbb{R}$:*

$$(\mathcal{N}_{\mathrm{cl}}(t)(f))(\eta) := \int_{\mathbb{R}^{2n}} e^{-\frac{\|\xi\|^2}{2t}} f^{(\xi)}(\eta) \frac{\mathrm{d}^{2n}\xi}{(2\pi t)^n},$$
$$\mathcal{N}_{\mathrm{cl}}(0) := \mathbb{1}.$$

We also have that for any s, $t \geq 0$,

$$\mathcal{N}_{\mathrm{cl}}(s) \circ \mathcal{N}_{\mathrm{cl}}(t) = \mathcal{N}_{\mathrm{cl}}(s + t).$$

We note again that we have

$$\mathcal{N}_{\text{cl}}(t)(f) = f_{Z,t} \star f,$$

where

$$(g \star f)(\eta) := \int_{\mathbb{R}^{2n}} g(\xi) f(\eta - \xi) \frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}$$

is the well-known classical convolution of the two functions g and f [with a factor of $(2\pi)^n$ in the Lebesgue measure on $\mathbb{R}^{2n}$ which we introduce purely for convenience].

The convolution (5) is compatible with displacements and with the heat semigroup evolution in a convenient way, which is stated in the following two lemmas:

Lemma 1 [Compatibility with displacements of the convolution (5)]. *(Ref. 13, Lemma 2) Let $f : \mathbb{R}^{2n} \to \mathbb{R}$ be a probability distribution and ρ be an n-mode quantum state. Then we have for any $\xi_1, \xi_2 \in \mathbb{R}^{2n}$,*

$$(f \star \rho)^{(\xi_1+\xi_2)} = f^{(\xi_1)} \star \rho^{(\xi_2)},$$

where $\rho^{(\xi)} = D(\xi)\rho D(\xi)^\dagger$.

Remark 1. Lemma 2 in Ref. 13 only states the compatibility for the case where ξ_1, ξ_2 are parallel. Nonetheless, the proof given there also works to prove the statement above.

Lemma 2 [Compatibility with the heat semigroup of the convolution (5)]. *(Ref. 13, Lemma 5) Assume the same prerequisites as in Lemma 1, and let t_1, $t_2 \geq 0$. Then we have*

$$\mathcal{N}(t_1 + t_2)(f \star \rho) = \mathcal{N}_{\text{cl}}(t_1)(f) \star \mathcal{N}(t_2)(\rho).$$

Definition 3 (Shannon differential entropy). *For a classical $\mathbb{R}^{2n}$-valued random variable X with a probability density function $f : \mathbb{R}^{2n} \to \mathbb{R}$, we define the Shannon differential entropy as*

$$S(X) = S(f) = -\int_{\mathbb{R}^{2n}} f(\xi) \log f(\xi) \frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}.$$

We continue with a short review of Gaussian quantum states. An n-mode quantum state ρ_G is called Gaussian if it has the following form:[16]

$$\rho_G = \frac{\exp\left[-\frac{1}{2}\sum_{k,l=1}^{2n}(R_k - d_k)h_{kl}(R_l - d_l)\right]}{\operatorname{tr}\exp\left[-\frac{1}{2}\sum_{k,l=1}^{2n}(R_k - d_k)h_{kl}(R_l - d_l)\right]},$$

where h is a positive definite real $2n \times 2n$ matrix and $d \in \mathbb{R}^{2n}$ is the vector of first moments of the state. The entropy of such a Gaussian state is given by

$$S(\rho_G) = \sum_{k=1}^{n} g\left(\nu_k - \frac{1}{2}\right),$$

where $g(N) := (N + 1)\log(N + 1) - N\log N$ and $\nu_1, \ldots, \nu_n$ are the symplectic eigenvalues of the covariance matrix $\Gamma = \frac{\Delta}{2}\left(\tan\frac{h\Delta}{2}\right)^{-1}$, i.e., the absolute values of the eigenvalues of $\Delta^{-1}\Gamma$.

A Gaussian state is called thermal if its first moments are zero, and the matrix h is proportional to the identity. Such thermal states have the special form

$$\omega_\beta = \frac{e^{-\beta H}}{\operatorname{tr} e^{-\beta H}}, \qquad h = \beta\mathbb{1}_{2n}, \qquad \beta > 0$$

for the Hamiltonian of n harmonic oscillators $H = \frac{1}{2}\sum_{k=1}^{2n} R_k^2 - \frac{n}{2}\mathbb{1}$. Gaussian states fulfill a special extremality property. Among all states ρ with a given average energy $\operatorname{tr}[H\rho]$, thermal states maximize the von Neumann entropy. Furthermore, among all states with a fixed covariance matrix, the Gaussian state is the one with maximal entropy.[19,20]

In our proofs, we are going to require the notion of quantum conditional Fisher information of quantum systems which was introduced in Ref. 11. We repeat the main properties of this quantity here. For a thorough definition and proofs, we refer to Ref. 11. Before giving this definition, we clarify the notion of "classical-quantum" states on a system RM if the classical system R is continuous. A state ρ_{RM} on RM is a probability measure on R which takes values in the trace class operators, i.e., a measurable collection of trace class operators on M $\{\rho_{MR}(\xi)\}_{\xi\in\mathbb{R}^{2n}}$ with the normalization condition

$$\int_{\mathbb{R}^{2n}} \mathrm{tr}_M[\rho_{MR}(\xi)]\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n} = 1.$$

This state "stores" a classical probability distribution ρ_R in the classical system R if its marginal on R has ρ_R as probability distribution. The marginals of ρ_{MR} are

$$\rho_M = \int_{\mathbb{R}^{2n}} \rho_{MR}(\xi)\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}, \qquad \rho_R(\xi) = \mathrm{tr}_M[\rho_{MR}(\xi)],$$

and the conditional states on M given the value of ξ are

$$\rho_{M|R=\xi} = \frac{\rho_{MR}(\xi)}{\rho_R(\xi)}.$$

We do not consider the case where the probability measure ρ_R is not absolutely continuous with respect to the Lebesgue measure since in this case, its Shannon differential entropy is not defined. For a more detailed discussion, we refer to Ref. 21, Sec. III.A.3, and the references therein (Refs. 22 and 23, Chaps. 4.6–4.7).

We can also define displacements of such a classical-quantum state: We write $\rho_{RM}^{(x,y)}$ to denote a state where the classical system R has been displaced by $x \in \mathbb{R}^{2n}$ and the quantum system M has been displaced by $y \in \mathbb{R}^{2n}$.

Definition 4 (Quantum integral conditional Fisher information). *(Ref. 11, Definition 6) Let A be an n-mode bosonic quantum system, and M be a generic quantum system. Let ρ_{AM} be a quantum state on AM. For any $t \geq 0$, the integral Fisher information of A conditioned on M is given by*

$$\Delta_{A|M}(\rho_{AM})(t) := I(A:Z|M)_{\sigma_{AMZ(t)}} \geq 0, \qquad t > 0,$$
$$\Delta_{A|M}(\rho_{AM})(0) := 0,$$

where Z is a classical Gaussian random variable with values in $\mathbb{R}^{2n}$ and probability density function

$$f_{Z,t}(z) = \frac{e^{-\frac{|z|^2}{2t}}}{t^n}, \qquad z \in \mathbb{R}^{2n},$$

and $\sigma_{AMZ}(t)$ is the quantum state on AMZ such that its marginal on Z is $f_{Z,t}$, and for any $z \in \mathbb{R}^{2n}$,

$$\sigma_{AM|Z=z}(t) = D_A(z)\rho_{AM}D_A(z)^\dagger.$$

Definition 5 (Quantum conditional Fisher information). *(Ref. 11, Definition 7, Proposition 1) Let ρ_{AM} be a quantum state on AM such that the marginal ρ_A has finite energy and the marginal ρ_M has finite entropy. Then we define the quantum conditional Fisher information of A conditioned on M as*

$$J(A|M)_{\rho_{AM}} := \lim_{t\to 0}\frac{\Delta_{A|M}(\rho_{AM})(t)}{t} = \frac{\mathrm{d}}{\mathrm{d}t}S(A|M)_{(\mathcal{N}_A(t)\otimes\mathbb{1}_M)(\rho_{AM})}\Big|_{t=0}.$$

As shown in Ref. 11, this limit always exists.

Finally, we are going to require a notion of conditional entropy of a classical system which is conditioned on a quantum system. If the system on which we condition is classical, the conditional entropy is simply

$$S(A|M) = \int_M S(A|M=m)\mathrm{d}p_M(m),$$

where p_M is the probability distribution of M. This definition is independent of whether the system A is classical or quantum. We now define the conditional entropy of a classical system which is conditioned on a quantum system in a way such that the chain rule for entropies is preserved.

Definition 6 (Quantum conditional entropy of classical-quantum systems). *Let R be a classical system and M be a quantum system. We define the conditional entropy of R given M as*

$$S(R|M) = S(M|R) + S(R) - S(M),$$

whenever the three quantities appearing on the right-hand side are finite.

The case where $S(M|R)$, $S(R)$, and $S(M)$ are not finite will not be part of our consideration.

III. QUANTUM INTEGRAL CONDITIONAL FISHER INFORMATION

In this section, we consider a generic quantum system M and a classical system R. We are going to define the quantum integral conditional Fisher information of R conditioned on M and prove a de Bruijn identity as well as a number of useful properties.

Definition 7 (Quantum integral conditional Fisher information). *For a quantum state ρ_{RM} on RM whose marginal on R is $\rho_R : \mathbb{R}^{2n} \to \mathbb{R}$ and $t \geq 0$, define the integral Fisher information of R conditioned on M as*

$$\begin{aligned} \Delta_{R|M}(\rho_{RM})(t) &:= I(R:Z|M)_{\sigma_{RZM}(t)}, \\ \Delta_{R|M}(\rho_{RM})(0) &:= 0, \end{aligned}$$

where Z is a classical Gaussian random variable with the probability density function equal to

$$f_{Z,t}(\xi) = \frac{e^{-\frac{|\xi|^2}{2t}}}{t^n}, \qquad \xi \in \mathbb{R}^{2n},$$

and $\sigma_{RZM}(t)$ is the quantum state on RZM such that its marginal on Z is equal to $f_{Z,t}$, and for any $z \in \mathbb{R}^{2n}$, we have

$$\sigma_{RM|Z=z}(t) = \rho_{RM}^{(z,0)}.$$

The marginal of $\sigma_{RZM}(t)$ on RM is equal to

$$\sigma_{RM}(t) = (\mathcal{N}_{\mathrm{cl}}(t) \otimes \mathbb{1}_M)(\rho_{RM}).$$

The marginal on R has probability density function $\mathcal{N}_{\mathrm{cl}}(t)(\rho_R)$.

Theorem 1 (Integral conditional de Bruijn identity).

$$\Delta_{R|M}(\rho_{RM})(t) = S(R|M)_{(\mathcal{N}_{\mathrm{cl}}(t)\otimes\mathbb{1}_M)(\rho_{RM})} - S(R|M)_{\rho_{RM}}.$$

Proof. We use the definition of the conditional mutual information as well as the definition of the conditional quantum entropy when the system on which we condition is classical. We calculate

$$\begin{aligned} I(R:Z|M)_{\sigma_{RMZ}} &= S(R|M)_{\sigma_{RMZ}} - S(R|MZ)_{\sigma_{RMZ}} \\ &= S(R|M)_{\sigma_{RM}} - \int_{\mathbb{R}^{2n}} S(R|M)_{\sigma_{RM|Z=z}} f_{Z,t}(z)\, \frac{\mathrm{d}^{2n}z}{(2\pi)^n} \\ &= S(R|M)_{\sigma_{RM}} - \int_{\mathbb{R}^{2n}} S(R|M)_{\rho_{RM}} f_{Z,t}(z)\, \frac{\mathrm{d}^{2n}z}{(2\pi)^n} \\ &= S(R|M)_{\sigma_{RM}} - S(R|M)_{\rho_{RM}}. \end{aligned}$$

The second to last step follows because the entropy is invariant under displacements of the classical system. □

We now show that the integral conditional Fisher information defined as above, as a function of t, is continuous, increasing, and concave. The proof strategy is similar to the proof of regularity for the quantum integral conditional Fisher information given in Ref. 11.

Lemma 3 (Continuity of the integral conditional Fisher information). *Let ρ_{RM} be a state such that the function $\mathbb{R}^{2n} \ni \xi \mapsto \rho_{M|R=\xi}$ is continuous with respect to the trace norm and the marginal ρ_R has finite average energy. Then, the function $t \mapsto \Delta_{R|M}(\rho_{RM})(t)$ is continuous for any $t \geq 0$.*

Proof. From the de Bruijn identity, Theorem 1, it is sufficient to prove that

$$\lim_{t\to 0} S(R|M)(\rho_{RM}(t)) = S(R|M)(\rho_{RM}),$$

where we have defined for any $t \geq 0$,

$$\rho_{RM}(t) = (\mathcal{N}_{\mathrm{cl}}(t) \otimes \mathbb{1}_M)(\rho_{RM}).$$

From the data processing inequality, for any $t \geq 0$,

$$S(R|M)(\rho_{RM}(t)) \geq S(R|M)(\rho_{RM}).$$

It is then sufficient to prove that

$$\limsup_{t\to 0} S(R|M)(\rho_{RM}(t)) \leq S(R|M)(\rho_{RM}).$$

We have from the chain rule

$$S(R|M)((\mathcal{N}_{\mathrm{cl}}(t) \otimes \mathbb{1}_M)(\rho_{RM})) = S(M|R)(\rho_{RM}(t)) + S(\rho_R(t)) - S(\rho_M).$$

From Ref. 24, Remark 9.3.8, and Refs. 25–27, the Shannon differential entropy is upper semicontinuous on the set of probability measures on $\mathbb{R}^{2n}$ absolutely continuous with respect to the Lebesgue measure and with finite average energy, and

$$\limsup_{t\to 0} S(\rho_R(t)) \leq S(\rho_R).$$

On the other hand, we have

$$S(\rho_M) - S(M|R)(\rho_{RM}(t)) = \int_{\mathbb{R}^{2n}} D(\rho_{M|R=\xi}(t)\|\rho_M)\,\rho_R(t)(\xi)\,\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}. \tag{9}$$

Since the function $t \mapsto \rho_{M|R=\xi}(t)$ is continuous with respect to the trace norm, we have for any $\xi \in \mathbb{R}^{2n}$,

$$\lim_{t\to 0} \|\rho_{M|R=\xi}(t) - \rho_{M|R=\xi}\|_1 = 0.$$

Because the relative entropy is positive, we get from Fatou's lemma

$$\begin{aligned}&\int_{\mathbb{R}^{2n}} \liminf_{t\to 0} D(\rho_{M|R=\xi}(t)\|\rho_M)\,\rho_R(t)(\xi)\,\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}\\ &\leq \liminf_{t\to 0} \int_{\mathbb{R}^{2n}} D(\rho_{M|R=\xi}(t)\|\rho_M)\,\rho_R(t)(\xi)\,\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}.\end{aligned} \tag{10}$$

Since the relative entropy is lower semicontinuous, we have for any $\xi \in \mathbb{R}^{2n}$,

$$D(\rho_{M|R=\xi}\|\rho_M) \leq \liminf_{t\to 0} D(\rho_{M|R=\xi}(t)\|\rho_M). \tag{11}$$

Combining (10), (11), and (9), we get

$$\limsup_{t\to 0} S(M|R)(\rho_{RM}(t)) \leq S(M|R)(\rho_{RM}).$$

□

Lemma 4. For any s, t ≥ 0,

$$\Delta_{R|M}((\mathcal{N}_{\mathrm{cl}}(s) \otimes \mathbb{1}_M)(\rho_{RM}))(t) = I(R:Z|M)_{(\mathcal{N}_{\mathrm{cl}}(s)\otimes\mathbb{1}_{MZ})(\sigma_{RMZ}(t))}.$$

Proof. Follows from the semigroup structure of $\mathcal{N}_{\mathrm{cl}}$. □

Lemma 5. For any s, t ≥ 0,

$$\Delta_{R|M}((\mathcal{N}_{\mathrm{cl}}(s) \otimes \mathbb{1}_M)(\rho_{RM})(t) \leq \Delta_{R|M}(\rho_{RM})(t).$$

Proof. Follows from the data processing inequality for the quantum mutual information. □

Lemma 6. For any s, t ≥ 0,

$$\begin{aligned}\Delta_{R|M}(\rho_{RM})(s+t) &= \Delta_{R|M}(\rho_{RM})(s) + \Delta_{R|M}((\mathcal{N}_{\rm cl}(s)\otimes\mathbb{1}_M)(\rho_{RM})(t)\\ &\geq \Delta_{R|M}(\rho_{RM})(s).\end{aligned}$$

Proof. Follows from Theorem 1. □

Theorem 2 (Regularity of the integral conditional Fisher information). *For any quantum state ρ_{RM} on RM such that the conditions of Lemma 3 are fulfilled, the integral conditional Fisher information $\Delta_{R|M}(\rho_{RM})(t)$ is a continuous, increasing, and concave function of t.*

Proof. Continuity was shown in Lemma 3 and the fact that the conditional Fisher information is increasing follows from Lemma 6.

For concavity, by continuity, it is enough to prove that for $0 \leq s \leq t$, we have

$$\Delta_{R|M}(\rho_{RM})\left(\frac{s+t}{2}\right) \geq \frac{\Delta_{R|M}(\rho_{RM})(s) + \Delta_{R|M}(\rho_{RM})(t)}{2}.$$

This can be written as

$$\begin{aligned}&\Delta_{R|M}(\rho_{RM})\left(\frac{s+t}{2}\right) - \Delta_{R|M}(\rho_{RM})(s)\\ &\geq \Delta_{R|M}(\rho_{RM})(t) - \Delta_{R|M}(\rho_{RM})\left(\frac{s+t}{2}\right).\end{aligned}$$

By Lemma 6, this can be restated as

$$\Delta_{R|M}(\rho_{RM}(s))\left(\frac{t-s}{2}\right) \geq \Delta_{R|M}\left(\left(\mathcal{N}_{\rm cl}\left(\frac{t-s}{2}\right)\otimes\mathbb{1}_M\right)(\rho_{RM}(s))\right)\left(\frac{t-s}{2}\right),$$

for $\rho_{RM}(s) := (\mathcal{N}_{\rm cl}(s)\otimes\mathbb{1}_M)(\rho_{RM})$. But this holds because of Lemma 5. □

IV. QUANTUM CONDITIONAL FISHER INFORMATION

Definition 8. For a quantum state ρ_{RM} on RM such that the conditions of Lemma 3 are fulfilled, we define the Fisher information of R conditioned on M as

$$J(R|M)_{\rho_{RM}} := \lim_{t\to 0}\frac{\Delta_{R|M}(\rho_{RM})(t)}{t}.$$

This limit always exists because the function $t \mapsto \Delta_{R|M}(\rho_{RM})(t)$ is continuous and concave by Theorem 2.

Proposition 1 (Quantum conditional de Bruijn). *Assume the hypotheses of Theorem 2. Then we have*

$$J(R|M)_{\rho_{RM}} = \frac{\mathrm{d}}{\mathrm{d}t}S(R|M)_{(\mathcal{N}_{cl}(t)\otimes\mathbb{1}_M)(\rho_{RM})}\Big|_{t=0}.$$

Proof. Follows from the integral conditional de Bruijn identity given in Theorem 1. □

A. Stam inequality

Theorem 3. *Let A be an n-mode quantum system, R be a classical system, and M be a generic quantum system. Let ρ_{ARM} be a quantum state on ARM such that its marginal on R has a probability density function $\rho_R : \mathbb{R}^{2n} \to \mathbb{R}$. Let ρ_{ARM} further fulfill*

$$\mathrm{tr}[H\rho_A] < \infty, \qquad E(\rho_R) < \infty, \qquad S(\rho_M) < \infty.$$

Let us suppose that A and R are conditionally independent given M,

$$I(A:R|M)_{\rho_{ARM}} = 0.$$

Then the linear conditional Stam inequality holds,

$$J(C|M)_{\rho_{CM}} \leq \lambda^2 J(A|M)_{\rho_{AM}} + (1-\lambda)^2 J(R|M)_{\rho_{RM}}, \qquad \forall \lambda \in [0,1],$$

where

$$\rho_{CM} \coloneqq (\mathcal{E} \otimes \mathbb{1}_M)(\rho_{ARM}) = \int_{\mathbb{R}^{2n}} D(\xi)\, \rho_{AM|R=\xi}\, D(\xi)^\dagger\, \rho_R(\xi)\, \frac{d^{2n}\xi}{(2\pi)^n}.$$

Choosing $\lambda = \frac{J(R|M)}{J(R|M)+J(A|M)}$*, we obtain the conditional Stam inequality*

$$\frac{1}{J(C|M)_{\rho_{CM}}} \geq \frac{1}{J(A|M)_{\rho_{AM}}} + \frac{1}{J(R|M)_{\rho_{RM}}}. \tag{12}$$

Proof. We prove the following:

$$\Delta_{C|M}(\rho_{CM})(t) \leq \Delta_{A|M}(\rho_{AM})(\lambda^2 t) + \Delta_{R|M}(\rho_{RM})((1-\lambda)^2 t).$$

Because Δ is increasing and concave, the Stam inequality follows taking the derivative at $t = 0$.

By definition, we have for any $t \geq 0$ that

$$\Delta_{C|M}(\rho_{CM})(t) = I(C:Z|M)_{\sigma_{CMZ}(t)},$$

for an $\mathbb{R}^{2n}$-valued Gaussian random variable Z with probability density function

$$f_{Z,t}(z) = \frac{e^{-\frac{\|z\|^2}{2t}}}{t^n}, \qquad z \in \mathbb{R}^{2n},$$

and $\sigma_{CMZ}(t)$ has $f_{Z,t}$ as marginal on Z, and for any $z \in \mathbb{R}^{2n}$, it fulfills

$$\sigma_{CM|Z=z}(t) = D_C(z)\rho_{CM}D_C(z)^\dagger.$$

We now define the state $\sigma_{ARMZ}(t)$ as the state with marginal on Z equal to $f_{Z,t}$, and for any $z \in \mathbb{R}^{2n}$,

$$\sigma_{ARM|Z=z} = \rho_{ARM}^{(\lambda z,(1-\lambda)z)},$$

i.e., the system A is displaced by λz and the system R is displaced by $(1-\lambda)z$. By compatibility of the convolution (5) with displacements, we have

$$\sigma_{CMZ}(t) = (\mathcal{E} \otimes \mathbb{1}_{MZ})(\sigma_{ARMZ}(t)).$$

We notice that

$$\begin{aligned} I(A:R|MZ)_{\sigma_{ARMZ}} &= \int_{\mathbb{R}^{2n}} I(A:R|M)_{\sigma_{ARM|Z=z}} f_{Z,t}(z)\, \frac{d^{2n}z}{(2\pi)^n} \\ &= \int_{\mathbb{R}^{2n}} I(A:R|M)_{\rho_{ARM}} f_{Z,t}(z)\, \frac{d^{2n}z}{(2\pi)^n} = 0. \end{aligned}$$

Now we obtain by data processing

$$\begin{aligned} I(C:Z|M)(t) &\leq I(AR:Z|M)(t) \\ &= I(A:Z|M)(t) + I(R:Z|M)(t) + I(A:R|MZ)(t) - I(A:R|M)(t) \\ &\leq I(A:Z|M)(t) + I(R:Z|M)(t). \end{aligned}$$

The last inequality follows because $I(A{:}\, R|M)(t) \geq 0$. In analogy to Ref. 11, Eqs. (79)–(81), we can show that

$$\begin{aligned} I(A:Z|M)_{\sigma_{AMZ}(t)} &= \Delta_{A|M}(\rho_{AM})\big(\lambda^2 t\big), \\ I(R:Z|M)_{\sigma_{RMZ}(t)} &= \Delta_{R|M}(\rho_{RM})\big((1-\lambda)^2 t\big). \end{aligned}$$

This follows from the definition of σ_{ARMZ} and the integral conditional de Bruijn identities. The claim follows. □

V. UNIVERSAL SCALING

Theorem 4. *Let R be a classical system and M be a quantum system. Let ρ_{RM} be a quantum state on RM such that its marginals have finite entropies. Then we have*

$$\lim_{t\to\infty}\left(S(R|M)_{(\mathcal{N}_{\mathrm{cl}}(t)\otimes\mathbb{1}_M)(\rho_{RM})} - n\log t - n\right) = 0.$$

Proof. **Upper bound.** We have

$$S(R|M)_{(\mathcal{N}_{\mathrm{cl}}(t)\otimes\mathbb{1}_M)(\rho_{RM})} \le S(R)_{\mathcal{N}_{\mathrm{cl}}(t)(\rho_R)}.$$

We know from the analysis of the classical heat flow[5] that the right-hand side scales as $n\log t + n$.

Lower bound. By concavity, we can restrict to pure ρ_{RM}. The pure states of the classical-quantum system RM are the tensor product of a Dirac delta on R with a pure state on M; hence, R and M are independent and

$$S(R|M)_{(\mathcal{N}_{\mathrm{cl}}(t)\otimes\mathbb{1}_M)(\rho_{RM})} = S(R)_{\mathcal{N}_{\mathrm{cl}}(t)(\rho_R)}.$$

Finally, the scaling of the classical entropy $S(R)_{\mathcal{N}_{\mathrm{cl}}(t)(\rho_R)}$ is known to be equal to $n\log t + n$ from Ref. 5, which concludes the proof. □

VI. ENTROPY POWER INEQUALITY

Theorem 5 [Conditional entropy power inequality for the convolution (5)]. *Let A be an n-mode quantum system, R be a classical system, and M be a generic quantum system. Let ρ_{ARM} be a quantum state on ARM such that its marginal on R has a probability density function $\rho_R : \mathbb{R}^{2n} \to \mathbb{R}$. Let ρ_{ARM} further fulfill*

$$\mathrm{tr}[H\rho_A] < \infty, \qquad E(\rho_R) < \infty, \qquad S(\rho_M) < \infty.$$

Let us suppose that A and R are conditionally independent given M,

$$I(A:R|M)_{\rho_{ARM}} = 0,$$

and let

$$\rho_{CM} := (\mathcal{E}\otimes\mathbb{1}_M)(\rho_{ARM}) = \int_{\mathbb{R}^{2n}} D(\xi)\,\rho_{AM|R=\xi}\,D(\xi)^\dagger\,\rho_R(\xi)\,\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}.$$

Then, for any $0 \le \lambda \le 1$, the linear conditional entropy power inequality holds,

$$\frac{S(C|M)}{n} \ge \lambda\frac{S(A|M)}{n} + (1-\lambda)\frac{S(R|M)}{n} - \lambda\log\lambda - (1-\lambda)\log(1-\lambda).$$

Optimizing over λ and choosing $\lambda = \frac{e^{S(A|M)/n}}{e^{S(A|M)/n}+e^{S(R|M)/n}}$, we obtain the conditional entropy power inequality for the convolution (5),

$$\boxed{\exp\frac{S(C|M)}{n} \ge \exp\frac{S(A|M)}{n} + \exp\frac{S(R|M)}{n}.} \tag{13}$$

In particular, if the classical system R is uncorrelated with the system M, we have the inequality

$$\exp\frac{S(C|M)}{n} \ge \exp\frac{S(A|M)}{n} + \exp\frac{S(\rho_R)}{n}.$$

Remark 2. An important case for applications is the case when R has the Gaussian probability density function $f_{Z,t} = \exp\left(-\frac{\|\xi\|^2}{2t}\right)/t^n$. In this special case, the inequality reads

$$\exp\frac{S(C|M)}{n} \ge \exp\frac{S(A|M)}{n} + et.$$

Proof. We define the evolution

$$\rho_{ARM}(t) = (\mathcal{N}(\lambda t)\otimes\mathcal{N}_{\mathrm{cl}}((1-\lambda)t)\otimes\mathbb{1}_M)(\rho_{ARM}).$$

Then, by compatibility with the heat semigroup, this amounts to an evolution of the C system given by

$$\rho_{CM}(t) = (\mathcal{N}(t) \otimes \mathbb{1}_M)(\rho_{CM}).$$

This evolution preserves the condition $I(A : R|M) = 0$ because of the data-processing inequality. We also define

$$\phi(t) = S(C|M)_{\rho_{CM}(t)} - \lambda S(A|M)_{\rho_{AM}(t)} - (1 - \lambda)S(R|M)_{\rho_{RM}(t)}.$$

Then we have, because of the de Bruijn identity and compatibility with the heat semigroup as well as the Stam inequality,

$$\phi'(t) = J(C|M)_{\rho_{CM}(t)} - \lambda^2 J(A|M)_{\rho_{AM}(t)} - (1 - \lambda)^2 J(R|M)_{\rho_{CM}(t)} \leq 0.$$

Since ϕ is a linear combination of continuous concave functions, we have for $t \geq 0$,

$$\phi(t) - \phi(0) = \int_0^t \phi'(s)\mathrm{d}s \leq 0.$$

Using the universal scaling, we obtain

$$\begin{aligned}\phi(0) &\geq \lim_{t\to\infty} \phi(t)\\ &= \lim_{t\to\infty}\left(S(C|M)_{\rho_{CM}(t)} - \lambda S(A|M)_{\rho_{AM}(t)} - (1 - \lambda)S(R|M)_{\rho_{RM}(t)}\right)\\ &= n(-\lambda \log \lambda - (1 - \lambda) \log(1 - \lambda)).\end{aligned}$$

The theorem follows. □

VII. OPTIMALITY OF THE QUANTUM CONDITIONAL ENTROPY POWER INEQUALITY

This section is dedicated to the study of the optimality of the quantum conditional entropy power inequality stated in Theorem 5. We show the following theorem:

Theorem 6 (Optimality of the conditional entropy power inequality). *For any $a, b \in \mathbb{R}$, there exists a sequence of states $\left\{\rho_{AM}^{(k)}\right\}_{k\in\mathbb{N}}$ and a probability density function $f : \mathbb{R}^2 \to \mathbb{R}$ such that the classical system R is uncorrelated with M and*

$$\lim_{k\to\infty} S(A|M)_{\rho_{AM}^{(k)}} = a, \qquad S(R|M)_f = b,$$

and

$$\lim_{k\to\infty} \exp S(C|M)_{\rho_{CM}^{(k)}} = \exp a + \exp b,$$

where $\rho_{CM}^{(k)} = (\mathcal{E}_f \otimes \mathbb{1}_M)(\rho_{AM})$ with $\mathcal{E}_f(\rho_A) = f \star \rho_A$.

Proof. Let $\sigma_{AM}^{(k)}$ be the Gaussian state with the covariance matrix

$$\Gamma_{AM}^{(k)} = \begin{pmatrix} k^2 & 0 & \sqrt{k^4 - \frac{1}{4}} & 0\\ 0 & k^2 & 0 & -\sqrt{k^4 - \frac{1}{4}}\\ \sqrt{k^4 - \frac{1}{4}} & 0 & k^2 & 0\\ 0 & -\sqrt{k^4 - \frac{1}{4}} & 0 & k^2 \end{pmatrix}.$$

Applying the heat semigroup on the quantum system A, we obtain the state $(\mathcal{N}(t) \otimes \mathbb{1}_M)(\sigma_{AM}^{(k)})$ which has the covariance matrix

$$\Gamma_{AM}^{(k)}(t) = \begin{pmatrix} k^2 + t & 0 & \sqrt{k^4 - \frac{1}{4}} & 0\\ 0 & k^2 + t & 0 & -\sqrt{k^4 - \frac{1}{4}}\\ \sqrt{k^4 - \frac{1}{4}} & 0 & k^2 & 0\\ 0 & -\sqrt{k^4 - \frac{1}{4}} & 0 & k^2 \end{pmatrix}.$$

The symplectic eigenvalues of this covariance matrix are

$$\nu_{\pm}^{(k)}(t) = \frac{1}{2}\sqrt{4k^2t \pm 2t\sqrt{4k^2t + t^2 + 1} + 2t^2 + 1} = k\sqrt{t} + \mathcal{O}(1) \qquad (k \to \infty).$$

Hence we have

$$\begin{aligned} S(AM)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\sigma_{AM}^{(k)})} &= g\left(\nu_+ - \frac{1}{2}\right) + g\left(\nu_- - \frac{1}{2}\right) \\ &= \log k^2 + \log t + 2 + \mathcal{O}\left(\frac{1}{k^2}\right), \\ S(M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\sigma_{AM}^{(k)})} &= g\left(k^2 - \frac{1}{2}\right) = \log k^2 + 1 + \mathcal{O}\left(\frac{1}{k^4}\right). \end{aligned}$$

It follows that

$$\lim_{k\to\infty} S(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\sigma_{AM}^{(k)})} = 1 + \log t.$$

We now choose $\rho_{AM}^{(k)} = (\mathcal{N}(e^{a-1}) \otimes \mathbb{1}_M)(\sigma_{AM}^{(k)})$, which fulfills

$$\lim_{k\to\infty} S(A|M)_{(\rho_{AM}^{(k)})} = a.$$

We further choose the classical system R to be uncorrelated with M and have probability density function

$$f = f_{Z,e^{b-1}} = \frac{e^{-\frac{\|\xi\|^2}{2e^{b-1}}}}{e^{b-1}}$$

of a Gaussian with covariance matrix $e^{b-1}\mathbb{1}_2$. Then we have for the entropy

$$S(R|M)_f = \log\left(ee^{b-1}\right) = b.$$

The state $\rho_{CM}^{(k)}$ has the covariance matrix

$$\Gamma_{CM}^{(k)} = \begin{pmatrix} k^2 + e^{a-1} + e^{b-1} & 0 & \sqrt{k^4 - \frac{1}{4}} & 0 \\ 0 & k^2 + e^{a-1} + e^{b-1} & 0 & -\sqrt{k^4 - \frac{1}{4}} \\ \sqrt{k^4 - \frac{1}{4}} & 0 & k^2 & 0 \\ 0 & -\sqrt{k^4 - \frac{1}{4}} & 0 & k^2 \end{pmatrix}.$$

Analogous to the calculation above, we obtain now

$$\lim_{k\to\infty} S(C|M)_{\rho_{CM}^{(k)}} = 1 + \log\left(e^{a-1} + e^{b-1}\right) = \log\left(e^a + e^b\right)$$

and finally

$$\lim_{k\to\infty} \exp S(C|M)_{\rho_{CM}^{(k)}} = e^a + e^b.$$

□

VIII. APPLICATIONS

The quantum conditional entropy power inequality (13) has various applications in the derivation of information-theoretic inequalities. We are going to show a variety of results regarding quantum conditional entropies. Many of these results have direct analogs in the case of unconditioned quantum entropies as well as in classical information theory.

A. Isoperimetric inequality for conditional entropies

Lemma 7 (Quantum conditional Fisher information isoperimetric inequality).

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{1}{n}J(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}\right]^{-1}\Bigg|_{t=0} \geq 1. \qquad (14)$$

Proof. We note that $(\mathcal{N}(t) \otimes \mathbb{1}_M)(\rho_{AM}) = (\mathcal{E}_{f_{Z,t}} \otimes \mathbb{1}_M)(\rho_{AM})$, where again $\mathcal{E}_f(\rho_A) = f \star \rho_A$ and $f_{Z,t}(\xi) = \exp\left(-\frac{\|\xi\|^2}{2t}\right)/t^n$. Applying the conditional Stam inequality (12), we obtain

$$\left(J(A|M)^{-1}_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})} - J(A|M)^{-1}_{\rho_A}\right) \geq J(R|M)^{-1}_{f_{Z,t}} = \frac{t}{n}.$$

This implies that

$$\frac{1}{t}\left(J(A|M)^{-1}_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})} - J(A|M)^{-1}_{\rho_{AM}}\right) \geq \frac{1}{n}.$$

Taking the limit $t \to 0$ implies the result. □

Theorem 7 (Isoperimetric inequality for quantum conditional entropies).

$$\frac{1}{n} J(A|M)_{\rho_{AM}} \exp\frac{S(A|M)_{\rho_{AM}}}{n} \geq e.$$

Proof. We apply the conditional de Bruijn identity[11] [Eq. (63)] and see that

$$\frac{\mathrm{d}}{\mathrm{d}t} \exp\frac{S(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}}{n}\bigg|_{t=0} = \frac{1}{n} J(A|M)_{\rho_{AM}} \exp\frac{S(A|M)_{\rho_{AM}}}{n}.$$

Recalling once again that $(\mathcal{N}(t) \otimes \mathbb{1}_M)(\rho_{AM}) = (\mathcal{E}_{f_{Z,t}} \otimes \mathbb{1}_M)(\rho_{AM})$ and inserting this into the conditional entropy power inequality (13) yields

$$\exp\frac{S(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}}{n} - \exp\frac{S(A|M)}{n} \geq \exp\frac{S(R|M)}{n} = et.$$

Dividing this equation by t and taking the limit $t \to 0$ concludes the proof of the theorem. □

B. Concavity of the quantum conditional entropy power along the heat flow

Theorem 8 (Concavity of the quantum conditional entropy power along the heat flow).

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} \exp\frac{S(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}}{n}\bigg|_{t=0} \leq 0.$$

Proof. We write here $P(t) = \exp\frac{S(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}}{n}$ and apply the de Bruijn identity[11] [Eq. (63)] twice to obtain

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} P(t)\bigg|_{t=0} = P(0)\left(\left[\frac{1}{n} J(A|M)\right]^2 + \frac{1}{n}\frac{\mathrm{d}}{\mathrm{d}t} J(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}\bigg|_{t=0}\right).$$

The quantum conditional Fisher information isoperimetric inequality stated in (14) can be restated as

$$\frac{1}{n^2} J(A|M)^2 + \frac{1}{n}\frac{\mathrm{d}}{\mathrm{d}t} J(A|M)_{(\mathcal{N}(t)\otimes\mathbb{1}_M)(\rho_{AM})}|_{t=0} \leq 0.$$

Since $P(0) \geq 0$, the concavity of the quantum conditional entropy power follows. □

C. Converse bound on the entanglement-assisted classical capacity for a non-Gaussian classical noise channel

The fact that conditional entropy power inequalities imply upper bounds on the entanglement-assisted classical capacity has been known since the first quantum conditional entropy power inequality has been proposed.[10,11] In this section, we use the conditional entropy power inequality (13) to prove such an upper bound for a classical noise channel which is not necessarily Gaussian.

We consider the classical noise channel with a given (possibly non-Gaussian) noise probability density function $f : \mathbb{R}^{2n} \to \mathbb{R}$. This channel is given by $\mathcal{E}_f : A \to C$,

$$\mathcal{E}_f(\rho_A) = f \star \rho_A .$$

The entanglement-assisted classical capacity[16,28,29] is then

$$C_{\mathrm{ea}}(\mathcal{E}_f) = \sup\Big\{ I(C:M)_{(\mathcal{E}_f \otimes \mathbb{1}_M)(\rho_{AM})} : \rho_{AM} \text{ pure}, \mathrm{tr}_A[H_A \rho_A] \le nE \Big\}.$$

The energy constraint $\mathrm{tr}_A[H_A\rho_A] \le nE$ amounts to the assumption that the sender can only use states of a finite average energy E per mode. This assumption is required to make the entanglement-assisted capacity finite. Indeed, the assumption that a sender can use an unlimited amount of energy is unphysical. Let

$$E_0 := \frac{E(f)}{2n}, \qquad S_0 := \frac{S(f)}{n}$$

be the average energy and entropy per mode of f. We can now bound the maximum output entropy as

$$\begin{aligned}
\sup_{\mathrm{tr}[H_A\rho_A]\le nE} S\Big(\mathcal{E}_f(\rho_A)\Big) &= \sup_{\mathrm{tr}[H_A\rho_A]\le nE} S\Big(\mathcal{E}_f(\rho_{A,0}^{(d(\rho_A))})\Big) \\
&= \sup_{\mathrm{tr}[H_A\rho_A]\le nE} S\Big(\mathcal{E}_f(\rho_{A,0})^{(d(\rho_A))}\Big) \\
&= \sup_{\mathrm{tr}[H_A\rho_A]\le nE} S\Big(\mathcal{E}_f(\rho_{A,0})\Big) \\
&= \sup_{\substack{\mathrm{tr}[H_A\rho_A]\le nE \\ d(\rho_A)=0}} S\Big(\mathcal{E}_f(\rho_A)\Big),
\end{aligned}$$

where we have written $\rho_{A,0} = D(-d(\rho_A))\rho_A D(-d(\rho_A))^\dagger$ for the state ρ_A which has been displaced by its first moments such that it is centered, i.e., the first moments of $\rho_{A,0}$ are zero. The first equality follows by this definition. In the second equality, we have used compatibility of the convolution (5) with displacements, and in the third equality, we have used the fact that the von Neumann entropy is invariant under conjugation with unitaries. In the fourth equality, we have used the fact that

$$\begin{aligned}
\mathrm{tr}[H_A\rho_{A,0}] &= \mathrm{tr}\left[\left(\frac{1}{2}\sum_{k=1}^{2n} R_k^2 - \frac{n}{2}\mathbb{1}_A\right)\rho_{A,0}\right] \\
&= \mathrm{tr}\left[\left(\frac{1}{2}\sum_{k=1}^{2n} D(-d(\rho_A))^\dagger R_k^2 D(-d(\rho_A)) - \frac{n}{2}\mathbb{1}_A\right)\rho_A\right] \\
&= \mathrm{tr}\left[\left(\frac{1}{2}\sum_{k=1}^{2n} (R_k - d_k(\rho_A))^2 - \frac{n}{2}\mathbb{1}_A\right)\rho_A\right] \\
&= \mathrm{tr}\left[\left(\frac{1}{2}\sum_{k=1}^{2n} R_k^2 - \frac{n}{2}\mathbb{1}_A\right)\rho_A\right] - \sum_{k=1}^{2n} d_k(\rho_A)\mathrm{tr}[R_k\rho_A] + \frac{1}{2}\sum_{k=1}^{2n} d_k(\rho_A)^2 \\
&= \mathrm{tr}[H_A\rho_A] - \frac{1}{2}\|d(\rho_A)\|^2 \le \mathrm{tr}[H_A\rho_A] \le nE.
\end{aligned}$$

Therefore in order to upper bound the output entropy, we can restrict our consideration to centered states, i.e., states which have zero first moments. The average energy per mode at the output $\mathcal{E}_f(\rho_A)$ is then bounded as

$$
\begin{aligned}
\frac{1}{n}\mathrm{tr}_C[H_C\mathcal{E}_f(\rho_A)] &= \mathrm{tr}\left[\left(\frac{1}{2n}\sum_{k=1}^{2n}R_k^2-\frac{\mathbb{1}}{2}\right)(f\star\rho_A)\right]\\
&= \mathrm{tr}\left[\int_{\mathbb{R}^{2n}}f(\xi)\left(\frac{1}{2n}\sum_{k=1}^{2n}R_k^2-\frac{\mathbb{1}}{2}\right)D(\xi)\rho_A D(\xi)^\dagger\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}\right]\\
&= \mathrm{tr}\left[\int_{\mathbb{R}^{2n}}f(\xi)\left(\frac{1}{2n}\sum_{k=1}^{2n}(R_k+\xi_k)^2-\frac{\mathbb{1}}{2}\right)\rho_A\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}\right]\\
&= \mathrm{tr}\left[\left(\frac{1}{2n}\sum_{k=1}^{2n}R_k^2-\frac{\mathbb{1}}{2}\right)\rho_A\right]+\frac{1}{2n}\sum_{k=1}^{2n}\int_{\mathbb{R}^{2n}}f(\xi)\xi_k^2\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}\\
&\quad+\frac{1}{n}\sum_{k=1}^{2n}\int_{\mathbb{R}^{2n}}f(\xi)\xi_k\mathrm{tr}[R_k\rho_A]\frac{\mathrm{d}^{2n}\xi}{(2\pi)^n}\\
&= \frac{1}{n}\mathrm{tr}[H_A\rho_A]+E_0\le E+E_0.
\end{aligned}
$$

In the last line, we have used that $\mathrm{tr}[R_k\rho_A]=0$ by assumption. Hence by the fact that thermal states maximize the von Neumann entropy among all states with a given average energy, we have that the maximum output entropy is bounded by

$$S(\mathcal{E}_f(\rho_A))\le ng(E+E_0).$$

From the conditional entropy power inequality (13), we obtain

$$
\begin{aligned}
\exp\frac{S(C|M)}{n} &\ge \exp\frac{S(A|M)}{n}+\exp S_0\\
&= \exp\frac{-S(A)}{n}+\exp S_0\\
&\ge \exp\left(-g(E)\right)+\exp S_0.
\end{aligned}
$$

This implies for the mutual information

$$
\begin{aligned}
I(C:M) &= S(\mathcal{E}_f(\rho_A))-S(C|M)_{(\mathcal{E}_f\otimes\mathbb{1}_M)(\rho_{AM})}\\
&\le ng(E+E_0)-n\log\left(e^{-g(E)}+e^{S_0}\right).
\end{aligned}
$$

Therefore, for the entanglement-assisted classical capacity, we have the upper bound

$$C_{\mathrm{ea}}(\mathcal{E}_f)\le ng(E+E_0)-n\log\left(e^{-g(E)}+e^{S_0}\right).$$

IX. A SIMPLE PROOF OF CONVERGENCE RATE OF THE QUANTUM ORNSTEIN-UHLENBECK SEMIGROUP

We consider the quantum Ornstein-Uhlenbeck semigroup which is a one-parameter semigroup of CPTP maps $\{\mathcal{P}^{(\mu,\lambda)}(t)=e^{t\mathcal{L}_{\mu,\nu}}\}_{t\ge0}$ on the one-mode Gaussian quantum system A generated by the Liouvillian

$$\mathcal{L}_{\mu,\lambda}=\mu^2\mathcal{L}_-+\lambda^2\mathcal{L}_+\qquad\text{for }\mu>\lambda>0,$$

where

$$\mathcal{L}_+(\rho)=a^\dagger\rho a-\frac{1}{2}\{aa^\dagger,\rho\}\qquad\text{and}\qquad\mathcal{L}_-(\rho)=a\rho a^\dagger-\frac{1}{2}\{a^\dagger a,\rho\},$$

where a is the ladder operator of A.

The map $\mathcal{P}^{(\mu,\lambda)}(t)$ is equivalent to a beam splitter with transmissivity $\eta=e^{-(\mu^2-\lambda^2)t}$ and environment state $\omega^{(\mu,\lambda)}:=\frac{\mu^2-\lambda^2}{\mu^2}\sum_{k=0}^\infty\left(\frac{\lambda^2}{\mu^2}\right)^k|k\rangle\langle k|$. This is a Gaussian thermal state with the covariance matrix equal to $\Gamma^{(\mu,\lambda)}=\frac{1}{2}\frac{\lambda^2+\mu^2}{\mu^2-\lambda^2}\mathbb{1}_2$. The state $\omega^{(\mu,\lambda)}$ is also the unique fixed point of the quantum Ornstein-Uhlenbeck semigroup with parameters μ and λ. It is known that the qOU semigroup converges in relative entropy to the fixed point at an exponential rate given by the exponent $\mu^2-\lambda^2$,

$$D\left(\mathcal{P}^{(\mu,\lambda)}(t)(\rho)\|\omega^{(\mu,\lambda)}\right)\le e^{-(\mu^2-\lambda^2)t}D\left(\rho\|\omega^{(\mu,\lambda)}\right)\qquad\text{for all }t\ge0.$$

This is a conjecture stated in Ref. 13, which was proven in Ref. 15 using methods of gradient flow.

Here we want to study a slightly different, more general scenario: We consider a bipartite quantum system AM, where the system A undergoes a qOU evolution. We are going to show a similar convergence statement in this situation, namely, that the system converges in relative entropy to the product state $\omega_A^{(\mu,\lambda)} \otimes \mathrm{tr}_A(\rho_{AM})$ at an exponential rate.

Theorem 9. *We have for any quantum state ρ_{AM},*

$$D\Big((\mathcal{P}^{(\mu,\lambda)}(t) \otimes \mathbb{1}_M)(\rho_{AM})\|\omega_A^{(\mu,\lambda)} \otimes \rho_M\Big) \le e^{-(\mu^2-\lambda^2)t} D\Big(\rho_{AM}\|\omega_A^{(\mu,\lambda)} \otimes \rho_M\Big),$$

where $\rho_M = tr_A(\rho_{AM})$ is the marginal state of ρ_{AM} on the system M. In particular, Eq. (7) holds.

Proof. Write $\rho_{AM}(t) = (\mathcal{P}_{\mu,\lambda}(t) \otimes \mathbb{1}_M)(\rho_{AM})$, we then have

$$\begin{aligned} D\Big(\rho_{AM}(t)\|\omega_A^{(\mu,\lambda)} \otimes \rho_M\Big) &= -S(A|M)_{\rho_{AM}(t)} - \mathrm{tr}\Big(\rho_A(t) \log \omega_A^{(\mu,\lambda)}\Big) \\ &\le -\eta S(A|M)_{\rho_{AM}} - (1-\eta) S(\omega_A^{(\mu,\lambda)}) \\ &\quad -\eta \mathrm{tr}\Big(\rho_A \log \omega_A^{(\mu,\lambda)}\Big) - (1-\eta)\mathrm{tr}\Big(\omega_A^{(\mu,\lambda)} \log \omega_A^{(\mu,\lambda)}\Big) \\ &= e^{-(\mu^2-\lambda^2)t} D\Big(\rho_{AM}\|\omega_A^{(\mu,\lambda)} \otimes \rho_M\Big). \end{aligned}$$

□

This implies exponential convergence to the fixed point both on bipartite systems and the result (7).

X. CONCLUSION

We have established a conditional entropy power inequality for classical noise channels in bosonic quantum systems, modeled by the convolution (5). This inequality implies the unconditioned entropy power inequality for this convolution and lifts regularity problems in previous proofs in this area. In the conditioned case, this inequality is optimal, while the optimal inequality in the unconditioned case remains unsolved. This situation is analogous to the situation for the beam splitter,[30] where the optimal unconditioned inequality is conjectured to be the entropy photon-number inequality,[31] which states that couples of thermal Gaussian input states minimize the output entropy of the beam splitter among all the couples of independent input states, each with a given entropy. The entropy photon-number inequality has been recently proven for the one-mode beam splitter in the particular case where one of the two inputs is a thermal Gaussian state[32–37] and in some very special cases for the multi-mode beam splitter,[38,39] and it otherwise remains an open challenging conjecture (see Ref. 40 for a review). Similarly, an analogous optimal inequality has been conjectured for the quantum additive noise channel.[30] While the validity of this inequality remains an open problem (besides the special case covered in Ref. 37), the conditional entropy power inequality proven in this paper is optimal and settles the problem in the presence of quantum memory.

We have used our new conditional entropy power inequality to provide upper bounds on the entanglement-assisted classical capacity of quantum non-Gaussian additive noise channels and to prove conditional quantum versions of various celebrated results from geometric analysis. Moreover, we have shown how conditional entropy power inequalities can be used to study the convergence rate of quantum dynamical semigroups, giving a simple and short proof of the exponential convergence of the quantum Ornstein-Uhlenbeck semigroup in relative entropy.

ACKNOWLEDGMENTS

G.D.P. thanks Dario Trevisan for useful comments.

G.D.P. acknowledges financial support from the European Research Council (ERC Grant Agreement Nos. 337603 and 321029), the Danish Council for Independent Research (Sapere Aude), VILLUM FONDEN via the QMATH Centre of Excellence (Grant No. 10059), and the Marie

Skłodowska-Curie Action GENIUS (Grant No. 792557). S.H. is supported by the Technische Universität München - Institute for Advanced Study, and funded by the German Excellence Initiative and the European Union Seventh Framework Programme under Grant Agreement No. 291763. S.H. acknowledges additional support by DFG Project No. K05430/1-1.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 792557.

[1] W. Beckner, "Inequalities in Fourier analysis," Ann. Math. **102**, 159–182 (1975).
[2] H. Brascamp and E. Lieb, "Best constants in Young's inequality, its converse and its generalization to more than three functions," Adv. Math. **20**, 151–172 (1976).
[3] C. E. Shannon, "A mathematical theory of communication," Bell Sys. Tech. J. **27**, 623–656 (1948).
[4] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," Inf. Control **2**, 101–112 (1959).
[5] N. Blachman, "The convolution inequality for entropy powers," IEEE Trans. Inf. Theory **11**, 267–271 (1965).
[6] A. Dembo, T. M. Cover, and J. A. Thomas, "Information theoretic inequalities," IEEE Trans. Inf. Theory **37**, 1501–1518 (1991).
[7] R. König and G. Smith, "The entropy power inequality for quantum systems," IEEE Trans. Inf. Theory **60**, 1536–1548 (2014).
[8] G. D. Palma, A. Mari, and V. Giovannetti, "A generalization of the entropy power inequality to bosonic quantum systems," Nat. Photonics **8**, 958–964 (2014).
[9] G. D. Palma, A. Mari, S. Lloyd, and V. Giovannetti, "Multimode quantum entropy power inequality," Phys. Rev. A **91**, 032320 (2015).
[10] R. König, "The conditional entropy power inequality for Gaussian quantum states," J. Math. Phys. **56**, 022201 (2015).
[11] G. De Palma and D. Trevisan, "The conditional entropy power inequality for bosonic quantum systems," Commun. Math. Phys. **360**, 639 (2018).
[12] R. Werner, "Quantum harmonic analysis on phase space," J. Math. Phys. **25**, 1404–1411 (1984).
[13] S. Huber, R. König, and A. Vershynina, "Geometric inequalities from phase space translations," J. Math. Phys. **58**, 012206 (2017).
[14] N. Datta, Y. Pautrat, and C. Rouzé, "Contractivity properties of a quantum diffusion semigroup," J. Math. Phys. **58**, 012205 (2017).
[15] E. A. Carlen and J. Maas, "Gradient flow and entropy inequalities for quantum Markov semigroups with detailed balance," J. Funct. Anal. **273**, 1810–1869 (2017).
[16] A. S. Holevo, *Quantum Systems, Channels, Information: A Mathematical Introduction* (De Gruyter, 2013).
[17] Y. Bardet, "Estimating the decoherence time using non-commutative functional inequalities," preprint arXiv:1710.01039 [quant-ph] (2017).
[18] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, 2017).
[19] A. S. Holevo, M. Sohma, and O. Hirota, "Capacity of quantum Gaussian channels," Phys. Rev. A **59**, 1820–1828 (1999).
[20] M. M. Wolf, G. Giedke, and J. I. Cirac, "Extremality of Gaussian quantum states," Phys. Rev. Lett. **96**, 080502 (2006).
[21] F. Furrer, M. Berta, M. Tomamichel, V. B. Scholz, and M. Christandl, "Position-momentum uncertainty relations in the presence of quantum memory," J. Math. Phys. **55**, 122205 (2014).
[22] G. J. Murphy, "Direct limits and tensor products," in *C^*–Algebras and Operator Theory*, edited by G. J. Murphy (Academic Press, San Diego, 1990), Chap. 6, pp. 173–216.
[23] M. Takesaki, *Theory of Operator Algebras I* (Springer, 2001).
[24] L. Ambrosio, N. Gigli, and G. Savare, "Gradient flows," in *Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics (Birkhäuser Basel, ETH Zürich, 2008).
[25] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of Bounded Variation and Free Discontinuity Problems* (Oxford Science Publications, Clarendon Press, 2000).
[26] G. Buttazzo, *Semicontinuity, Relaxation, and Integral Representation in the Calculus of Variations*, Pitman Research Notes in Mathematics Series (Longman Scientific & Technical, 1989).
[27] C. Goffman, J. Serrin *et al.*, "Sublinear functions of measures and variational integrals," Duke Math. J. **31**, 159–178 (1964).
[28] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2017).
[29] M. M. Wilde and H. Qi, "Energy-constrained private and quantum capacities of quantum channels," preprint arXiv:1609.01997 [quant-ph] (2016).
[30] S. Huber and R. König, "Coherent state coding approaches the capacity of non-Gaussian bosonic noise channels," J. Phys. A: Math. Theor. **51**, 184001 (2018).
[31] S. Guha, B. Erkmen, and J. Shapiro, "The entropy photon-number inequality and its consequences," in *Information Theory and Applications Workshop* (IEEE, 2008), Vol. 2008, pp. 128–130.
[32] G. D. Palma, D. Trevisan, and V. Giovannetti, "Passive states optimize the output of bosonic Gaussian quantum channels," IEEE Trans. Inf. Theory **62**, 2895–2906 (2016).
[33] G. D. Palma, D. Trevisan, and V. Giovannetti, "Gaussian states minimize the output entropy of the one-mode quantum attenuator," IEEE Trans. Inf. Theory **63**, 728–737 (2017).
[34] G. D. Palma, "Gaussian optimizers and other topics in quantum information," Ph.D. thesis, Scuola Normale Superiore, Pisa, Italy, 2016, supervisor: Professor Vittorio Giovannetti; e-print arXiv:1710.09395.
[35] H. Qi, M. M. Wilde, and S. Guha, "On the minimum output entropy of single-mode phase-insensitive Gaussian channels," preprint arXiv:1607.05262 [quant-ph] (2016).

[36] G. De Palma, D. Trevisan, and V. Giovannetti, "The one-mode quantum-limited Gaussian attenuator and amplifier have Gaussian maximizers," Ann. Henri Poincaré **19**, 2919–2953 (2018).
[37] G. D. Palma, D. Trevisan, and V. Giovannetti, "Gaussian states minimize the output entropy of one-mode quantum Gaussian channels," Phys. Rev. Lett. **118**, 160503 (2017).
[38] G. De Palma, D. Trevisan, and V. Giovannetti, "Multimode Gaussian optimizers for the Wehrl entropy and quantum Gaussian channels," preprint arXiv:1705.00499 [quant-ph] (2017).
[39] V. Giovannetti, A. Holevo, and R. García-Patron, "A solution of Gaussian optimizer conjecture for quantum channels," Commun. Math. Phys. **334**, 1553–1571 (2015).
[40] G. De Palma, D. Trevisan, V. Giovannetti, and L. Ambrosio, "Gaussian optimizers for entropic inequalities in quantum information," J. Math. Phys. **59**, 081101 (2018).

B Further articles

B.1 Uncertainty relations: An operational approach to the error-disturbance tradeoff

Uncertainty relations: An operational approach to the error-disturbance tradeoff

Joseph M. Renes, Volkher B. Scholz, and Stefan Huber

The notions of error and disturbance appearing in quantum uncertainty relations are often quantified by the discrepancy of a physical quantity from its ideal value. However, these real and ideal values are not the outcomes of simultaneous measurements, and comparing the values of unmeasured observables is not necessarily meaningful according to quantum theory. To overcome these conceptual difficulties, we take a different approach and define error and disturbance in an operational manner. In particular, we formulate both in terms of the probability that one can successfully distinguish the actual measurement device from the relevant hypothetical ideal by any experimental test whatsoever. This definition itself does not rely on the formalism of quantum theory, avoiding many of the conceptual difficulties of usual definitions. We then derive new Heisenberg-type uncertainty relations for both joint measurability and the error-disturbance tradeoff for arbitrary observables of finite-dimensional systems, as well as for the case of position and momentum. Our relations may be directly applied in information processing settings, for example to infer that devices which can faithfully transmit information regarding one observable do not leak any information about conjugate observables to the environment. We also show that Englert's wave-particle duality relation [PRL 77, 2154 (1996)] can be viewed as an error-disturbance uncertainty relation.

B.1.1 Definitions of error and disturbance

There is no canonical consensus regarding the definition of error and disturbance of a measurement apparatus. The definitions we make here are particular choices, which have some conceptual advantages compared to other approaches, and for which we can prove uncertainty relations, as we show later.

We want to define the error ε_X an apparatus $\mathcal{E}$ makes relative to an ideal measurement $\mathcal{Q}_X$ of an observable X. For this, we use the distinguishability of the two channels, taking only the classical output of the apparatus. We want to allow for classical postprocessing. Let us describe the ideal measurement and the apparatus by the Heisenberg-picture channels $\mathcal{Q}_X : L^\infty(\mathrm{X}) \to \mathcal{B}(\mathcal{H}_A)$ and $\mathcal{E} : \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathrm{Y}) \to \mathcal{B}(\mathcal{H}_A)$. Let us further allow for an arbitrary classical postprocessing operation $\mathcal{R} : L^\infty(\mathrm{X}) \to L^\infty(\mathrm{Y})$. Writing $\mathcal{T}_B : L^\infty(\mathrm{X}) \to \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathrm{X})$ for the partial trace map, $\mathcal{T}_B(f) = \mathbb{1}_B \otimes f$, we define

$$\varepsilon_X(\mathcal{E}) := \inf_{\mathcal{R}} \delta(\mathcal{Q}_X, \mathcal{E}\mathcal{R}\mathcal{T}_B) \ .$$

Here the distinguishability δ is defined via the *completely bounded norm* (cb norm), given by $\delta(\mathcal{E}_1, \mathcal{E}_2) = \frac{1}{2} \|\mathcal{E}_1 - \mathcal{E}_2\|_{\mathrm{cb}}$, where $\|T\|_{\mathrm{cb}} := \sup_{n\in\mathbb{N}} \|\mathbb{1}_{\mathbb{C}^n} \otimes T\|_\infty$.

In order to define the disturbance an apparatus $\mathcal{E} : \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(X) \to \mathcal{B}(\mathcal{H}_A)$ causes to an observable Z, let $\mathcal{Q}_Z : L^\infty(\mathrm{Z}) \to \mathcal{B}(\mathcal{H}_A)$ be the ideal Z measurement and let $\mathcal{R} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathrm{X})$ be a recovery map which acts on the output of $\mathcal{E}$ conditioned on the value of the classical output X. We then define the measurement disturbance as the measurement error after using the best recovery map,

$$\nu_Z(\mathcal{E}) := \inf_{\mathcal{R}} \delta(\mathcal{Q}_Z, \mathcal{E}\mathcal{R}\mathcal{T}_{\mathrm{Y}}\mathcal{Q}_Z) \ .$$

Finally, we want to define *preparation disturbance.* Consider a channel $\mathcal{P}_Z : \mathcal{B}(\mathcal{H}_A) \to L^\infty(\mathrm{Z})$ whose dual in the Schrödinger picture prepares the eigenstates of Z, and consider recovery operations $\mathcal{R} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathrm{X})$. Then we define the preparation disturbance as

$$\eta_Z(\mathcal{E}) := \inf_{\mathcal{R}} \delta(\mathcal{P}_Z, \mathcal{P}_Z \mathcal{E} \mathcal{R} \mathcal{T}_{\mathrm{Y}}) \ .$$

This disturbance measure is related to the distinguishability of the ideal preparation device $\mathcal{P}_Z$ and $\mathcal{P}_Z$ followed by the apparatus $\mathcal{E}$ and the best possible recovery operation $\mathcal{R}$.

In the case of finite-dimensional state preparation where $\mathcal{H} = \mathbb{C}^d$, we also define a figure of "demerit" which compares the functionality of our apparatus with the worst-case behavior, instead of the best-case behavior. This is defined as

$$\hat{\eta}_Z(\mathcal{E}) := \frac{d-1}{d} - \inf_{\mathcal{C}\text{: constant}} \delta(\mathcal{C}, \mathcal{P}_Z \mathcal{E}) \ ,$$

where the optimization is carried out over all constant maps $\mathcal{C} : \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathrm{X}) \to L^\infty(\mathrm{Z})$. In this measure, the disturbance is small if it is easy to distinguish the action of $\mathcal{P}_Z\mathcal{E}$ from having a constant output.

B.1.2 Uncertainty relations in finite dimensions

We define the following measures of complementarity which appear in our uncertainty relations: $c_M(X, Z) := \nu_Z(\mathcal{Q}_X)$, $c_P(X, Z) := \eta_Z(\mathcal{Q}_X)$, and $\hat{c}_P(X, Z) := \hat{\eta}_Z(\mathcal{Q}_X)$. We have the following closed-form lower bounds on these measures of complementarity, where we write $\{|\phi_x\rangle\}_{x\in\mathrm{X}}$ and $\{|\theta_z\rangle\}_{z\in\mathrm{Z}}$ for the eigenvectors of the nondegenerate observables X and Z:

$$\begin{aligned} c_M(X, Z) &\geq 1 - \frac{1}{d} \sum_x \max_z |\langle \phi_x | \theta_z \rangle|^2 \ , \\ c_P(X, Z) &\geq 1 - \frac{1}{d} \sum_x \max_z |\langle \phi_x | \theta_z \rangle|^2 \ , \\ \hat{c}_P(X, Z) &\geq \frac{d-1}{d} - \max_z \frac{1}{2} \sum_x \left| \frac{1}{d} - |\langle \phi_x | \theta_z \rangle|^2 \right| \ . \end{aligned}$$

Then we have the following uncertainty relations:

Theorem B.1.1. *For any two observables X and Z and any quantum instrument $\mathcal{E}$,*

$$\begin{aligned} \sqrt{2\varepsilon_X(\mathcal{E})} + \nu_Z(\mathcal{E}) &\geq c_M(X, Z) \qquad \text{and} \\ \varepsilon_X(\mathcal{E}) + \sqrt{2\nu_Z(\mathcal{E})} &\geq c_M(Z, X) \ . \end{aligned}$$

B.1.3 Uncertainty relations for position and momentum

In this setup, we want to consider instruments which measure position or momentum with a finite precision. For a bounded function $\alpha \in L^2(\mathrm{Q})$ (where Q corresponds to the outcome set $\mathbb{R}$ of the position measurement), we define the finite-precision position measurement instrument $\mathcal{E}_\alpha : L^\infty(\mathrm{Q}) \otimes \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ by

$$\mathcal{E}_\alpha(f \otimes a) = \int \mathrm{d}q f(q) A_{q;\alpha}^\dagger a A_{q;\alpha} \ ,$$

where $A_{q;\alpha}\psi(q') = \alpha(q - q')\psi(q')$ for all $\psi \in L^2(\mathrm{Q})$ and $q' \in \mathrm{Q}$. In this model, setting the width of the function α sets the precision limit of the instrument. We want to focus on Gaussian limits on precision, where α is given by the square root of a normalized Gaussian of variance σ^2. In this picture, the Stinespring dilation of an ideal σ-limited measurement device is a Gaussian unitary. The definitions of measurement error and disturbance are the same as in the previous section, except that the factor $\frac{d-1}{d}$ in this setting is replaced by 1. We will also not focus on the figure of "demerit" $\hat{\eta}$ in this setting, since any non-constant channel can be distinguished from a constant one by putting in states of arbitrarily high momentum. We have the following uncertainty relations for position and momentum:

Theorem B.1.2. *For any two observables X and Z and any quantum instrument $\mathcal{E}$,*

$$\sqrt{2\varepsilon_X(\mathcal{E})} + \eta_Z(\mathcal{E}) \geq c_P(X,Z) \qquad \textit{and}$$
$$\sqrt{2\varepsilon_X(\mathcal{E})} + \hat{\eta}_Z(\mathcal{E}) \geq \hat{c}_P(X,Z) \ .$$

Theorem B.1.3. *Set $c = 2\sigma_Q\sigma_P$ for any precision values $\sigma_P, \sigma_Q > 0$. Then for any quantum instrument $\mathcal{E}$,*

$$\left.\begin{array}{r}\sqrt{2\varepsilon_Q(\mathcal{E})} + \nu_P(\mathcal{E}) \\ \varepsilon_Q(\mathcal{E}) + \sqrt{2\nu_Q(\mathcal{E})}\end{array}\right\} \geq \frac{1-c^2}{(1+c^{2/3}+c^{4/3})^{3/2}} \qquad \textit{and}$$
$$\sqrt{2\varepsilon_Q(\mathcal{E})} + \eta_P(\mathcal{E}) \geq \frac{(1+c^2)^{1/2}}{\big((1+c^2) + c^{2/3}(1+c^2)^{2/3} + c^{4/3}(1+c^2)^{1/3}\big)^{3/2}} \ .$$

B.1.4 Applications

Given an ideal measurement $\mathcal{Q}_Z$, we define the map $\mathcal{Q}_Z^{\natural} = \mathcal{W}_Z\mathcal{T}_Z$ with $\mathcal{W}_Z : a \to W_Z^{\dagger} a W_Z$, where W_Z is the Stinespring isometry of $\mathcal{Q}_Z$. This map can be seen as performing a Z measurement and immediately forgetting the result. We can then apply our uncertainty relation to make an information-disturbance statement: If acting with a channel $\mathcal{N}$ does not substantially affect the possibility of performing an X measurement, then Z-basis inputs to the complementary channel $\mathcal{N}^{\#}$ result in an essentially constant output. This is formalized in the following corollary.

Corollary B.1.4. *Given a channel $\mathcal{N}$ and complementary channel $\mathcal{N}^{\#}$, suppose that there exists a measurement Λ_X such that $\delta(Q_X, \mathcal{N}\Lambda_X) \leq \varepsilon$. Then there exists a constant channel $\mathcal{C}$ such that*

$$\delta(Q_Z^{\natural}\mathcal{N}^{\#}, \mathcal{C}) \leq \sqrt{2\varepsilon} + \frac{d-1}{d} - \hat{c}_P(X,Z) \ .$$

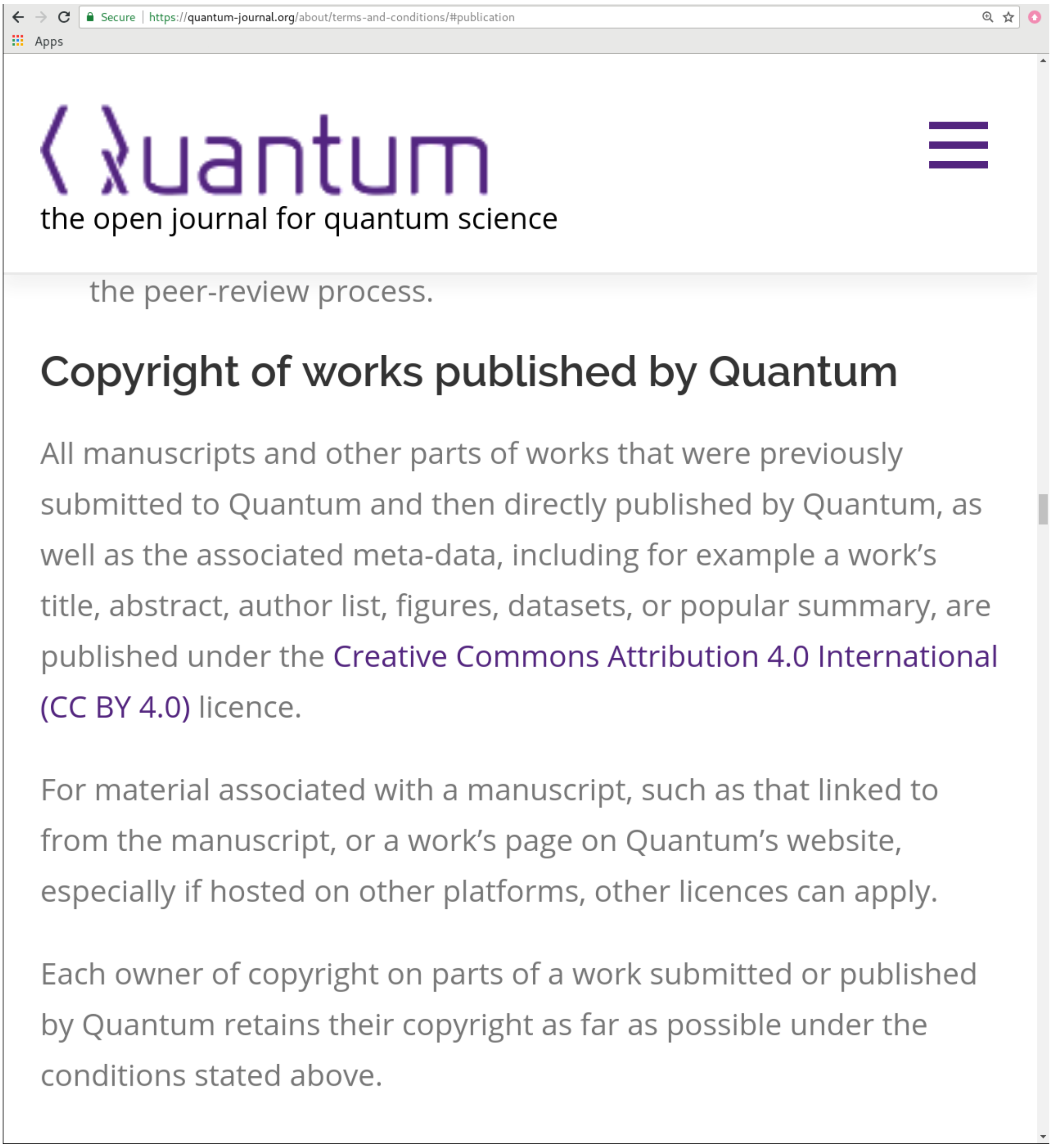
For maximally complementary X and Z, i.e., if

$$|\langle\phi_x|\theta_z\rangle|^2 = \frac{1}{d} \qquad \textit{for all } x \in X, z \in Z \ ,$$

we have $\delta(Q_Z^{\natural}\mathcal{N}^{\#}, \mathcal{C}) \leq \sqrt{2\varepsilon}$.

Permission to include:

Joseph M. Renes, Volkher B. Scholz, and Stefan Huber
Uncertainty relations: An operational approach to the error-disturbance tradeoff.
Quantum 1, 20 (2017).



Secure | https://quantum-journal.org/about/terms-and-conditions/#publication
Apps
Quantum
the open journal for quantum science

the peer-review process.

Copyright of works published by Quantum

All manuscripts and other parts of works that were previously submitted to Quantum and then directly published by Quantum, as well as the associated meta-data, including for example a work's title, abstract, author list, figures, datasets, or popular summary, are published under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

For material associated with a manuscript, such as that linked to from the manuscript, or a work's page on Quantum's website, especially if hosted on other platforms, other licences can apply.

Each owner of copyright on parts of a work submitted or published by Quantum retains their copyright as far as possible under the conditions stated above.

Uncertainty relations: An operational approach to the error-disturbance tradeoff

Joseph M. Renes[1], Volkher B. Scholz[1,2], and Stefan Huber[1,3]

[1] *Institute for Theoretical Physics, ETH Zürich, Switzerland*
[2] *Department of Physics, Ghent University, Belgium*
[3] *Department of Mathematics, Technische Universität München, Germany*

Accepted in Quantum July 10, 2017

The notions of error and disturbance appearing in quantum uncertainty relations are often quantified by the discrepancy of a physical quantity from its ideal value. However, these real and ideal values are not the outcomes of simultaneous measurements, and comparing the values of unmeasured observables is not necessarily meaningful according to quantum theory. To overcome these conceptual difficulties, we take a different approach and define error and disturbance in an operational manner. In particular, we formulate both in terms of the probability that one can successfully distinguish the actual measurement device from the relevant hypothetical ideal by any experimental test whatsoever. This definition itself does not rely on the formalism of quantum theory, avoiding many of the conceptual difficulties of usual definitions. We then derive new Heisenberg-type uncertainty relations for both joint measurability and the error-disturbance tradeoff for arbitrary observables of finite-dimensional systems, as well as for the case of position and momentum. Our relations may be directly applied in information processing settings, for example to infer that devices which can faithfully transmit information regarding one observable do not leak any information about conjugate observables to the environment. We also show that Englert's wave-particle duality relation [Phys. Rev. Lett. **77**, 2154 (1996)] can be viewed as an error-disturbance uncertainty relation.

1 Introduction

It is no overstatement to say that the uncertainty principle is a cornerstone of our understanding of quantum mechanics, clearly marking the departure of quantum physics from the world of classical physics. Heisenberg's original formulation in 1927 mentions two facets to the principle. The first restricts the joint measurability of observables, stating that noncommuting observables such as position and momentum can only be simultaneously determined with a characteristic amount of indeterminacy [1, p. 172] (see [2, p. 62] for an English translation). The second describes an error-disturbance tradeoff, noting that the more precise a measurement of one observable is made, the greater the disturbance to noncommuting observables [1, p. 175] ([2, p. 64]). The two are of course closely related, and Heisenberg argues for the former on the basis of the latter. Neither version can be taken merely as a limitation on measurement of otherwise well-defined values of position and momentum, but rather as questioning the sense in which values of two noncommuting observables can even be said to simultaneously exist. Unlike classical mechanics, in the framework of quantum mechanics we cannot necessarily regard unmeasured quantities as physically meaningful.

More formal statements were constructed only much later, due to the lack of a precise mathematical description of the measurement process in quantum mechanics. Here we must be careful to draw a distinction between statements addressing Heisenberg's original notions of uncertainty from those, like the standard Kennard-Robertson uncertainty relation [3, 4], which address the impossibility of finding a quantum state with well-defined values for noncommuting observables. Entropic uncertainty relations [5, 6] are also an example of this class; see [7] for a review. Joint measurability has a longer history, going back at least to the seminal work of Arthurs and Kelly [8] and continuing in [9–27]. Quantitative error-disturbance relations have only been formulated relatively recently, going back at least to Braginsky and Khalili [28, Chap. 5] and continuing in [20, 29–35].

Beyond technical difficulties in formulating uncertainty relations, there is a perhaps more difficult conceptual hurdle in that the intended consequences of the uncertainty principle seem to preclude their own straightforward formalization. To find a relation between, say, the error of a position measurement and its disturbance to momentum in a given experimental setup like the gamma ray microscope would seem to require comparing the actual values of position and momentum with their supposed ideal values. However, according to the uncertainty principle itself, we should be wary of simultaneously ascribing well-defined values to the actual and ideal position and momentum since they do not correspond to commuting observables. Thus, it is not immediately clear how to formulate either meaningful measures of error and disturbance, for instance as mean-square deviations between real and ideal values, or a meaningful relation between them.[1] This question is the subject of much ongoing debate [25, 30, 36–39].

[1] Uncertainty relations like the Kennard-Robertson bound or entropic relations do not face this issue as they do not attempt to compare actual and ideal values of the observables.

Without drawing any conclusions as to the ultimate success or failure of this program, in this paper we propose a completely different approach which we hope sheds new light on these conceptual difficulties. Here, we define error and disturbance in an operational manner and ask for uncertainty relations that are statements about the properties of measurement devices, not of fixed experimental setups or of physical quantities themselves. More specifically, we define error and disturbance in terms of the *distinguishing probability*, the probability that the actual behavior of the measurement apparatus can be distinguished from the relevant ideal behavior in any single experiment whatsoever. To characterize measurement error, for example, we imagine a black box containing either the actual device or the ideal device. By controlling the input and observing the output we can make an informed guess as to which is the case. We then attribute a large measurement error to the measurement apparatus if it is easy to tell the difference, so that there is a high probability of correctly guessing, and a low error if not; of course we pick the optimal input states and output measurements for this purpose. In this way we do not need to attribute a particular ideal value of the observable to be measured, we do not need to compare actual and ideal values themselves (nor do we necessarily even care what the possible values are), and instead we focus squarely on the properties of the device itself. Intuitively, we might expect that calibration provides the strictest test, i.e. inputting states with a known value of the observable in question. But in fact this is not the case, as entanglement at the input can increase the distinguishability of two measurements. The merit of this approach is that the notion of distinguishability itself does not rely on any concepts or formalism of quantum theory, which helps avoid conceptual difficulties in formalizing the uncertainty principle.

Defining the disturbance an apparatus causes to an observable is more delicate, as an observable itself does not have a directly operational meaning (as opposed to the measurement of an observable). But we can consider the disturbance made either to an ideal measurement of the observable or to ideal preparation of states with well-defined values of the observable. In all cases, the error and disturbance measures we consider are directly linked to a well-studied norm on quantum channels known as the completely bounded norm or diamond norm. We can then ask for bounds on the error and disturbance quantities for two given observables that every measurement apparatus must satisfy. In particular, we are interested in bounds depending only on the chosen observables and not the particular device. Any such relation is a statement about measurement devices themselves and is not specific to the particular experimental setup in which they are used. Nor are such relations statements about the values or behavior of physical quantities themselves. In this sense, we seek statements of the uncertainty principle akin to Kelvin's form of the second law of thermodynamics as a constraint on thermal machines, and not like Clausius's or Planck's form involving the behavior of physical quantities (heat and entropy, respectively). By appealing to a fundamental constraint on quantum dynamics, the continuity (in the completely bounded norm) of the Stinespring dilation [40, 41], we find error-disturbance uncertainty relations for arbitrary observables in finite dimensions, as well as for position and momentum. Furthermore, we show how the relation for measurement error and measurement disturbance can be transformed into a joint-measurability uncertainty relation. Interestingly, we also find that Englert's wave-particle duality relation [42] can be viewed as an error-disturbance relation.

The case of position and momentum illustrates the stark difference between the kind of uncertainty statements we can make in our approach with one based on the notion of comparing real and ideal values. Take the notion of joint measurability, where we would like to formalize the notion that no device can accurately measure both position and momentum. In the latter approach one would first try to quantify the amount of position or momentum error made by a device as the discrepancy to the true value, and then show that they cannot both be small. The errors would be in units of position or momentum, respectively, and the hoped-for uncertainty relation would pertain to these values. Here, in contrast, we focus on the performance of the actual device relative to fixed ideal devices, in this case idealized separate measurements of position or momentum. Importantly, we need not think of the ideal measurement as having infinite precision. Instead, we can pick any desired precision and ask if the behavior of the actual device is essentially the same as this precision-limited ideal. Now the position and momentum errors do not have units of these quantities (they are unitless and always lie between zero and one), but instead *depend on the desired precision*. Our uncertainty relation then implies that both errors cannot be small if we demand high precision in both position and momentum. In particular, when the product of the scales of the two precisions is small compared to Planck's constant, then the errors will be bounded away from zero (see Theorem 3 for a precise statement). It is certainly easier to have a small error in this sense when the demanded precision is low, and this accords nicely with the fact that sufficiently-inaccurate joint measurement is possible. Indeed, we find no bound on the errors for low precision.

An advantage and indeed a separate motivation of an operational approach is that bounds involving operational quantities are often useful in analyzing information processing protocols. For example, entropic uncertainty relations, which like the Robertson relation characterize quantum states, have proven very useful in establishing simple proofs of the security of quantum key distribution [6, 7, 43–45]. Here we show that the error-disturbance relation implies that quantum channels which can faithfully transmit information regarding one observable do not leak any information whatsoever about conjugate observables to the environment. This statement cannot be derived from entropic relations, as it holds for all channel inputs. It can be used to construct leakage-resilient classical computers from fault-tolerant quantum computers [46], for instance.

The remainder of the paper is structured as follows. In the next section we give the mathematical background necessary to state our results, and describe how the general notion of distinguishability is related to the completely bounded norm (cb norm) in this setting. In Section 3 we define our error and disturbance measures precisely. Section 4 presents the error-disturbance tradeoff relations for finite dimensions, and details how joint measurability relations can be obtained from them. Section 5 considers the error-disturbance tradeoff relations for position and momentum. Two applications of the tradeoffs are given in Section 6: a formal statement of the information disturbance tradeoff for information about noncommuting observables and the connection between error-disturbance tradeoffs and Englert's wave-particle duality relations. In Section 7 we compare our results to previous approaches in more detail, and finally we finish with open questions in Section 8.

2 Mathematical setup

2.1 Distinguishability

The notion of the distinguishing probability is independent of the mathematical framework needed to describe quantum systems, so we give it first. Consider an apparatus $\mathcal{E}$ which in some way transforms an input A into an output B. To describe how different $\mathcal{E}$ is from another such apparatus $\mathcal{E}'$, we can imagine the following scenario. Suppose that we randomly place either $\mathcal{E}$ or $\mathcal{E}'$ into a black box such that we no longer have any access to the inner workings of the device, only its inputs and outputs. Now our task is to guess which device is actually in the box by performing a single experiment, feeding in any desired input and observing the output in any manner of our choosing. In particular, the inputs and measurements can and should depend on $\mathcal{E}$ and $\mathcal{E}'$. The probability of making a correct guess, call it $p_{\text{dist}}(\mathcal{E},\mathcal{E}')$, ranges from $\frac{1}{2}$ to 1, since we can always just make a random guess without doing any experiment on the box at all. Therefore it is more convenient to work with the distinguishability measure

$$\delta(\mathcal{E},\mathcal{E}') := 2p_{\text{dist}}(\mathcal{E},\mathcal{E}') - 1\,, \tag{1}$$

which ranges from zero (completely indistinguishable) to one (completely distinguishable). Later on we will show this quantity takes a specific mathematical form in quantum mechanics. But note that the definition implies that the distinguishability is monotonic under concatenation with a channel $\mathcal{F}$ to both $\mathcal{E}$ and $\mathcal{E}'$, since this just restricts the possible tests. That is, both $\delta(\mathcal{E}\mathcal{F},\mathcal{E}'\mathcal{F}) \leq \delta(\mathcal{E},\mathcal{E}')$ and $\delta(\mathcal{F}\mathcal{E},\mathcal{F}\mathcal{E}') \leq \delta(\mathcal{E},\mathcal{E}')$ hold for all channels $\mathcal{F}$ whose inputs and outputs are such that the channel concatenation is sensible. Here and in the remainder of the paper, we denote concatenation of channels by juxtaposition, while juxtaposition of operators denotes multiplication as usual.

2.2 Systems, algebras, channels, and measurements

In the finite-dimensional case we will be interested in two arbitrary nondegenerate observables denoted X and Z. Only the eigenvectors of the observables will be relevant, call them $|\varphi_x\rangle$ and $|\theta_z\rangle$, respectively. In infinite dimensions we will confine our analysis to position Q and momentum P, taking $\hbar = 1$. The analog of Q and P in finite dimensions are canonically conjugate observables X and Z for which $|\varphi_x\rangle = \frac{1}{\sqrt{d}}\sum_z \omega^{xz}\,|\theta_z\rangle$, where d is the dimension and ω is a primitive dth root of unity.

It will be more convenient for our purposes to adopt the algebraic framework and use the Heisenberg picture, though we shall occasionally employ the Schrödinger picture. In the Heisenberg picture we describe systems chiefly by the algebra of observables on them and describe transformations of systems by quantum channels, completely positive and unital maps from the algebra of observables of the output to the observables of the input [10, 47–50]. This allows us to treat classical and quantum systems on an equal footing within the same framework. When the input or output system is quantum mechanical, the observables are the bounded operators $\mathcal{B}(\mathcal{H})$ from the Hilbert space $\mathcal{H}$ associated with the system to itself. Classical systems, such as the results of measurement or inputs to a state preparation device, take values in a set, call it Y. The relevant algebra of observables here is $L^\infty(\mathsf{Y})$, the (bounded, measureable) functions on Y. Hybrid systems are described by tensor products, so an apparatus $\mathcal{E}$ which measures a quantum system has an output algebra described by $L^\infty(\mathsf{Y})\otimes\mathcal{B}(\mathcal{H})$. To describe just the measurement result, we keep only $L^\infty(\mathsf{Y})$. We shall occasionally denote the input and output spaces explicitly as $\mathcal{E}_{A\to\mathsf{Y}B}$ when useful.

For arbitrary input and output algebras $\mathcal{A}_A$ and $\mathcal{A}_B$, quantum channels are precisely those maps $\mathcal{E}$ which are unital, $\mathcal{E}(\mathbb{1}_B) = \mathbb{1}_A$, and completely positive, meaning that not only does $\mathcal{E}$ map positive elements of $\mathcal{A}_B$ to positive elements of $\mathcal{A}_A$, it also maps positive elements of $\mathcal{A}_B \otimes \mathcal{B}(\mathbb{C}^n)$ to positive elements of $\mathcal{A}_A \otimes \mathcal{B}(\mathbb{C}^n)$ for all integer n. This requirement is necessary to ensure that channels act properly on entangled systems.

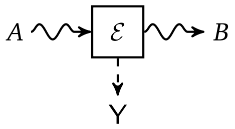


Figure 1: A general quantum apparatus $\mathcal{E}$. The apparatus measures a quantum system A giving the output Y. In so doing, $\mathcal{E}$ also transforms the input A into the output system B. Here the wavy lines denote quantum systems, the dashed lines classical systems. Formally, the apparatus is described by a quantum instrument.

A general measurement apparatus has both classical and quantum outputs, corresponding to the measurement result and the post-measurement quantum system. Channels describing such devices are called *quantum instruments*; we will call the channel describing just the measurement outcome a *measurement*. In finite dimensions any measurement can be seen as part of a quantum instrument, but not so for idealized position or momentum measurements, as shown in Theorem 3.3 of [10] (see page 57). Technically, we may anticipate the result since the post-measurement state of such a device would presumably be a delta function located at the value of the measurement, which is not an element of $L^2(\mathrm{Q})$. This need not bother us, though, since it is not operationally meaningful to consider a position measurement instrument of infinite precision. And indeed there is no mathematical obstacle to describing finite-precision position measurement by quantum instruments, as shown in Theorem 6.1 (page 67 of [10]). For any bounded function $\alpha \in L^2(\mathrm{Q})$ we can define the instrument $\mathcal{E}_\alpha : L^\infty(\mathrm{Q}) \otimes \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ by

$$\mathcal{E}_\alpha(f \otimes a) = \int \mathrm{d}q\, f(q) A^*_{q;\alpha} a A_{q;\alpha}\,, \tag{2}$$

where $A_{q;\alpha}\psi(q') = \alpha(q-q')\psi(q')$ for all $\psi \in L^2(\mathrm{Q})$. The classical output of the instrument is essentially the ideal value convolved with the function α. Thus, setting the width of α sets the precision limit of the instrument.

2.3 Distinguishability as a channel norm

The distinguishability measure is actually a norm on quantum channels, equal (apart from a factor of one half) to the so-called norm of complete boundedness, the cb norm [51–53]. The cb norm is defined as an extension of the operator norm, similar to the extension of positivity above, as

$$\|T\|_{\mathrm{cb}} := \sup_{n\in\mathbb{N}} \|\mathbb{1}_n \otimes T\|_\infty\,, \tag{3}$$

where $\|T\|_\infty$ is the operator norm. Then

$$\delta(\mathcal{E}_1, \mathcal{E}_2) = \tfrac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_{\mathrm{cb}}\,. \tag{4}$$

In the Schrödinger picture we instead extend the trace norm $\|\cdot\|_1$, and the result is usually called the diamond norm [51, 53]. In either case, the extension serves to account for entangled inputs in the experiment to test whether $\mathcal{E}_1$ or $\mathcal{E}_2$ is the actual channel. In fact, entanglement is helpful even when the channels describe projective measurements, as shown by an example given in Appendix A. This expression for the cb or diamond norm is not closed-form, as it requires an optimization. However, in finite dimensions the cb norm can be cast as a convex optimization, specifically as a semidefinite program [54, 55], which makes numerical computation tractable. Further details are given in Appendix B.

2.4 The Stinespring representation and its continuity

According to the Stinespring representation theorem [52, 56], any channel $\mathcal{E}$ mapping an algebra $\mathcal{A}$ to $\mathcal{B}(\mathcal{H})$ can be expressed in terms of an isometry $V : \mathcal{H} \to \mathcal{K}$ to some Hilbert space $\mathcal{K}$ and a representation π of $\mathcal{A}$ in $\mathcal{B}(\mathcal{K})$ such that, for all $a \in \mathcal{A}$,

$$\mathcal{E}(a) = V^* \pi(a) V\,. \tag{5}$$

The isometry in the Stinespring representation is usually called the *dilation* of the channel, and $\mathcal{K}$ the dilation space. In finite-dimensional settings, calling the input A and the output B, one usually considers maps taking $\mathcal{A} = \mathcal{B}(\mathcal{H}_B)$ to $\mathcal{B}(\mathcal{H}_A)$. Then one can choose $\mathcal{K} = \mathcal{H}_B \otimes \mathcal{H}_E$, where $\mathcal{H}_E$ is a suitably large Hilbert space associated to the "environment" of the transformation ($\mathcal{H}_E$ can always be chosen to have dimension $\dim(\mathcal{H}_A)\dim(\mathcal{H}_B)$). The representation π is just $\pi(a) = a \otimes \mathbb{1}_E$. Using the isometry V, we can also construct a channel from $\mathcal{B}(\mathcal{H}_E)$ to $\mathcal{B}(\mathcal{H}_A)$ in the same manner; this is known as the complement $\mathcal{E}^\sharp$ of $\mathcal{E}$.

The advantage of the general form of the Stinespring representation is that we can easily describe measurements, possibly continuous-valued, as well. For the case of finite outcomes, consider the ideal projective measurement $\mathcal{Q}_X$ of the observable X. Choosing a basis $\{|b_x\rangle\}$ of $L^2(\mathsf{X})$ and defining $\pi(\delta_x) = |b_x\rangle\langle b_x|$ for δ_x the function taking the value 1 at x and zero elsewhere, the canonical dilation isometry $W_X : \mathcal{H} \to L^2(\mathsf{X}) \otimes \mathcal{H}$ is given by

$$W_X = \sum_x |b_x\rangle \otimes |\varphi_x\rangle\langle\varphi_x|\,. \tag{6}$$

Note that this isometry defines a quantum instrument, since it can describe both the measurement outcome and the post-measurement quantum system. If we want to describe just the measurement result, we could simply use $W_X = \sum_x |b_x\rangle\,\langle\varphi_x|$ with the same π. More generally, a POVM with elements Λ_x has the isometry $W_X = \sum_x |b_x\rangle \otimes \sqrt{\Lambda_x}$.

For finite-precision measurements of position or momentum, the form of the quantum instrument in (2) immediately gives a Stinespring dilation $W_Q : \mathcal{H} \to \mathcal{K}$ with $\mathcal{K} = L^2(\mathsf{Q}) \otimes \mathcal{H}$ whose action is defined by

$$(W_Q\psi)(q,q') = \alpha(q-q')\psi(q')\,, \tag{7}$$

and where π is just pointwise multiplication on the $L^\infty(\mathsf{Q})$ factor, i.e. for $f \in L^\infty(\mathsf{Q})$, and $a \in \mathcal{B}(\mathcal{H})$, $[\pi(f \otimes a)(\xi \otimes \psi)](q,q') = f(q)\xi(q)\cdot(a\psi)(q')$ for all $\xi \in L^2(\mathsf{Q})$ and $\psi \in \mathcal{H}$.

A slight change to the isometry in (6) gives the dilation of the device which prepares the state $|\varphi_x\rangle$ for classical input x. Formally the device is described by the map $\mathcal{P} : \mathcal{B}(\mathcal{H}) \to L^2(\mathsf{X})$ for which $\mathcal{P}(\Lambda) = \sum_x |b_x\rangle\langle b_x|\,\langle\varphi_x|\Lambda|\varphi_x\rangle$. Now consider $W'_X : L^2(\mathsf{X}) \to \mathcal{H} \otimes L^2(\mathsf{X})$ given by

$$W'_X = \sum_x |\varphi_x\rangle \otimes |b_x\rangle\langle b_x|\,. \tag{8}$$

Choosing $\pi(\Lambda) = \Lambda \otimes \mathbb{1}_{\mathsf{X}}$, we have $\mathcal{P}(\Lambda) = W'^{*}_X \pi(\Lambda) W'_X$.

The Stinespring representation is not unique [41]. Given two representations $(\pi_1, V_1, \mathcal{K}_1)$ and $(\pi_2, V_2, \mathcal{K}_2)$ of the same channel $\mathcal{E}$, there exists a partial isometry $U : \mathcal{K}_1 \to \mathcal{K}_2$ such that $UV_1 = V_2$, $U^*V_2 = V_1$, and $U\pi_1(a) = \pi_2(a)U$ for all $a \in \mathcal{A}$. For the representations π as usually employed for the finite-dimensional case, this last condition implies that U is a partial isometry from one environment to the other, for $U(a \otimes \mathbb{1}_E) = (a \otimes \mathbb{1}_{E'})U$ can only hold for all a if U acts trivially on B. For channels describing measurements, finite or continuous, the last condition implies that any such U is a conditional partial isometry, dependent on the outcome of the measurement result. Thus, for any set of isometries $U_x : \mathcal{H}_S \to \mathcal{H}_R$, $\sum_x |b_x\rangle \otimes U_x|\varphi_x\rangle\langle\varphi_x|U_x^*$ is a valid dilation of $\mathcal{Q}_X$, just as is W_X in (6). Similarly, $(W'_Q\psi)(q,q') = \alpha(q-q')[U_q\psi](q')$ is a valid dilation of $\mathcal{E}_\alpha$ in (2).

The main technical ingredient required for our results is the continuity of the Stinespring representation in the cb norm [40, 41]. That is, channels which are nearly indistinguishable have Stinespring dilations which are close and vice versa. For completely positive and unital maps $\mathcal{E}_1$ and $\mathcal{E}_2$, [40, 41] show that

$$\tfrac{1}{2}\|\mathcal{E}_1 - \mathcal{E}_2\|_{\mathrm{cb}} \leq \inf_{\pi_i, V_i} \|V_1 - V_2\|_\infty \leq \sqrt{\|\mathcal{E}_1 - \mathcal{E}_2\|_{\mathrm{cb}}}\,, \tag{9}$$

where the infimum is taken over all Stinespring representations $(\pi_i, V_i, \mathcal{K}_i)$ of $\mathcal{E}_i$.

2.5 Sequential and joint measurements

Using the Stinespring representation we can easily show that, in principle, any joint measurement can always be decomposed into sequential measurement.

Lemma 1. *Suppose that $\mathcal{E} : L^\infty(\mathsf{X}) \otimes L^\infty(\mathsf{Z}) \to \mathcal{B}(\mathcal{H})$ is a channel describing a joint measurement. Then there exists an apparatus $\mathcal{A} : L^\infty(\mathsf{X}) \otimes \mathcal{B}(\mathcal{H}') \to \mathcal{B}(\mathcal{H})$ and a conditional measurement $\mathcal{M} : L^\infty(\mathsf{X}) \otimes L^\infty(\mathsf{Z}) \to L^\infty(\mathsf{X}) \otimes \mathcal{B}(\mathcal{H}')$ such that $\mathcal{E} = \mathcal{A}\mathcal{M}$.*

Proof. Define $\mathcal{M}' : L^\infty(\mathsf{X}) \to \mathcal{B}(\mathcal{H})$ to be just the X output of $\mathcal{E}$, i.e. $\mathcal{M}'(f) = \mathcal{E}(f \otimes 1)$. Now suppose that $V : \mathcal{H} \to L^2(\mathsf{X}) \otimes L^2(\mathsf{Z}) \otimes \mathcal{H}''$ is a Stinespring representation of $\mathcal{E}$ and $V_X : \mathcal{H} \to L^2(\mathsf{X}) \otimes \mathcal{H}'$ is a representation of $\mathcal{M}'$, both with the standard representation π of L^∞ into L^2. By construction, V is also a dilation of $\mathcal{M}'$, and therefore there exists a partial isometry U_X such that $V = U_X V_X$. More specifically, conditional on the value $\mathsf{X} = x$, each U_x sends $\mathcal{H}'$ to $L^2(\mathsf{Z}) \otimes \mathcal{H}''$. Thus, setting $\mathcal{A}(f \otimes a) = V_X^*(\pi(f) \otimes a)V_X$ and $\mathcal{M}_x(f) = U_x^*(\pi(f) \otimes \mathbb{1})U_x$, we have $\mathcal{E} = \mathcal{A}\mathcal{M}$. □

3 Definitions of error and disturbance

3.1 Measurement error

To characterize the error ε_X an apparatus $\mathcal{E}$ makes relative to an ideal measurement $\mathcal{Q}_X$ of an observable X, we can simply use the distinguishability of the two channels, taking only the classical output of $\mathcal{E}$. Suppose that the apparatus is described by the channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathsf{X}) \to \mathcal{B}(\mathcal{H}_A)$ and the ideal measurement by the channel $\mathcal{Q}_X : L^\infty(\mathsf{X}) \to \mathcal{B}(\mathcal{H}_A)$. To ignore the output system B, we make use of the partial trace map $\mathcal{T}_B : L^\infty(\mathsf{X}) \to \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathsf{X})$ given by $\mathcal{T}_B(f) = \mathbb{1}_B \otimes f$. Then a sensible notion of error is given by $\varepsilon_X(\mathcal{E}) = \delta(\mathcal{Q}_X, \mathcal{E}\mathcal{T}_B)$. If it is easy to tell the ideal measurement apart from the actual device, then the error is large; if it is difficult, then the error is small.

As a general definition, though, this quantity is deficient to two respects. First, we could imagine an apparatus which performs an ideal $\mathcal{Q}_X$ measurement, but simply mislabels the outputs. This leads to $\varepsilon_X(\mathcal{E}) = 1$, even though the ideal measurement is actually performed. Second, we might wish to consider the case that the classical output set of the apparatus is not equal to X itself. For instance, perhaps $\mathcal{E}$ delivers much more output than is expected from $\mathcal{Q}_X$. In this case we also formally have $\varepsilon_X(\mathcal{E}) = 1$, since we can just examine the output to distinguish the two devices.

We can remedy both of these issues by describing the apparatus by the channel $\mathcal{E} : \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathsf{Y}) \to \mathcal{B}(\mathcal{H}_A)$ and just including a further classical postprocessing operation $\mathcal{R} : L^\infty(\mathsf{X}) \to L^\infty(\mathsf{Y})$ in the distinguishability step. Since we are free to choose the best such map, we define

$$\varepsilon_X(\mathcal{E}) := \inf_{\mathcal{R}} \delta(\mathcal{Q}_X, \mathcal{E}\mathcal{R}\mathcal{T}_B). \tag{10}$$

The setup of the definition is depicted in Figure 2.

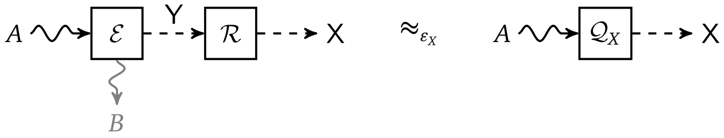


Figure 2: Measurement error. The error made by the apparatus $\mathcal{E}$ in measuring X is defined by how distinguishable the actual device is from the ideal measurement $\mathcal{Q}_X$ in any experiment whatsoever, after suitably processing the classical output Y of $\mathcal{E}$ with the map $\mathcal{R}$. To enable a fair comparison, we ignore the quantum output of the apparatus, indicated in the diagram by graying out B. If the actual and ideal devices are difficult to tell apart, the error is small.

3.2 Measurement disturbance

Defining the disturbance an apparatus $\mathcal{E}$ causes to an observable, say Z, is more delicate, as an observable itself does not have a directly operational meaning. But there are two straightforward ways to proceed: we can either associate the observable with measurement or with state preparation. In the former, we compare how well we can mimic the ideal measurement $\mathcal{Q}_Z$ of the observable after employing the apparatus $\mathcal{E}$, quantifying this using measurement error as before. Additionally, we should allow the use of recovery operations in which we attempt to “restore” the input state as well as possible, possibly conditional on the output of the measurement. Formally, let $\mathcal{Q}_Z : L^\infty(\mathsf{Z}) \to \mathcal{B}(\mathcal{H}_A)$ be the ideal Z measurement and $\mathcal{R}$ be a recovery map $\mathcal{R} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathsf{X})$ which acts on the output of $\mathcal{E}$ conditional on the value of the classical output X (which it then promptly forgets). As depicted in Figure 3, the measurement disturbance is then the measurement error after using the best recovery map:

$$\nu_Z(\mathcal{E}) := \inf_{\mathcal{R}} \delta(\mathcal{Q}_Z, \mathcal{E}\mathcal{R}\mathcal{T}_{\mathsf{Y}}\mathcal{Q}_Z). \tag{11}$$

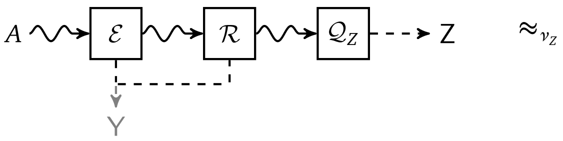


Figure 3: Measurement disturbance. To define the disturbance imparted by an apparatus $\mathcal{E}$ to the measurement of an observable Z, consider performing the ideal $\mathcal{Q}_Z$ measurement on the output B of $\mathcal{E}$. First, however, it may be advantageous to "correct" or "recover" the original input A by some operation $\mathcal{R}$. In general, $\mathcal{R}$ may depend on the output X of $\mathcal{E}$. The distinguishability between the resulting combined operation and just performing $\mathcal{Q}_Z$ on the original input defines the measurement disturbance.

3.3 Preparation disturbance

For state preparation, consider a device with classical input and quantum output that prepares the eigenstates of Z. We can model this by a channel $\mathcal{P}_Z$, which in the Schrödinger picture produces $|\theta_z\rangle$ upon receiving the input z. Now we compare the action of $\mathcal{P}_Z$ to the action of $\mathcal{P}_Z$ followed by $\mathcal{E}$, again employing a recovery operation. Formally, let $\mathcal{P}_Z : \mathcal{B}(\mathcal{H}_A) \to L^\infty(\mathsf{Z})$ be the ideal Z preparation device and consider recovery operations $\mathcal{R}$ of the form $\mathcal{R} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B) \otimes L^\infty(\mathsf{X})$. Then the preparation disturbance is defined as

$$\eta_Z(\mathcal{E}) := \inf_{\mathcal{R}} \delta(\mathcal{P}_Z, \mathcal{P}_Z \mathcal{E} \mathcal{R} \mathcal{T}_{\mathsf{Y}}). \tag{12}$$

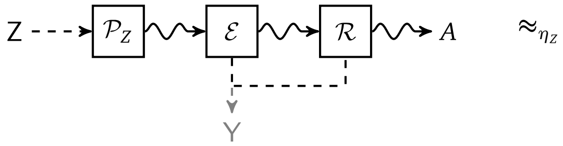


Figure 4: Preparation disturbance. The ideal preparation device $\mathcal{P}_Z$ takes a classical input Z and creates the corresponding Z eigenstate. As with measurement disturbance, the preparation disturbance is related to the distinguishability of the ideal preparation device $\mathcal{P}_Z$ and $\mathcal{P}_Z$ followed by the apparatus $\mathcal{E}$ in question and the best possible recovery operation $\mathcal{R}$.

All of the measures defined so far are "figures of merit", in the sense that we compare the actual device to the ideal, perfect functionality. In the case of state preparation we can also define a disturbance measure as a "figure of demerit", by comparing the actual functionality not to the best-case behavior but to the worst. To this end, consider a state preparation device $\mathcal{C}$ which just ignores the classical input and always prepares the same fixed output state. These are constant (output) channels, and clearly $\mathcal{E}$ disturbs the state preparation $\mathcal{P}_Z$ considerably if $\mathcal{P}_Z\mathcal{E}$ has effectively a constant output. Based on this intuition, we can then make the following formal definition:

$$\widehat{\eta}_Z(\mathcal{E}) := \tfrac{d-1}{d} - \inf_{\mathcal{C}:\text{const.}} \delta(\mathcal{C}, \mathcal{P}_Z\mathcal{E}). \tag{13}$$

The disturbance is small according to this measure if it is easy to distinguish the action of $\mathcal{P}_Z\mathcal{E}$ from having a constant output, and large otherwise. To see that $\widehat{\eta}_Z$ is positive, use the Schrödinger picture and let the output of $\mathcal{C}^*$ be the state σ for all inputs. Then note that $\inf_{\mathcal{C}} \delta(\mathcal{C}, \mathcal{P}_Z\mathcal{E}) = \min_{\mathcal{C}} \max_z \delta(\sigma, \mathcal{E}^*(\theta_z))$, where the latter δ is the trace distance. Choosing $\sigma = \frac{1}{d}\sum_z \mathcal{E}^*(\theta_z)$ and using joint convexity of the trace distance, we have $\inf_{\mathcal{C}} \delta(\mathcal{C}, \mathcal{P}_Z\mathcal{E}) \leq \frac{d-1}{d}$.

We remark that while this disturbance measure leads to finite bounds in the case of finite dimensions, it is less well behaved in the case of position and momentum measurements: Without any bound on the energy of the test states, two channels tend to be as distinguishable as possible, unless they are already constant channels. To be more precise, any non-constant channel which only changes the energy by a fixed amount can be differentiated from a constant channel by inputing states of very high energy. Roughly speaking, even an arbitrarily strongly disturbing operation can be used to gain *some* information about the input and hence a constant channel is not a good "worst case" scenario. This is in sharp contrast to the finite-dimensional case, and supports the view that the disturbance measures $\nu_Z(\mathcal{E})$ and $\eta_Z(\mathcal{E})$ are physically more sensible.

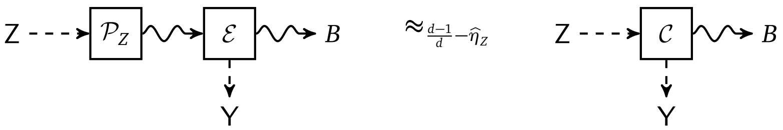


Figure 5: Figure of "demerit" version of preparation disturbance. Another approach to defining preparation disturbance is to consider distinguishability to a non-ideal device instead of an ideal device. The apparatus $\mathcal{E}$ imparts a large disturbance to the preparation $\mathcal{P}_Z$ if the output of the combination $\mathcal{P}_Z\mathcal{E}$ is essentially independent of the input. Thus we consider the distinguishability of $\mathcal{P}_Z\mathcal{E}$ and a constant preparation $\mathcal{C}$ which outputs a fixed state regardless of the input Z.

For finite-dimensional systems, all the measures of error and disturbance can be expressed as semidefinite programs, as detailed in Appendix B. As an example, we compute these measures for the simple case of a non-ideal X measurement on a qubit; we will meet this example later in assessing the tightness of the uncertainty relations and their connection to wave-particle duality relations in the Mach-Zehnder interferometer. Consider the ideal measurement isometry (6), and suppose that the basis states $|b_x\rangle$ are replaced by two pure states $|\gamma_x\rangle$ which have an overlap $\langle\gamma_0|\gamma_1\rangle = \sin\theta$. Without loss of generality, we can take $|\gamma_x\rangle = \cos\frac{\theta}{2}|b_x\rangle + \sin\frac{\theta}{2}|b_{x+1}\rangle$. The optimal measurement $\mathcal{Q}$ for distinguishing these two states is just projective measurement in the $|b_x\rangle$ basis, so let us consider the channel $\mathcal{E}_{\mathrm{MZ}} = \mathcal{W}\mathcal{Q}$. Then, as detailed in Appendix B, for Z canonically conjugate to X we find

$$\varepsilon_X(\mathcal{E}_{\mathrm{MZ}}) = \tfrac{1}{2}(1-\cos\theta) \qquad \text{and} \tag{14}$$

$$\nu_Z(\mathcal{E}_{\mathrm{MZ}}) = \eta_Z(\mathcal{E}) = \widehat{\eta}_Z(\mathcal{E}) = \tfrac{1}{2}(1-\sin\theta)\,. \tag{15}$$

In all of the figures of merit, the optimal recovery map $\mathcal{R}$ is to do nothing, while in $\widehat{\eta}_Z$ the optimal channel $\mathcal{C}$ outputs the average of the two outputs of $\mathcal{P}_Z\mathcal{E}$.

4 Uncertainty relations in finite dimensions

4.1 Complementarity measures

Before turning to the uncertainty relations, we first present several measures of complementarity that will appear therein. Indeed, we can use the above notions of disturbance to define several measures of complementarity that will later appear in our uncertainty relations. For instance, we can measure the complementarity of two observables just by using the measurement disturbance ν. Specifically, treating $\mathcal{Q}_X$ as the actual measurement and $\mathcal{Q}_Z$ as the ideal measurement, we define $c_M(X,Z) := \nu_Z(\mathcal{Q}_X)$. This quantity is equivalent to $\varepsilon_Z(\mathcal{Q}_X)$ since any recovery map $\mathcal{R}_{\mathsf{X}\to\mathsf{Z}}$ in ε_Z can be used to define $\mathcal{R}'_{\mathsf{X}\to A}$ in ν_Z by $\mathcal{R}' = \mathcal{R}\mathcal{P}_Z$. Similarly, we could treat one observable as defining the ideal state preparation device and the other as the measurement apparatus, which leads to $c_P(X,Z) := \eta_Z(\mathcal{Q}_X)$. Here we could also use the "figure of demerit" and define $\widehat{c}_P(X,Z) := \widehat{\eta}_Z(\mathcal{Q}_X)$.

Though the three complementarity measures are conceptually straightforward, it is also desireable to have closed-form expressions, particularly for the bounds in the uncertainty relations. To this end, we derive lower bounds as follows. First, consider c_M and choose as inputs Z basis states. This gives, for random choice of input,

$$\begin{align} c_M(X,Z) &\geq \inf_{\mathcal{R}} \delta(\mathcal{P}_Z\mathcal{Q}_Z, \mathcal{P}_Z\mathcal{Q}_X\mathcal{R}) \tag{16a}\\ &\geq 1 - \max_R \tfrac{1}{d}\sum_{xz} |\langle\varphi_x|\theta_z\rangle|^2 R_{zx} \tag{16b}\\ &\geq 1 - \max_R \tfrac{1}{d}\sum_x \max_z |\langle\varphi_x|\theta_z\rangle|^2 \sum_{z'} R_{z'x} \tag{16c}\\ &= 1 - \tfrac{1}{d}\sum_x \max_z |\langle\varphi_x|\theta_z\rangle|^2\,, \tag{16d} \end{align}$$

where the maximization is over stochastic matrices R, and we use the fact that $\sum_z R_{zx} = 1$ for all x. For c_P we can proceed similarly. Again replacing the recovery map $\mathcal{R}_{\mathsf{X}\to A}$ followed by $\mathcal{Q}_Z$ with a classical postprocessing map $\mathcal{R}_{\mathsf{X}\to\mathsf{Z}}$, we have

$$\begin{aligned} c_P(X,Z) &\geq \inf_{\mathcal{R}_{\mathsf{X}\to\mathsf{A}}} \delta(\mathcal{P}_Z\mathcal{Q}_Z, \mathcal{P}_Z\mathcal{Q}_X\mathcal{R}\mathcal{Q}_Z) && \text{(17a)}\\ &= \inf_{\mathcal{R}_{\mathsf{X}\to\mathsf{Z}}} \delta(\mathcal{P}_Z\mathcal{Q}_Z, \mathcal{P}_Z\mathcal{Q}_X\mathcal{R}) && \text{(17b)}\\ &\geq 1 - \tfrac{1}{d}\sum_x \max_z |\langle\varphi_x|\theta_z\rangle|^2 \,. && \text{(17c)} \end{aligned}$$

For $\widehat{c}_P(X,Z)$ we have

$$\begin{aligned} \widehat{c}_P(X,Z) &= \tfrac{d-1}{d} - \inf_{\mathcal{C}:\text{const.}} \delta(\mathcal{C}, \mathcal{P}_Z\mathcal{Q}_X) && \text{(18a)}\\ &= \tfrac{d-1}{d} - \min_P \max_z \delta(P, \mathcal{Q}_X^*(\theta_z)) && \text{(18b)}\\ &\geq \tfrac{d-1}{d} - \max_z \tfrac{1}{2}\sum_x \left|\tfrac{1}{d} - |\langle\varphi_x|\theta_z\rangle|^2\right|, && \text{(18c)} \end{aligned}$$

where the bound comes from choosing P to be the uniform distribution. We could also choose $P(x) = |\langle\varphi_x|\theta_{z'}\rangle|^2$ for some z' to obtain the bound $\widehat{c}_P(X,Z) \geq \frac{d-1}{d} - \min_{z'}\max_z \frac{1}{2}\sum_x \left|\mathrm{Tr}[\varphi_x(\theta_z - \theta_{z'})]\right|$. However, from numerical investigation of random bases, it appears that this bound is rarely better than the previous one.

Let us comment on the properties of the complementarity measures and their bounds in (16d), (17c), and (18c). Both expressions in the bounds are, properly, functions only of the two orthonormal bases involved, depending only on the set of overlaps. In particular, both are invariant under relabelling the bases. Uncertainty relations formulated in terms of conditional entropy typically only involve the largest overlap or largest two overlaps [7, 57], but the bounds derived here are yet more sensitive to the structure of the overlaps. Interestingly, the quantity in (16d) appears in the information exclusion relation of [57], where the sum of mutual informations different systems can have about the observables X and Z is bounded by $\log_2 d\sum_x \max_z |\langle\varphi_x|\theta_z\rangle|^2$.

The complementarity measures themselves all take the same value in two extreme cases: zero in the trivial case of identical bases, $(d-1)/d$ in the case that the two bases are conjugate, meaning $|\langle\varphi_x|\theta_z\rangle|^2 = 1/d$ for all x,z. In between, however, the separation between the two can be quite large. Consider two observables that share two eigenvectors while the remainder are conjugate. The bounds (16d) and (17c) imply that c_M and c_P are both greater than $(d-3)/d$. The bound on $\widehat{c}_P$ from (18c) is zero, though a better choice of constant channel can easily be found in this case. In dimensions $d = 3k+2$, fix the constant channel to output the distribution P with probability $1/3$ of being either of the last two outputs, $1/3k$ for any k of the remainder, and zero otherwise. Then we have $\hat{c}_P \geq \frac{d-1}{d} - \max_z \delta(P, \mathcal{Q}_X^*\mathcal{P}_Z^*(z))$. It is easy to show the optimal value is $2/3$ so that $\hat{c}_P \geq (d-3)/3d$. Hence, in the limit of large d, the gap between the two measures can be at least $2/3$. This example also shows that the gap between the complementary measures and the bounds can be large, though we will not investigate this further here.

4.2 Results

We finally have all the pieces necessary to formally state our uncertainty relations. The first relates measurement error and measurement disturbance, where we have

Theorem 1. *For any two observables X and Z and any quantum instrument $\mathcal{E}$,*

$$\sqrt{2\varepsilon_X(\mathcal{E})} + \nu_Z(\mathcal{E}) \geq c_M(X,Z) \quad \textit{and} \tag{19}$$

$$\varepsilon_X(\mathcal{E}) + \sqrt{2\nu_Z(\mathcal{E})} \geq c_M(Z,X)\,. \tag{20}$$

Due to Lemma 1, any joint measurement of two observables can be decomposed into a sequential measurement, which implies that these bounds hold for joint measurement devices as well. Indeed, we will make use of that lemma to derive (20) from (19) in the proof below. Of course we can replace the c_M quantities with closed-form expressions using the bound in (16d). Figure 6 shows the bound for the case of conjugate observables of a qubit, for which $c_M(X,Z) = c_M(Z,X) = \frac{1}{2}$. It also shows the particular relation between error and measurement disturbance achieved by the apparatus $\mathcal{E}_{\mathsf{MZ}}$ mentioned at the end of §3, from which we can conclude the that bound is tight in the region of vanishing error or vanishing disturbance.

For measurement error and preparation disturbance we find the following relations

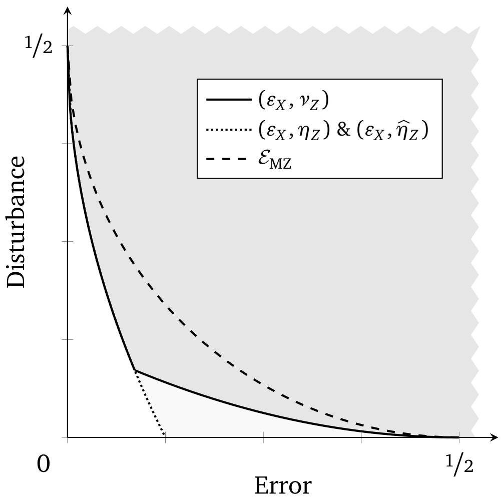


Figure 6: Error versus disturbance bounds for conjugate qubit observables. Theorem 1 restricts the possible combinations of measurement error ε_X and measurement disturbance ν_Z to the dark gray region bounded by the solid line. Theorem 2 additionally includes the light gray region. Also shown are the error and disturbance values achieved by $\mathcal{E}_{\text{MZ}}$ from §3.

Theorem 2. *For any two observables X and Z and any quantum instrument $\mathcal{E}$,*

$$\sqrt{2\varepsilon_X(\mathcal{E})} + \eta_Z(\mathcal{E}) \geq c_P(X,Z) \quad \textit{and} \tag{21}$$

$$\sqrt{2\varepsilon_X(\mathcal{E})} + \widehat{\eta}_Z(\mathcal{E}) \geq \widehat{c}_P(X,Z)\,. \tag{22}$$

Returning to Figure 6 but replacing the vertical axis with η_Z or $\widehat{\eta}_Z$, we now have only the upper branch of the bound, which continues to the horizontal axis as the dotted line. Here we can only conclude that the bounds are tight in the region of vanishing error.

4.3 Proofs

The proofs of all three uncertainty relations are just judicious applications of the triangle inequality, and the particular bound comes from the setting in which $\mathcal{P}_Z$ meets $\mathcal{Q}_X$. We shall make use of the fact that an instrument which has a small error in measuring $\mathcal{Q}_X$ is close to one which actually employs the instrument associated with $\mathcal{Q}_X$. This is encapsulated in the following

Lemma 2. *For any apparatus $\mathcal{E}_{A\to \mathsf{Y}B}$ there exists a channel $\mathcal{F}_{\mathsf{X}A\to\mathsf{Y}B}$ such that $\delta(\mathcal{E}, \mathcal{Q}'_X\mathcal{F}) \leq \sqrt{2\varepsilon_X(\mathcal{E})}$, where $\mathcal{Q}'_X$ is a quantum instrument associated with the measurement $\mathcal{Q}_X$. Furthermore, if $\mathcal{Q}_X$ is a projective measurement, then there exists a state preparation $\mathcal{P}_{\mathsf{X}\to\mathsf{Y}B}$ such that $\delta(\mathcal{E},\mathcal{Q}_X\mathcal{P}) \leq \sqrt{2\varepsilon_X(\mathcal{E})}$.*

Proof. Let $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_E \otimes L^2(\mathsf{X})$ and $W_X : \mathcal{H}_A \to L^2(\mathsf{X}) \otimes \mathcal{H}_A$ be respective dilations of $\mathcal{E}$ and $\mathcal{Q}_X$. Using the dilation W_X we can define the instrument $\mathcal{Q}'_X$ as

$$\begin{aligned} \mathcal{Q}'_X : L^\infty(\mathsf{X}) \otimes \mathcal{B}(\mathcal{H}_B) &\to \mathcal{B}(\mathcal{H}_A) \\ g\otimes A &\mapsto W_X^*(\pi(g)\otimes A)W_X\,. \end{aligned} \tag{23}$$

Suppose $\mathcal{R}_{\mathsf{Y}\to\mathsf{X}}$ is the optimal map in the definition of $\varepsilon_X(\mathcal{E})$, and let $\mathcal{R}'_{\mathsf{Y}\to\mathsf{XY}}$ be the extension of $\mathcal{R}$ which keeps the input Y; it has a dilation $V' : L^2(\mathsf{Y}) \to L^2(\mathsf{Y})\otimes L^2(\mathsf{X})$. By Stinespring continuity, in finite dimensions there exists a conditional isometry $U_X : L^2(\mathsf{X})\otimes\mathcal{H}_A \to L^2(\mathsf{X})\otimes L^2(\mathsf{Y})\otimes\mathcal{H}_B\otimes\mathcal{H}_E$ such that

$$\left\|V'V - U_X W_X\right\|_\infty \leq \sqrt{2\varepsilon_X(\mathcal{E})}\,. \tag{24}$$

Now consider the map

$$\begin{aligned} \mathcal{E}' : L^\infty(\mathsf{Y}) \otimes \mathcal{B}(\mathcal{H}_B) &\to \mathcal{B}(\mathcal{H}_A) \\ f\otimes A &\mapsto W_X^* U_X^*(\mathbb{1}_\mathsf{X}\otimes\pi(f)\otimes A\otimes \mathbb{1}_E)U_X W_X\,. \end{aligned} \tag{25}$$

By the other bound in Stinespring continuity we thus have $\delta(\mathcal{E},\mathcal{E}') \le \sqrt{2\varepsilon_X(\mathcal{E})}$. Furthermore, as described in §2.4, U_X is a conditional isometry, i.e. a collection of isometries $U_x : \mathcal{H}_A \to L^2(\mathsf{Y}) \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ for each measurement outcome x. Note that we may regard elements of $L^\infty(\mathsf{X}) \otimes \mathcal{B}(\mathcal{H})$ as sequences $(A_x)_{x\in\mathsf{X}}$ with $A_x \in \mathcal{B}(\mathcal{H})$ for all $x \in \mathsf{X}$ such that ess $\sup_x \|A_x\|_\infty < \infty$. Therefore we may define

$$\begin{aligned}\mathcal{F} : L^\infty(\mathsf{Y}) \otimes \mathcal{B}(\mathcal{H}_B) &\to L^\infty(\mathsf{X}) \otimes \mathcal{B}(\mathcal{H}_A)\\ f \otimes A &\mapsto (U_x^*(\pi(f) \otimes A \otimes \mathbb{1}_E)U_x)_{x\in\mathsf{X}}\,,\end{aligned} \tag{26}$$

so that $\mathcal{E}' = \mathcal{Q}'_X\mathcal{F}$. This completes the proof of the first statement.

If $\mathcal{Q}_X$ is a projective measurement, then the output B of $\mathcal{Q}'_X$ can just as well be prepared from the X output. Describing this with the map $\mathcal{P}'_{\mathsf{X}\to\mathsf{X}A}$ which prepares states in A given the value of X and retains X at the output, we have $\mathcal{Q}'_X = \mathcal{Q}_X\mathcal{P}'$. Setting $\mathcal{P} = \mathcal{P}'\mathcal{F}$ completes the proof of the second statement. □

Now, to prove (19), start with the triangle inequality and monotonicity. Suppose $\mathcal{P}_{\mathsf{X}\to\mathsf{Y}B}$ is the state preparation map from Lemma 2. Then, for any $\mathcal{R}_{\mathsf{Y}B\to A}$,

$$\begin{aligned}\delta(\mathcal{Q}_Z, \mathcal{Q}_X\mathcal{P}\mathcal{R}\mathcal{Q}_Z) &\le \delta(\mathcal{Q}_Z, \mathcal{E}\mathcal{R}\mathcal{Q}_Z) + \delta(\mathcal{E}\mathcal{R}\mathcal{Q}_Z, \mathcal{Q}_X\mathcal{P}\mathcal{R}\mathcal{Q}_Z) && (27a)\\ &\le \delta(\mathcal{Q}_Z, \mathcal{E}\mathcal{R}\mathcal{Q}_Z) + \delta(\mathcal{E}, \mathcal{Q}_X\mathcal{P}) && (27b)\\ &= \delta(\mathcal{Q}_Z, \mathcal{E}\mathcal{R}\mathcal{Q}_Z) + \sqrt{2\varepsilon_X(\mathcal{E})}\,. && (27c)\end{aligned}$$

Observe that $\mathcal{P}\mathcal{R}\mathcal{Q}_Z$ is just a map $\mathcal{R}'_{\mathsf{X}\to\mathsf{Z}}$. Taking the infimum over $\mathcal{R}$ we then have

$$\begin{aligned}\sqrt{2\varepsilon_X(\mathcal{E})} + \nu_Z(\mathcal{E}) &\ge \inf_{\mathcal{R}} \delta(\mathcal{Q}_Z, \mathcal{Q}_X\mathcal{P}\mathcal{R}\mathcal{Q}_Z) && (28a)\\ &\ge \inf_{\mathcal{R}} \delta(\mathcal{Q}_Z, \mathcal{Q}_X\mathcal{R})\,. && (28b)\end{aligned}$$

To show (20), let $\mathcal{R}_{\mathsf{Y}B\to A}$ and $\mathcal{R}'_{\mathsf{Y}\to\mathsf{X}}$ be the optimal maps in $\nu_Z(\mathcal{E})$ and $\varepsilon_X(\mathcal{E})$, respectively. Now apply Lemma 1 to $\mathcal{M} = \mathcal{E}\mathcal{R}'\mathcal{R}\mathcal{Q}_Z$ and suppose that $\mathcal{E}'_{A\to\mathsf{Z}B}$ is the resulting instrument and $\mathcal{M}_{\mathsf{Z}B\to\mathsf{X}}$ is the conditional measurement. By the above argument, $\sqrt{2\varepsilon_Z(\mathcal{E}')} + \nu_X(\mathcal{E}') \ge \inf_{\mathcal{R}} \delta(\mathcal{Q}_X, \mathcal{Q}_Z\mathcal{R})$. But $\varepsilon_Z(\mathcal{E}') \le \delta(\mathcal{Q}_Z, \mathcal{E}'\mathcal{T}_B) = \nu_Z(\mathcal{E})$ and $\nu_X(\mathcal{E}') \le \delta(\mathcal{Q}_X, \mathcal{E}'\mathcal{M}) = \varepsilon_X(\mathcal{E})$, where in the latter we use the fact that we could always reprepare an X eigenstate and then let $\mathcal{Q}_X$ measure it. Therefore the desired bound holds.

To establish (21), we proceed just as above to obtain

$$\delta(\mathcal{P}_Z, \mathcal{P}_Z\mathcal{Q}_X\mathcal{P}\mathcal{R}) \le \delta(\mathcal{P}_Z, \mathcal{P}_Z\mathcal{E}\mathcal{R}) + \sqrt{2\varepsilon_X(\mathcal{E})}\,. \tag{29}$$

Now $\mathcal{P}_{\mathsf{X}\to\mathsf{Y}B}\mathcal{R}_{\mathsf{Y}B\to A}$ is a preparation map $\mathcal{P}_{\mathsf{X}\to A}$, and taking the infimum over $\mathcal{R}$ gives

$$\begin{aligned}\sqrt{2\varepsilon_X(\mathcal{E})} + \eta_Z(\mathcal{E}) &\ge \inf_{\mathcal{R}} \delta(\mathcal{P}_Z, \mathcal{P}_Z\mathcal{Q}_X\mathcal{P}\mathcal{R}) && (30a)\\ &\ge \inf_{\mathcal{P}} \delta(\mathcal{P}_Z, \mathcal{P}_Z\mathcal{Q}_X\mathcal{P})\,. && (30b)\end{aligned}$$

Finally, (22). Since the $\widehat{\eta}_Z$ disturbance measure is defined “backwards”, we start the triangle inequality with the distinguishability quantity related to disturbance, rather than the eventual constant of the bound. For any channel $\mathcal{C}_{\mathsf{Z}\to\mathsf{X}}$ and $\mathcal{P}_{\mathsf{X}\to\mathsf{Y}B}$ from Lemma 2, just as before we have

$$\begin{aligned}\delta(\mathcal{C}\mathcal{P}, \mathcal{P}_Z\mathcal{E}) &\le \delta(\mathcal{C}\mathcal{P}, \mathcal{P}_Z\mathcal{Q}_X\mathcal{P}) + \delta(\mathcal{P}_Z\mathcal{Q}_X\mathcal{P}, \mathcal{P}_Z\mathcal{E}) && (31a)\\ &\le \delta(\mathcal{C}, \mathcal{P}_Z\mathcal{Q}_X) + \sqrt{2\varepsilon_X(\mathcal{E})}\,. && (31b)\end{aligned}$$

Now we take the infimum over constant channels $\mathcal{C}_{\mathsf{Z}\to\mathsf{X}}$. Note that

$$\inf_{\mathcal{C}_{\mathsf{Z}\to\mathsf{Y}B}} \delta(\mathcal{C}, \mathcal{P}_Z\mathcal{E}) \le \inf_{\mathcal{C}_{\mathsf{Z}\to\mathsf{X}}} \delta(\mathcal{C}\mathcal{P}, \mathcal{P}_Z\mathcal{E})\,. \tag{32}$$

Therefore, we have

$$\sqrt{2\varepsilon_X(\mathcal{E})} + \widehat{\eta}_Z(\mathcal{E}) \ge \tfrac{d-1}{d} - \inf_{\mathcal{C}} \delta(\mathcal{C}, \mathcal{P}_Z\mathcal{Q}_X)\,. \tag{33}$$

This last proof also applies to a more general definition of disturbance which does not use $\mathcal{P}_Z$ at the input, but rather diagonalizes or “pinches” any input quantum system in the Z basis. Such a transformation can be thought of as the result of performing an ideal Z measurement, but forgetting the result. More formally, letting $\mathcal{Q}_Z^\natural = \mathcal{W}_Z \mathcal{T}_Z$ with $\mathcal{W}_Z : a \to W_Z^* a W_Z$, we can define

$$\widetilde{\eta}_Z(\mathcal{E}) = \tfrac{d-1}{d} - \inf_{\mathcal{C}} \delta(\mathcal{C}, \mathcal{Q}_Z^\natural \mathcal{E})\,. \tag{34}$$

Though perhaps less conceptually appealing, this is a more general notion of disturbance, since now we can potentially use entanglement at the input to increase distinguishability of $\mathcal{Q}_Z^\natural \mathcal{E}$ from any constant channel. However, due to the form of $\mathcal{Q}_Z^\natural$, entanglement will not help. Applied to any bipartite state, the map $\mathcal{Q}_Z^\natural$ produces a state of the form $\sum_z p_z |\theta_z\rangle\langle\theta_z| \otimes \sigma_z$ for some probability distribution p_z and set of normalized states σ_z, and therefore the input to $\mathcal{E}$ itself is again an output of $\mathcal{P}_Z$. Since classical correlation with ancillary systems is already covered in $\widehat{\eta}_Z(\mathcal{E})$, it follows that $\widetilde{\eta}_Z(\mathcal{E}) = \widehat{\eta}_Z(\mathcal{E})$.

5 Position & momentum

5.1 Gaussian precision-limited measurement and preparation

Now we turn to the infinite-dimensional case of position and momentum measurements. Let us focus on Gaussian limits on precision, where the convolution function α described in §2.2 is the square root of a normalized Gaussian of width σ, and for convenience define

$$g_\sigma(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}\,. \tag{35}$$

One advantage of the Gaussian choice is that the Stinespring dilation of the ideal σ-limited measurement device is just a canonical transformation. Thus, measurement of position Q just amounts to adding this value to an ancillary system which is prepared in a zero-mean Gaussian state with position standard deviation σ_Q, and similarly for momentum. The same interpretation is available for precision-limited state preparation. To prepare a momentum state of width σ_P, we begin with a system in a zero-mean Gaussian state with momentum standard deviation σ_P and simply shift the momentum by the desired amount.

Given the ideal devices, the definitions of error and disturbance are those of §3, as in the finite-dimensional case, with the slight change that the first term of $\widehat{\eta}$ is now 1. To reduce clutter, we do not indicate σ_Q and σ_P specifically in the error and disturbance functions themselves.

Since our error and disturbance measures are based on possible state preparations and measurements in order to best distinguish the two devices, in principle one ought to consider precision limits in the distinguishability quantity δ. However, we will not follow this approach here, and instead we allow test of arbitrary precision in order to preserve the link between distinguishability and the cb norm. This leads to bounds that are perhaps overly pessimistic, but nevertheless limit the possible performance of any device.

5.2 Results

As discussed previously, the disturbance measure of demerit $\widehat{\eta}$ cannot be expected to lead to uncertainty relations for position and momentum observables, as any non-constant channel can be perfectly differentiated from a constant one by inputting states of arbitrarily high momentum. We thus focus on the disturbance measures of merit.

Theorem 3. *Set $c = 2\sigma_Q\sigma_P$ for any precision values $\sigma_Q, \sigma_P > 0$. Then for any quantum instrument $\mathcal{E}$,*

$$\left.\begin{array}{r} \sqrt{2\varepsilon_Q(\mathcal{E})} + \nu_P(\mathcal{E}) \\ \varepsilon_Q(\mathcal{E}) + \sqrt{2\nu_Q(\mathcal{E})} \end{array}\right\} \geq \frac{1-c^2}{(1+c^{2/3}+c^{4/3})^{3/2}} \quad \textit{and} \tag{36}$$

$$\sqrt{2\varepsilon_Q(\mathcal{E})} + \eta_P(\mathcal{E}) \geq \frac{(1+c^2)^{1/2}}{((1+c^2)+c^{2/3}(1+c^2)^{2/3}+c^{4/3}(1+c^2)^{1/3})^{3/2}}\,. \tag{37}$$

Before proceeding to the proofs, let us comment on the properties of the two bounds. As can be seen in Figure 7, the bounds take essentially the same values for $\sigma_Q\sigma_P \ll \frac{1}{2}$, and indeed both evaluate to unity at $\sigma_Q\sigma_P = 0$. This is the region of combined position and momentum precision far smaller than the natural scale set by $\hbar$, and the limit of infinite precision accords with the finite-dimensional bounds for conjugate observables in the limit $d \to \infty$. Otherwise, though, the bounds differ remarkably. The measurement disturbance bound in (36) is positive only when $\sigma_Q\sigma_P \leq \frac{1}{2}$, which is the Heisenberg precision limit. In contrast, the preparation disturbance bound in (37) is always positive, though it decays roughly as $(\sigma_Q\sigma_P)^2$.

The distinction between these two cases is a result of allowing arbitrarily precise measurements in the distinguishability measure. It can be understood by the following heuristic argument. Consider an experiment in which a momentum state of width σ_P^{in} is subjected to a position measurement of resolution σ_Q and then a momentum measurement of resolution σ_P^{out}. From the uncertainty principle, we expect the position measurement to change the momentum by an amount $\sim 1/\sigma_Q$. Thus, to reliably detect the change in momentum, σ_P^{out} must fulfill the condition $\sigma_P^{\text{out}} \ll \sigma_P^{\text{in}} + 1/\sigma_Q$. The Heisenberg limit in the measurement disturbance scenario is $\sigma_P^{\text{out}} = 2/\sigma_Q$, meaning this condition cannot be met no matter how small we choose σ_P^{in}. This is consistent with no nontrivial bound in (36) in this region. On the other hand, for preparation disturbance the Heisenberg limit is $\sigma_P^{\text{in}} = 2/\sigma_Q$, so detecting the change in momentum simply requires $\sigma_P^{\text{out}} \ll 1/\sigma_Q$. A more satisfying approach would be to include the precision limitation in the distinguishability measure to restore the symmetry of the two scenarios, but this requires significant changes to the proof and is left for future work.

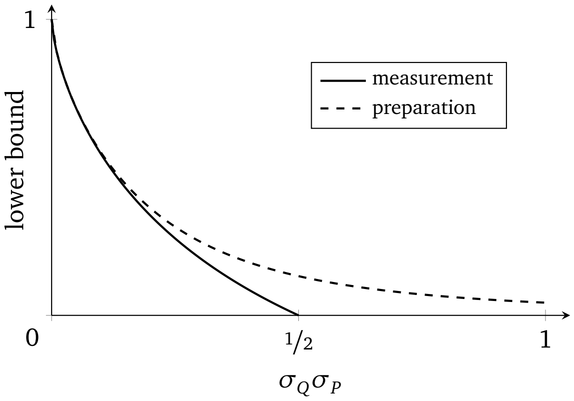


Figure 7: Uncertainty bounds appearing in Theorem 3 in terms of the combined precision $\sigma_Q\sigma_P$. The solid line corresponds to the bound involving measurement disturbance, (36), the dashed line to the bound involving preparation disturbance, (37).

5.3 Proofs

The proof of Theorem 3 is broadly similar to the finite-dimensional case. We would again like to begin with $\mathcal{F}_{QA\to YB}$ from Lemma 2 such that $\delta(\mathcal{E}, \mathcal{Q}'_Q\mathcal{F}) \leq \sqrt{2\varepsilon_Q(\mathcal{E})}$. However, the argument does not quite go through, as in infinite dimensions we cannot immediately ensure that the infimum in Stinespring continuity is attained. Nonetheless, we can consider a sequence of maps $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that the desired distinguishability bound holds in the limit $n \to \infty$.

To show (36), we follow the steps in (27). Now, though, consider the map $\mathcal{F}'_n$ which just appends Q to the output of $\mathcal{F}_n$, and define $\mathcal{N} = \mathcal{Q}'_Q\mathcal{F}_n\mathcal{R}\mathcal{Q}_P$, where $\mathcal{Q}'_Q$ is the instrument associated with position measurement $\mathcal{Q}_Q$. Then we have

$$\delta(\mathcal{Q}_P, \mathcal{N}\mathcal{T}_{\mathrm{Q}}) \leq \delta(\mathcal{Q}_P, \mathcal{E}\mathcal{R}\mathcal{Q}_P) + \delta(\mathcal{E}\mathcal{R}\mathcal{Q}_P, \mathcal{N}\mathcal{T}_{\mathrm{Q}}) \tag{38a}$$

$$\leq \delta(\mathcal{Q}_P, \mathcal{E}\mathcal{R}\mathcal{Q}_P) + \delta(\mathcal{E}, \mathcal{Q}'_Q\mathcal{F}_n). \tag{38b}$$

Taking the limit $n \to \infty$ and the infimum over recovery maps $\mathcal{R}$ produces $\sqrt{2\varepsilon_Q(\mathcal{E})} + \nu_P(\mathcal{E})$ on the righthand side. We can bound the lefthand side by testing with pure unentangled inputs:

$$\delta(\mathcal{Q}_P, \mathcal{N}\mathcal{T}_{\mathrm{Q}}) \geq \sup_{\psi, f} \langle \psi, \big(\mathcal{Q}_P(f) - [\mathcal{N}\mathcal{T}_{\mathrm{Q}}](f)\big)\psi\rangle\,. \tag{39}$$

Now we want to show that, since $\mathcal{Q}_P$ is covariant with respect to phase space translations, without loss of generality we can take $\mathcal{N}$ to be covariant as well. Consider the translated version of both $\mathcal{Q}_P$ and $\mathcal{N}\mathcal{T}_{\mathrm{Q}}$, obtained by shifting their inputs and outputs correspondingly by some amount $z = (q, p)$. For the states ψ this shift is implemented by the Weyl-Heisenberg operators V_z, while for tests f only the value of p is relevant. Any such shift does not change the distinguishability, because we can always shift ψ and f as well to recover the original quantity. Averaging over the translated versions therefore also leads to the same distinguishability, and since $\mathcal{Q}_P$ is itself covariant, the averaging results in a covariant $\mathcal{N}\mathcal{T}_{\mathrm{Q}}$. The details of the averaging require some care in this noncompact setting, but are standard by now, and we refer the reader to the work of Werner [22] for furter details. Since $\mathcal{T}_{\rm Q}$ just ignores the Q output of the measurement $\mathcal{N}$, we may thus proceed by assuming that $\mathcal{N}$ is a covariant measurement.

Any covariant $\mathcal{N}$ has the form

$$\mathcal{N}(f)=\int_{\mathbb{R}^2}\frac{\mathrm{d}z}{2\pi}\,f(z)V_z m V_z^*, \tag{40}$$

for some positive operator m such that $\mathrm{Tr}[m]=1$. Due to the definition of $\mathcal{N}$, the position measurement result is precisely that obtained from $\mathcal{Q}_Q$. By the covariant form of $\mathcal{N}$, this implies that the position width of m is just σ_Q (or rather that of the parity version of m, see [22]). Suppose the momentum distribution has standard deviation $\widehat{\sigma}_P$; then $\sigma_Q\widehat{\sigma}_P \geq 1/2$ follows from the Kennard uncertainty relation [3].

Now we can evaluate the lower bound term by term. Let us choose a Gaussian state in the momentum representation and test function: $\psi = g^{\frac{1}{2}}_{\sigma_\psi}$ and $f=\sqrt{2\pi}\sigma_f g_{\sigma_f}$. Then the first term is a straightforward Gaussian integral, since the precision-limited measurement just amounts to the ideal measurement convolved with g_{σ_P}:

$$\begin{aligned}\langle\psi,\mathcal{Q}_P(f)\psi\rangle &= \int_{\mathbb{R}^2} dp'dp\; g_{\sigma_\psi}(p')g_{\sigma_P}(p'-p)f(p) &(41a)\\ &= \frac{\sigma_f}{\sqrt{\sigma_f^2+\sigma_P^2+\sigma_\psi^2}}. &(41b)\end{aligned}$$

The second term is the same, just with $\widehat{\sigma}_P$ instead of σ_P, so we have

$$\delta(\mathcal{Q}_P,\mathcal{N}\mathcal{T}_{\rm Q}) \geq \frac{\sigma_f}{\sqrt{\sigma_f^2+\sigma_P^2+\sigma_\psi^2}} - \frac{\sigma_f}{\sqrt{\sigma_f^2+\widehat{\sigma}_P^2+\sigma_\psi^2}}. \tag{42}$$

The tightest possible bound comes from the smallest $\widehat{\sigma}_P$, which is $1/2\sigma_Q$, and the bound is clearly trivial if $\sigma_Q\sigma_P \geq 1/2$. If this is not the case, we can optimize our choice of σ_f. To simplify the calculation, assume that σ_ψ is small compared to σ_f (so that we are testing with a very narrow momentum state). Then, with $c=2\sigma_Q\sigma_P$, the optimal σ_f is given by

$$\sigma_f^2=\frac{\sigma_P^2}{c^{2/3}(1+c^{2/3})}. \tag{43}$$

Using this in (42) gives (36).

For preparation disturbance, proceed as before to obtain

$$\begin{aligned}\delta(\mathcal{P}_P,\mathcal{P}_P\mathcal{Q}'_{\rm Q}\mathcal{F}'_n\mathcal{R}\mathcal{T}_{\rm Q}) &\leq \delta(\mathcal{P}_P,\mathcal{P}_P\mathcal{E}\mathcal{R})+\delta(\mathcal{P}_P\mathcal{E}\mathcal{R},\mathcal{P}_P\mathcal{Q}'_{\rm Q}\mathcal{F}'_n\mathcal{R}\mathcal{T}_{\rm Q}) &(44a)\\ &\leq \delta(\mathcal{P}_P,\mathcal{P}_P\mathcal{E}\mathcal{R})+\delta(\mathcal{E},\mathcal{Q}'_{\rm Q}\mathcal{F}_n) &(44b)\end{aligned}$$

Now the limit $n\to\infty$ and the infimum over recovery maps $\mathcal{R}$ produces $\sqrt{2\varepsilon_Q(\mathcal{E})}+\eta_P(\mathcal{E})$ on the righthand side. A lower bound on the quantity on the lefthand side can be obtained by using $\mathcal{P}_P$ to prepare a σ_P-limited input state and making a σ_m-limited momentum measurement $\bar{\mathcal{Q}}_P$ measurement on the output, so that, for $\mathcal{N}$ as before,

$$\delta(\mathcal{P}_P,\mathcal{P}_P\mathcal{Q}'_{\rm Q}\mathcal{F}'_n\mathcal{R}\mathcal{T}_{\rm Q}) \geq \sup_{\psi:\text{Gaussian};f}\,\langle\psi,\big(\bar{\mathcal{Q}}_P(f)-[\mathcal{N}\mathcal{T}_{\rm Q}](f)\big)\psi\rangle\,. \tag{45}$$

The only difference to (39) is that the supremum is restricted to Gaussian states of width σ_P. The covariance argument nonetheless goes through as before, and we can proceed to evaluate the lower bound as above. This yields

$$\delta(\mathcal{P}_P,\mathcal{P}_P\mathcal{Q}'_{\rm Q}\mathcal{F}'_n\mathcal{R}\mathcal{T}_{\rm Q}) \geq \frac{\sigma_f}{\sqrt{\sigma_f^2+\sigma_m^2+\sigma_P^2}} - \frac{\sigma_f}{\sqrt{\sigma_f^2+\frac{1}{4\sigma_Q^2}+\sigma_P^2}}\,. \tag{46}$$

We may as well consider $\sigma_m\to 0$ so as to increase the first term. The optimal σ_f is then given by the optimizer above, replacing c with $c/\sqrt{1+c^2}$. Making the same replacement in (36) yields (37).

6 Applications

6.1 No information about Z without disturbance to X

A useful tool in the construction of quantum information processing protocols is the link between reliable transmission of X eigenstates through a channel $\mathcal{N}$ and Z eigenstates through its complement $\mathcal{N}^\sharp$, particularly when the observables X and Z are maximally complementary, i.e. $|\langle\varphi_x|\vartheta_z\rangle|^2 = \frac{1}{d}$ for all x, z. Due to the uncertainty principle, we expect that a channel cannot reliably transmit the bases to different outputs, since this would provide a means to simultaneously measure X and Z. This link has been used by Shor and Preskill to prove the security of quantum key distribution [58] and by Devetak to determine the quantum channel capacity [59]. Entropic state-preparation uncertainty relations from [6, 44] can be used to understand both results, as shown in [60, 61].

However, the above approach has the serious drawback that it can only be used in cases where the specific X-basis transmission over $\mathcal{N}$ and Z-basis transmission over $\mathcal{N}^\sharp$ are in some sense compatible and not *counterfactual*; because the argument relies on a state-dependent uncertainty principle, both scenarios must be compatible with the same quantum state. Fortunately, this can be done for both QKD security and quantum capacity, because at issue is whether X-basis (Z-basis) transmission is reliable (unreliable) on average *when the states are selected uniformly at random*. Choosing among either basis states at random is compatible with a random measurement in either basis of half of a maximally entangled state, and so both X and Z basis scenarios are indeed compatible. The same restriction to choosing input states uniformly appears in the recent result of [33], as it also ultimately relies on a state-preparation uncertainty relation.

Using Theorem 2 we can extend the method above to counterfactual uses of arbitrary channels $\mathcal{N}$, in the following sense: If acting with the channel $\mathcal{N}$ does not substantially affect the possibility of performing an X measurement, then Z-basis inputs to $\mathcal{N}^\sharp$ result in an essentially constant output. More concretely, we have

Corollary 1. *Given a channel $\mathcal{N}$ and complementary channel $\mathcal{N}^\sharp$, suppose that there exists a measurement Λ_X such that $\delta(\mathcal{Q}_X, \mathcal{N}\Lambda_X) \leq \varepsilon$. Then there exists a constant channel $\mathcal{C}$ such that*

$$\delta(\mathcal{Q}_Z^\natural \mathcal{N}^\sharp, \mathcal{C}) \leq \sqrt{2\varepsilon} + \tfrac{d-1}{d} - \widehat{c}_P(X,Z). \tag{47}$$

For maximally complementary X and Z, $\delta(\mathcal{Q}_Z^\natural \mathcal{N}^\sharp, \mathcal{C}) \leq \sqrt{2\varepsilon}$.

Proof. Let V be the Stinespring dilation of $\mathcal{N}$ such that $\mathcal{N}^\sharp$ is the complementary channel and define $\mathcal{E} = \mathcal{V}_{\mathcal{N}}\Lambda_X$. For $\mathcal{C}$ the optimal choice in the definition of $\widehat{\eta}_Z(\mathcal{E})$, (22), (34), and $\widetilde{\eta}_Z = \widehat{\eta}_Z$ imply $\delta(\mathcal{Q}_Z^\natural \mathcal{E}, \mathcal{C}) \leq \sqrt{2\varepsilon} + \frac{d-1}{d} - \widehat{c}_P(X,Z)$. Since $\mathcal{N}^\sharp$ is obtained from $\mathcal{E}$ by ignoring the Λ_X measurement result, $\delta(\mathcal{Q}_Z^\natural \mathcal{N}^\sharp, \mathcal{C}) \leq \delta(\mathcal{Q}_Z^\natural \mathcal{E}, \mathcal{C})$. □

This formulation is important because in more general cryptographic and communication scenarios we are interested in the worst-case behavior of the protocol, not the average case under some particular probability distribution. For instance, in [46] the goal is to construct a classical computer resilient to leakage of Z-basis information by establishing that reliable X basis measurement is possible despite the interference of the eavesdropper. However, such an X measurement is entirely counterfactual and cannot be reconciled with the actual Z-basis usage, as the Z-basis states will be chosen *deterministically* in the classical computer.

It is important to point out that, unfortunately, calibration testing is in general completely insufficient to establish a small value of $\delta(\mathcal{Q}_X, \mathcal{N}\Lambda_X)$. More specifically, the following example shows that there is no dimension-independent bound connecting $\inf_{\Lambda_X} \delta(\mathcal{Q}_X, \mathcal{N}\Lambda_X)$ to the worst case probability of incorrectly identifying an X eigenstate input to $\mathcal{N}$, for arbitrary $\mathcal{N}$. Let the quantities p_{yz} be given by $p_{y,0} = 2/d$ for $y = 0, \dots, d/2-1$, $p_{y,1} = 2/d$ for $y = d/2, \dots, d-1$, and $p_{y,z} = 1/d$ otherwise, where we assume d is even, and then define the isometry $V : \mathcal{H}_A \to \mathcal{H}_B \otimes \mathcal{H}_C \otimes \mathcal{H}_D$ as the map taking $|z\rangle_A$ to $\sum_y \sqrt{p_{yz}}\, |y\rangle_B |z\rangle_C |y\rangle_D$. Finally, let $\mathcal{N} : \mathcal{B}(\mathcal{H}_B) \otimes \mathcal{B}(\mathcal{H}_C) \to \mathcal{B}(\mathcal{H}_A)$ be the channel obtained by ignoring D, i.e. in the Schrödinger picture $\mathcal{N}^*(\varrho) = \mathrm{Tr}_D[V\varrho V^*]$. Now consider inputs in the X basis, with X canonically conjugate to Z. As shown in Appendix C, the probability of correctly determining any particular X input is the same for all values, and is equal to $\frac{1}{d^2}\sum_y \left(\sum_z \sqrt{p_{y,z}}\right)^2 = (d + \sqrt{2} - 2)^2/d^2$. The worst case X error probability therefore tends to zero like $1/d$ as $d \to \infty$. On the other hand, Z-basis inputs 0 and 1 to the complementary channel $\mathcal{E}^\sharp$ result in completely disjoint output states due to the form of p_{yz}. Thus, if we consider a test which inputs one of these randomly and checks for agreement at the output, we find $\inf_{\mathcal{C}} \delta(\mathcal{Q}_Z^\natural \mathcal{N}^\sharp, \mathcal{C}) \geq \frac{1}{2}$. Using the bound above, this implies $\inf_{\Lambda_X} \delta(\mathcal{Q}_X, \mathcal{N}\Lambda_X) \geq \frac{1}{8}$. This is not 1, but the point is it is bounded away from zero and independent of d: There must be a factor of d when converting between the worst case error probability and the distinguishability.

We can appreciate the failure of calibration in this example from a different point of view, by appealing to the information-disturbance tradeoff of [40]. Since $\mathcal{N}$ transmits Z eigenstates perfectly to BC and X eigenstates almost perfectly, we might be tempted to conclude that the channel is close to the identity channel. However, the information-disturbance tradeoff implies that complements of channels close to the identity are close to constant channels. Clearly this is not the case here, since $\mathcal{N}^*(|0\rangle\langle 0|)$ is distinguishable from $\mathcal{N}^*(|1\rangle\langle 1|)$. This point is discussed further by one of us in [62]. The counterexample constructed above it not symmetric for Z inputs, and it is an open question if calibration is sufficient in the symmetric case. For channels that are covariant with respect to the Weyl-Heisenberg group (also known as the generalized Pauli group), it is not hard to show that calibration *is* in fact sufficient.

6.2 Connection to wave-particle duality relations

In [42] Englert presents a wave-particle complementarity relation in a Mach-Zehnder interferometer, quantifying the extent to which “the observations of an interference pattern and the acquisition of which-way information are mutually exclusive”. The particle-like “which-way” information is obtained by additional detectors in the arms of the interferometer, while fringe visibility is measured by the population difference between the two output ports of the interferometer. The detectors can be thought of as producing different states in an ancilla system, depending on the path taken by the light. Englert shows the following tradeoff between the visibility V and distinguishability D of the which-way detector states:

$$V^2 + D^2 \leq 1. \tag{48}$$

We may regard the entire interferometer plus which-way detector as an apparatus $\mathcal{E}_{\text{MZ}}$ with quantum and classical output. It turns out that $\mathcal{E}_{\text{MZ}}$ is precisely the nonideal qubit X measurement considered in §3 and that path distinguishability is related error of X and visibility to disturbance (all of which are equal in this case by (15)) of a conjugate observable Z. More specifically, as shown in Appendix D,

$$\varepsilon_X(\mathcal{E}_{\text{MZ}}) = \tfrac{1}{2}(1-D) \quad \text{and} \quad \nu_Z(\mathcal{E}_{\text{MZ}}) = \eta_Z(\mathcal{E}_{\text{MZ}}) = \widehat{\eta}_Z(\mathcal{E}_{\text{MZ}}) = \tfrac{1}{2}(1-V). \tag{49}$$

Therefore, (48) is also an error-preparation disturbance relation. By the same token, the uncertainty relations in Theorems 1 and 2 imply wave-particle duality relations.

Let us comment on other connections between uncertainty and duality relations. Recently, [63] showed a relation between wave-particle duality relations and entropic uncertainty relations. As discussed above, the latter are state-dependent state-preparation relations, and so the interpretation of the wave-particle duality relation is somewhat different. Here we have shown that Englert's relation can actually be understood as a state-independent relation.

Each of the disturbance measures are related to visibility in Englert's setup. It is an interesting question to consider a multipath interferometer to settle the question of which disturbance measure should be associated to visibility in general. From the discussion of [64], it would appear that visibility ought to be related to measurement disturbance ν_Z, but we leave a complete analysis to future work.

7 Comparison to previous work

Broadly speaking, there are two main kinds of uncertainty relations: those which are constraints on fixed experiments, including the details of the input quantum state, and those that are constraints on quantum devices themselves, independent of the particular input. All of our relations are of the latter type, in contrast to entropic relations, which are typically of the former type. At a formal level, this distinction appears in whether or not the quantities involved in the precise relation depend on the input state or not.[2] Each type of relation certainly has its use, though when considering error-disturbance uncertainty relations, we argued in the introduction that the conceptual underpinnings of state-dependent relations describing fixed experiments are unclear. Indeed, it is precisely because of the uncertainty principle that trouble arises in defining error and disturbance in this case. Worse still, there can be no nontrivial bound relating error and disturbance which applies universally, i.e. to all states [65].

Independent of the previous question, another major contrast between different kinds of uncertainty relations is whether they depend on the values taken by the observables, or only the configuration of their eigenstates. Again, our relations are all of the latter type, but now we share this property with entropic relations. That is not to say that the observable values are completely irrelevant in our setting, merely that they

[2] This is separate from the issue of whether the bound depends on the state, as for instance in the Robertson relation [4].

are not necessarily relevant. In distinguishing the outputs of an ideal position measurement of given precision from the outputs of the actual device, one may indeed make use of the difference in measurement values. But this need not be the only kind of comparison.

In the recent work of Busch, Lahti, and Werner [25], the authors used the Wasserstein metric of order two, corresponding to the mean squared error, as the underlying distance $D(.,.)$ to measure the closeness of probability distributions. If $\mathcal{M}^Q$, $\mathcal{M}^P$ are the marginals of some joint measurement of position Q and momentum P, and X_ϱ denotes the distribution coming from applying the measurement X to the state ϱ, their relation reads

$$\sup_\varrho D(\mathcal{M}^Q_\varrho, Q_\varrho) \cdot \sup_\varrho D(\mathcal{M}^P_\varrho, P_\varrho) \geq c\,, \tag{50}$$

for some universal constant c. In [27], the authors generalize their results to arbitrary Wasserstein metrics. As in our case, the two distinguishability quantities in (50) are separately maximized over all states, and hence the resulting expression characterizes the goodness of the approximate measurement.

One could instead ask for a "coupled optimization", a relation of the form

$$\sup_\varrho \Big[D(\mathcal{M}^Q_\varrho, Q_\varrho) D(\mathcal{M}^P_\varrho, P_\varrho) \Big] \geq c', \tag{51}$$

for some other constant c'.[3] This approach is taken in [66] for the question of joint measurability. While this statement certainly tells us that no device can accurately measure both position and momentum for all input states, the bound c' only holds (and can only hold) for the worst possible input state. In contrast, our bounds, as well as in (50) are state-independent in the sense that the bound holds for all states. Indeed, the two approaches are more distinct than the similarities between (50) and (51) would suggest. By optimizing over input states separately, our results and those of [22, 25, 27] are statements about the properties of measurement devices themselves, independent of any particular experimental setup. State-dependent settings capture the behavior of measurement devices in specific experimental setups and must therefore account for the details of the input state.

The same set of authors also studied the case of finite-dimensional systems, in particular qubit systems, again using the Wasserstein metric of order two [26]. Their results for this case are similar, with the product in (50) replaced by a sum. Perhaps most closely related to our results is the recent work by Ipsen [34], who uses the variational distance as the underlying distinguishability measure to derive similar additive uncertainty relations. We note, however, that both [26] and [34] only consider joint measurability and do not consider the change to the state after the approximate measurement is performed, as it is done in our error-disturbance relation. Furthermore, both base their distinguishability measures on the measurement statistics of the devices alone. But this does not necessarily tell us how distinguishable two devices ultimately are, as we could employ input states entangled with ancilla systems to test them. These two measures can be different [51], even for entanglement-breaking channels [67]. In Appendix A we give an example which shows that this is also true of quantum measurements, a specific kind of entanglement-breaking channel.

Entropic quantities are another means of comparing two probability distributions, an approach taken recently by Buscemi *et al.* [33] and Coles and Furrer [35] (see also Martens and de Muynck [29]). Both contributions formalize error and disturbance in terms of relative or conditional entropies, and derive their results from entropic uncertainty relations for state preparation which incorporate the effects of quantum entanglement [6, 44]. They differ in the choice of the entropic measure and the choice of the state on which the entropic terms are evaluated. Buscemi *et al.* find state-independent error-disturbance relations involving the von Neumann entropy, evaluated for input states which describe observable eigenstates chosen uniformly at random. As described in Sec. 6, the restriction to uniformly-random inputs is significant, and leads to a characterization of the average-case behavior of the device (averaged over the choice of input state), not the worst-case behavior as presented here. Meanwhile, Coles and Furrer make use of general Rényi-type entropies, hence also capturing the worst-case behavior. However, they are after a state-dependent error-disturbance relation which relates the amount of information a measurement device can extract from a state about the results of a *future* measurement of one observable to the amount of disturbance caused to other observable.

An important distinction between both these results and those presented here is the quantity appearing in the uncertainty bound, i.e. the quantification of complementarity of two observables. As both the aforementioned results are based on entropic state-preparation uncertainty relations, they each quantify complementarity by the largest overlap of the eigenstates of the two observables. This bound is trivial should the two

[3] Such an approach has been advocated by David Reeb (private communication).

observables share an eigenstate. However, a perfect joint measurement is clearly impossible even if the observables share all but two eigenvectors (if they share all but one, they necessarily share all eigenvectors). All three complementarity measures used here are nontrivial whenever not all eigenvectors are shared between the observables.

8 Conclusions

We have formulated simple, operational definitions of error and disturbance based on the probability of distinguishing the actual measurement apparatus from the relevant ideal apparatus by any testing procedure whatsoever. The resulting quantities are conceptually straightfoward properties of the measurement apparatus, not any particular fixed experimental setup. We presented uncertainty relations for both joint measurability and the error-disturbance tradeoff, for both arbitrary finite-dimensional systems and for position and momentum. In the former case the bounds involve simple measures of the complementarity of two observables, while the latter involve the ratio of the desired position and momentum precisions σ_Q and σ_P to Planck's constant $\hbar$. We further showed that this operational approach has applications to quantum information processing and to wave-particle duality relations. Finally, we presented a detailed comparison of the relation of our results to previous work on uncertainty relations.

Several interesting questions remain open. One may inquire about the tightness of the bounds. The qubit example for conjugate observables discussed at the end of §3 shows that the finite-dimensional bounds of Theorem 2 are tight for small error ε_X, though no conclusion can be drawn from this example for small preparation disturbance. It would be interesting to check the tightness of the position and momentum bounds by computing the error and disturbance measures for a device described by a covariant measurement. For reasons of simplicity, we have not attempted to incorporate precision limits into the definitions of error and distinguishability of position and momentum. Doing so would lead to more conceptually satisfying bounds and perhaps remedy the fact that the measurement error-preparation disturbance bound is nontrivial even outside the Heisenberg limit. Bounds for other observables in infinite dimensions would also be quite interesting, for instance the mixed discrete/continuous case of energy and position of a harmonic oscillator. Restricting to covariant measurements, in finite or infinite dimensions, it would also be interesting to determine if entangled inputs improve the distinguishability measures, or whether calibration testing is sufficient. From the application in Corollary 1, it would appear that calibration is sufficient, but we have not settled the matter conclusively.

Acknowledgements: The authors are grateful to David Sutter, Paul Busch, Omar Fawzi, Fabian Furrer, Michael Walter and especially David Reeb and Reinhard Werner for helpful discussions. This work was supported by the Swiss National Science Foundation (through the NCCR 'Quantum Science and Technology' and grant no. 200020-135048) and the European Research Council (grant no. 258932). SH is funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763. He acknowledges additional support by DFG project no. K05430/1-1.

References

[1] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik", Zeitschrift für Physik **43**, 172–198 (1927) (page 1).

[2] J. A. Wheeler and W. H. Zurek, *Quantum Theory and Measurement* (Princeton University Press, 1984) (page 1).

[3] E. H. Kennard, "Zur Quantenmechanik einfacher Bewegungstypen", Zeitschrift für Physik **44**, 326–352 (1927) (pages 1, 14).

[4] H. P. Robertson, "The Uncertainty Principle", Physical Review **34**, 163 (1929) (pages 1, 16).

[5] H. Maassen and J. B. M. Uffink, "Generalized entropic uncertainty relations", Physical Review Letters **60**, 1103 (1988) (page 1).

[6] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, "The uncertainty principle in the presence of quantum memory", Nature Physics **6**, 659–662 (2010), arXiv:0909.0950 [quant-ph] (pages 1, 3, 15, 17).

[7] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, "Entropic uncertainty relations and their applications", Reviews of Modern Physics **89**, 015002 (2017), arXiv:1511.04857 [quant-ph] (pages 1, 3, 9).

[8] E. Arthurs and J. L. Kelly, "On the Simultaneous Measurement of a Pair of Conjugate Observables", Bell System Technical Journal **44**, 725–729 (1965) (page 1).

[9] C. Y. She and H. Heffner, "Simultaneous Measurement of Noncommuting Observables", Physical Review **152**, 1103–1110 (1966) (page 1).

[10] E. B. Davies, *Quantum theory of open systems* (Academic Press, London, 1976) (pages 1, 3, 4).

[11] S. T. Ali and E. Prugovečki, "Systems of imprimitivity and representations of quantum mechanics on fuzzy phase spaces", Journal of Mathematical Physics **18**, 219–228 (1977) (page 1).

[12] E. Prugovečki, “On fuzzy spin spaces”, Journal of Physics A: Mathematical and General **10**, 543 (1977) (page 1).

[13] P. Busch, “Indeterminacy relations and simultaneous measurements in quantum theory”, International Journal of Theoretical Physics **24**, 63–92 (1985) (page 1).

[14] P. Busch, “Unsharp reality and joint measurements for spin observables”, Physical Review D **33**, 2253–2261 (1986) (page 1).

[15] E. Arthurs and M. S. Goodman, “Quantum Correlations: A Generalized Heisenberg Uncertainty Relation”, Physical Review Letters **60**, 2447–2449 (1988) (page 1).

[16] H. Martens and W. M. de Muynck, “Towards a new uncertainty principle: quantum measurement noise”, Physics Letters A **157**, 441–448 (1991) (page 1).

[17] S. Ishikawa, “Uncertainty relations in simultaneous measurements for arbitrary observables”, Reports on Mathematical Physics **29**, 257–273 (1991) (page 1).

[18] M. G. Raymer, “Uncertainty principle for joint measurement of noncommuting variables”, American Journal of Physics **62**, 986–993 (1994) (page 1).

[19] U. Leonhardt, B. Böhmer, and H. Paul, “Uncertainty relations for realistic joint measurements of position and momentum in quantum optics”, Optics Communications **119**, 296–300 (1995) (page 1).

[20] D. M. Appleby, “Concept of Experimental Accuracy and Simultaneous Measurements of Position and Momentum”, International Journal of Theoretical Physics **37**, 1491–1509 (1998), arXiv:quant-ph/9803046 (page 1).

[21] M. J. W. Hall, “Prior information: How to circumvent the standard joint-measurement uncertainty relation”, Physical Review A **69**, 052113 (2004), arXiv:quant-ph/0309091 (page 1).

[22] R. F. Werner, “The uncertainty relation for joint measurement of position and momentum”, Quantum Information and Computation **4**, 546–562 (2004), arXiv:quant-ph/0405184 (pages 1, 14, 17).

[23] M. Ozawa, “Uncertainty relations for joint measurements of noncommuting observables”, Physics Letters A **320**, 367–374 (2004) (page 1).

[24] Y. Watanabe, T. Sagawa, and M. Ueda, “Uncertainty relation revisited from quantum estimation theory”, Physical Review A **84**, 042121 (2011), arXiv:1010.3571 [quant-ph] (page 1).

[25] P. Busch, P. Lahti, and R. F. Werner, “Proof of Heisenberg’s Error-Disturbance Relation”, Physical Review Letters **111**, 160405 (2013), arXiv:1306.1565 [quant-ph] (pages 1, 17).

[26] P. Busch, P. Lahti, and R. F. Werner, “Heisenberg uncertainty for qubit measurements”, Physical Review A **89**, 012129 (2014), arXiv:1311.0837 [quant-ph] (pages 1, 17).

[27] P. Busch, P. Lahti, and R. F. Werner, “Measurement uncertainty relations”, Journal of Mathematical Physics **55**, 042111 (2014), arXiv:1312.4392 [quant-ph] (pages 1, 17).

[28] V. B. Braginsky and F. Y. Khalili, *Quantum Measurement* (Cambridge University Press, 1992) (page 1).

[29] H. Martens and W. M. de Muynck, “Disturbance, conservation laws and the uncertainty principle”, Journal of Physics A: Mathematical and General **25**, 4887 (1992) (pages 1, 17).

[30] M. Ozawa, “Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement”, Physical Review A **67**, 042105 (2003), arXiv:quant-ph/0207121 (page 1).

[31] Y. Watanabe and M. Ueda, “Quantum Estimation Theory of Error and Disturbance in Quantum Measurement”, (2011), arXiv:1106.2526 [quant-ph] (page 1).

[32] C. Branciard, “Error-tradeoff and error-disturbance relations for incompatible quantum measurements”, Proceedings of the National Academy of Sciences **110**, 6742–6747 (2013), arXiv:1304.2071 [quant-ph] (page 1).

[33] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde, “Noise and Disturbance in Quantum Measurements: An Information-Theoretic Approach”, Physical Review Letters **112**, 050401 (2014), arXiv:1310.6603 [quant-ph] (pages 1, 15, 17).

[34] A. C. Ipsen, “Error-disturbance relations for finite dimensional systems”, (2013), arXiv:1311.0259 [math-ph] (pages 1, 17).

[35] P. J. Coles and F. Furrer, “State-dependent approach to entropic measurement–disturbance relations”, Physics Letters A **379**, 105–112 (2015), arXiv:1311.7637 [quant-ph] (pages 1, 17).

[36] M. Ozawa, “Uncertainty relations for noise and disturbance in generalized quantum measurements”, Annals of Physics **311**, 350–416 (2004) (page 1).

[37] P. Busch, P. Lahti, and R. F. Werner, “Quantum root-mean-square error and measurement uncertainty relations”, Reviews of Modern Physics **86**, 1261–1281 (2014), arXiv:1312.4393 [quant-ph] (page 1).

[38] D. M. Appleby, “Quantum Errors and Disturbances: Response to Busch, Lahti and Werner”, Entropy **18**, 174 (2016), arXiv:1602.09002 [quant-ph] (page 1).

[39] M. Ozawa, “Disproving Heisenberg’s error-disturbance relation”, (2013), arXiv:1308.3540 [quant-ph] (page 1).

[40] D. Kretschmann, D. Schlingemann, and R. Werner, “The Information-Disturbance Tradeoff and the Continuity of Stinespring’s Representation”, IEEE Transactions on Information Theory **54**, 1708–1717 (2008), arXiv:quant-ph/0605009 (pages 2, 5, 16).

[41] D. Kretschmann, D. Schlingemann, and R. F. Werner, “A continuity theorem for Stinespring’s dilation”, Journal of Functional Analysis **255**, 1889–1904 (2008), arXiv:0710.2495 [quant-ph] (pages 2, 5).

[42] B.-G. Englert, “Fringe Visibility and Which-Way Information: An Inequality”, Physical Review Letters **77**, 2154 (1996) (pages 2, 16, 25).

[43] J. M. Renes and J.-C. Boileau, “Conjectured Strong Complementary Information Tradeoff”, Physical Review Letters **103**, 020402 (2009), arXiv:0806.3984 [quant-ph] (page 3).

[44] M. Tomamichel and R. Renner, “Uncertainty Relation for Smooth Entropies”, Physical Review Letters **106**, 110506 (2011), arXiv:1009.2015 [quant-ph] (pages 3, 15, 17).

[45] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, “Tight finite-key analysis for quantum cryptography”, Nature Communications **3**, 634 (2012), arXiv:1103.4130 [quant-ph] (page 3).

[46] F. G. Lacerda, J. M. Renes, and R. Renner, “Classical leakage resilience from fault-tolerant quantum computation”, (2014), arXiv:1404.7516 [quant-ph] (pages 3, 15).

[47] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, Lecture notes in physics 190 (Springer-Verlag, Berlin, 1983) (page 3).

[48] R. F. Werner, “Quantum Information Theory — an Invitation”, in *Quantum Information*, Springer Tracts in Modern Physics 173 (Springer Berlin Heidelberg, 2001), pp. 14–57, arXiv:quant-ph/0101061 (page 3).

[49] M. M. Wolf, “Quantum Channels and Operations: A Guided Tour”, (2012) (pages 3, 21).

[50] C. Bény and F. Richter, “Algebraic approach to quantum theory: a finite-dimensional guide”, (2015), arXiv:1505.03106 [quant-ph] (page 3).

[51] A. Y. Kitaev, “Quantum computations: Algorithms and error correction”, Russian Mathematical Surveys **52**, 1191–1249 (1997) (pages 4, 17).

[52] V. Paulsen, *Completely Bounded Maps and Operator Algebras*, Vol. 78, Cambridge Studies in Advanced Mathematics (Cambridge University Press, Jan. 15, 2003), 320 pp. (page 4).

[53] A. Gilchrist, N. K. Langford, and M. A. Nielsen, “Distance measures to compare real and ideal quantum processes”, Physical Review A **71**, 062310 (2005), arXiv:quant-ph/0408063 (page 4).

[54] J. Watrous, “Semidefinite Programs for Completely Bounded Norms”, Theory of Computing **5**, 217–238 (2009), arXiv:0901.4709 [quant-ph] (pages 4, 21).

[55] J. Watrous, “Simpler semidefinite programs for completely bounded norms”, Chicago Journal of Theoretical Computer Science **2013**, 8 (2013), arXiv:1207.5726 [quant-ph] (page 4).

[56] W. F. Stinespring, “Positive functions on C*-algebras”, Proceedings of the American Mathematical Society **6**, 211–216 (1955) (page 4).

[57] P. J. Coles and M. Piani, “Improved entropic uncertainty relations and information exclusion relations”, Physical Review A **89**, 022112 (2014), arXiv:1307.4265 [quant-ph] (page 9).

[58] P. W. Shor and J. Preskill, “Simple Proof of Security of the BB84 Quantum Key Distribution Protocol”, Physical Review Letters **85**, 441 (2000), arXiv:quant-ph/0003004 (page 15).

[59] I. Devetak, “The private classical capacity and quantum capacity of a quantum channel”, IEEE Transactions on Information Theory **51**, 44–55 (2005), arXiv:quant-ph/0304127 (page 15).

[60] J. M. Renes, “Duality of privacy amplification against quantum adversaries and data compression with quantum side information”, Proceedings of the Royal Society A **467**, 1604–1623 (2011), arXiv:1003.0703 [quant-ph] (page 15).

[61] J. M. Renes, “The Physics of Quantum Information: Complementarity, Uncertainty, and Entanglement”, Habilitation (TU Darmstadt, 2012), arXiv:1212.2379 [quant-ph] (page 15).

[62] J. M. Renes, “Uncertainty relations and approximate quantum error correction”, Physical Review A **94**, 032314 (2016), arXiv:1605.01420 [quant-ph] (page 16).

[63] P. J. Coles, J. Kaniewski, and S. Wehner, “Equivalence of wave–particle duality to entropic uncertainty”, Nature Communications **5**, 5814 (2014), arXiv:1403.4687 [quant-ph] (page 16).

[64] P. J. Coles, “Entropic framework for wave-particle duality in multipath interferometers”, Physical Review A **93**, 062111 (2016), arXiv:1512.09081 [quant-ph] (page 16).

[65] K. Korzekwa, D. Jennings, and T. Rudolph, “Operational constraints on state-dependent formulations of quantum error-disturbance trade-off relations”, Physical Review A **89**, 052108 (2014), arXiv:1311.5506 [quant-ph] (page 16).

[66] A. Barchielli, M. Gregoratti, and A. Toigo, “Measurement uncertainty relations for discrete observables: Relative entropy formulation”, (2016), arXiv:1608.01986 [math-ph] (page 17).

[67] M. F. Sacchi, “Entanglement can enhance the distinguishability of entanglement-breaking channels”, Physical Review A **72**, 014305 (2005), arXiv:quant-ph/0505174 (page 17).

[68] V. P. Belavkin, “Optimal multiple quantum statistical hypothesis testing”, Stochastics **1**, 315 (1975) (page 24).

[69] P. Hausladen and W. K. Wootters, “A ‘Pretty Good’ Measurement for Distinguishing Quantum States”, Journal of Modern Optics **41**, 2385 (1994) (page 24).

A Entanglement improves the distinguishability of measurements

Here we give an example of two measurements whose distinguishability is improved by entanglement. Let $\mathcal{E}_1$ be a measurement in an arbitrary chosen basis $|b_0\rangle$, $|b_1\rangle$, and $|b_2\rangle$, and define $\mathcal{E}_2$ be measurement in the basis given by $|\theta_0\rangle = \frac{1}{3}(2|b_0\rangle + 2|b_1\rangle - |b_2\rangle)$, $|\theta_1\rangle = \frac{1}{3}(-1|b_0\rangle + 2|b_1\rangle + 2|b_2\rangle)$ and $|\theta_2\rangle = \frac{1}{3}(2|b_0\rangle - |b_1\rangle + 2|b_2\rangle)$. Using $T_k = |b_k\rangle\langle b_k| - |\theta_k\rangle\langle\theta_k|$, the largest distinguishability to be had without entanglement is given by

$$\delta'(\mathcal{E}_1,\mathcal{E}_2) = \max_{\varrho} \tfrac{1}{2}\sum_{k=0}^{2}\left|\mathrm{Tr}[\varrho T_k]\right| \tag{52a}$$

$$= \max_{\varrho}\max_{\{s_k=\pm1\}} \tfrac{1}{2}\sum_{k=0}^{2}\mathrm{Tr}[s_k T_k \varrho] \tag{52b}$$

$$= \max_{\{s_k=\pm1\}} \Big\|\sum_{k=0}^{2} s_k T_k\Big\|_\infty . \tag{52c}$$

Checking the eight combinations of s_k, one easily finds that the maximimum value is $\sqrt{5}/3$.

Meanwhile, if we use the state

$$\varrho = \frac{1}{6}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \tag{53}$$

to define $\Psi = (\mathbb{1}\otimes\sqrt{\varrho})\Omega(\mathbb{1}\otimes\sqrt{\varrho})$ for Ω the projector onto $|\Omega\rangle = \sum_k |b_k\rangle\otimes|b_k\rangle$, then

$$\delta(\mathcal{E}_1,\mathcal{E}_2) \geq \tfrac{1}{2}\sum_{k=0}^{2}\left\|\mathrm{Tr}_1[(T_k\otimes\mathbb{1})\Psi]\right\|_1 . \tag{54}$$

Direct calculation shows that $\delta(\mathcal{E}_1,\mathcal{E}_2) \geq \sqrt{3}/2$. Thus, there exist projective measurements for which $\delta(\mathcal{E}_1,\mathcal{E}_2) > \delta'(\mathcal{E}_1,\mathcal{E}_2)$.

B Computing error and disturbance by convex optimization

Here we detail how to compute the error and disturbance quantities via semidefinite programming and calculate these for the nonideal qubit X measurement example. Given a Hilbert space $\mathcal{H}$ with basis $\{|k\rangle\}_{k=1}^{d}$, define, just as above, $|\Omega\rangle = \sum_{k=1}^{d}|k\rangle\otimes|k\rangle \in \mathcal{H}\otimes\mathcal{H}$. Then, for any channel $\mathcal{E}$, let $\mathtt{C}$ denote the Choi mapping of $\mathcal{E}^*$ to an unnormalized bipartite state,

$$\mathtt{C}(\mathcal{E}) := \mathcal{E}^*\otimes\mathcal{I}(|\Omega\rangle\langle\Omega|) \in \mathcal{B}(\mathcal{H}_B\otimes\mathcal{H}_A) . \tag{55}$$

The action of the channel can be compactly expressed in terms of the Choi operator as $\mathcal{E}_{A\to B}(\Lambda_B) = \mathrm{Tr}_B[\Lambda_B \mathtt{C}(\mathcal{E})_{BA}]$ or in the Schrödinger picture as $\mathcal{E}^*_{A\to B}(\varrho_A) = \mathrm{Tr}_A[\mathtt{C}(\mathcal{E})_{BA}\varrho_A^T]$, where the transpose is taken in the basis defining $\mathtt{C}$ (see, e.g. [49]). The cb norm can then be expressed in primal and dual form as [54]

$$\begin{aligned} \tfrac{1}{2}\|\mathcal{E}_{A\to B}\|_{\mathrm{cb}} = \ & \underset{K,\varrho}{\text{maximum}} && \mathrm{Tr}[\mathtt{C}(\mathcal{E})_{BA}K_{BA}] \\ & \text{subject to} && K_{BA} - \mathbb{1}_B\otimes\varrho_A \leq 0,\ \mathrm{Tr}[\varrho_A]\leq 1, \\ & && \varrho_A, K_{BA} \geq 0, \end{aligned} \tag{56}$$

$$\begin{aligned} = \ & \underset{T,\lambda}{\text{minimum}} && \lambda \\ & \text{subject to} && T_{BA} \geq \mathtt{C}(\mathcal{E})_{BA},\ \lambda\mathbb{1}_A - T_A \geq 0, \\ & && T_{BA}, \lambda \geq 0 . \end{aligned} \tag{57}$$

Note that in the dual formulation the objective function is just the operator norm $\|T_A\|_\infty$. For infinite-dimensional systems the Choi operator does not have such a nice form, though it might be possible to formulate the cb norm of Gaussian channels as a tractable optimization.

The additional optimizations involving $\mathcal{R}$ in the measures of error and disturbance are immediately compatible with the dual formulation in (57), and so these quantities can be cast as semidefinite programs. To start, consider the error in measuring X. With $Q_{\mathsf{X}A} = \mathsf{C}(\mathcal{Q}_X)$ and $E_{\mathsf{Y}BA} = \mathsf{C}(\mathcal{E}_{A\to\mathsf{Y}B})$, we have

$$\begin{aligned} \varepsilon_X(\mathcal{E}_{A\to\mathsf{Y}B}) = \underset{T,\lambda,R}{\text{minimum}} \quad & \lambda \\ \text{subject to} \quad & T_{\mathsf{X}A} + \mathrm{Tr}_{\mathsf{Y}}[R_{\mathsf{XY}}E_{\mathsf{Y}A}] \geq Q_{\mathsf{X}A},\ \lambda\mathbb{1}_A - T_A \geq 0,\ R_{\mathsf{Y}} = \mathbb{1}_{\mathsf{Y}}, \\ & \lambda, T_{\mathsf{X}A}, R_{\mathsf{XY}} \geq 0\,. \end{aligned} \tag{58}$$

Without loss of generality, we may restrict the operator $T_{\mathsf{X}A}$ to be a hybrid classical-quantum operator, classical on X, and of course R_{XY} is classical on both systems. This is also the reason it is unnecessary to transpose Y in $\mathrm{Tr}_{\mathsf{Y}}[R_{\mathsf{XY}}E_{\mathsf{Y}A}]$. Further symmetries of $Q_{\mathsf{X}A}$ and $E_{\mathsf{X}A}$ can be quite helpful in simplifying the program, but we will not pursue this further here. The associated primal form is as follows.

$$\begin{aligned} \varepsilon_X(\mathcal{E}_{A\to\mathsf{Y}B}) = \underset{K,\varrho,L}{\text{maximum}} \quad & \mathrm{Tr}[Q_{\mathsf{X}A}K_{\mathsf{X}A}] - \mathrm{Tr}[L_{\mathsf{Y}}] \\ \text{subject to} \quad & K_{\mathsf{X}A} - \mathbb{1}_{\mathsf{X}} \otimes \varrho_A \leq 0,\ \mathrm{Tr}[\varrho_A] \leq 1,\ \mathrm{Tr}_A[E_{\mathsf{Y}A}K_{\mathsf{X}A}] - L_{\mathsf{Y}} \otimes \mathbb{1}_{\mathsf{X}} \leq 0, \\ & \varrho_A, K_{\mathsf{X}A} \geq 0, L_{\mathsf{Y}} = L_{\mathsf{Y}}^*. \end{aligned} \tag{59}$$

In writing an equality we have assumed that the duality gap is zero. But this is easy enough to show using the Slater condition, namely by ensuring that the value of the minimization is finite and that there exists a strictly feasible set of maximization variables. The former holds because ε_X is the infinimum of the distinguishability, and hence $\varepsilon_X(\mathcal{E}) \geq 0$. Meanwhile, a strictly feasible set of variables in (59) is given by $K = \frac{1}{2}k\mathbb{1}$, $\varrho = k\mathbb{1}$, and $L = kE_{\mathsf{Y}}$ for $k < 1/\dim(A)$.

To formulate the measurement disturbance $\nu_Z(\mathcal{E}_{A\to\mathsf{Y}B})$ we are interested in $\mathsf{C}(\mathcal{E}\mathcal{R}\mathcal{T}_{\mathsf{Y}}\mathcal{Q}_Z)$, which can be expressed as a linear map on $R_{AB\mathsf{Y}}$:

$$\mathsf{C}(\mathcal{E}\mathcal{R}\mathcal{T}_{\mathsf{Y}}\mathcal{Q}_Z) = \mathrm{Tr}_{A'\mathsf{Y}B}[Q_{\mathsf{Z}A'}R_{A'\mathsf{Y}B}^{T_{A'}}E_{\mathsf{Y}BA}^{T_B}] \tag{60a}$$

$$= \mathrm{Tr}_{A'\mathsf{Y}B}[R_{A'\mathsf{Y}B}Q_{\mathsf{Z}A'}^{T_{A'}}E_{\mathsf{Y}BA}^{T_B}]\,. \tag{60b}$$

In the second step we have transposed the A' system in the first. Then we have

$$\begin{aligned} \nu_Z(\mathcal{E}_{A\to\mathsf{Y}B}) = \underset{T,\lambda,R}{\text{minimum}} \quad & \lambda \\ \text{subject to} \quad & T_{\mathsf{Z}A} + \mathrm{Tr}_{A'\mathsf{Y}B}[R_{A'\mathsf{Y}B}Q_{\mathsf{Z}A'}^{T_{A'}}E_{\mathsf{Y}BA}^{T_B}] \geq Q_{\mathsf{Z}A},\ \lambda\mathbb{1}_A - T_A \geq 0,\ R_{\mathsf{Y}B} = \mathbb{1}_{\mathsf{Y}B}, \\ & \lambda, T_{\mathsf{Z}A}, R_{A'\mathsf{Y}B} \geq 0\,, \end{aligned} \tag{61}$$

$$\begin{aligned} = \underset{K,\varrho,L}{\text{maximum}} \quad & \mathrm{Tr}[Q_{\mathsf{Z}A}K_{\mathsf{Z}A}] - \mathrm{Tr}[L_{\mathsf{Y}B}] \\ \text{subject to} \quad & K_{\mathsf{Z}A} - \mathbb{1}_Z \otimes \varrho_A \leq 0,\ \mathrm{Tr}[\varrho_A] \leq 1,\ \mathrm{Tr}_{\mathsf{Z}A}[Q_{\mathsf{Z}A'}E_{\mathsf{Y}BA}K_{\mathsf{Z}A}] - \mathbb{1}_{A'} \otimes L_{\mathsf{Y}B} \leq 0, \\ & \varrho_A, K_{\mathsf{Z}A} \geq 0, L_{\mathsf{Y}B} = L_{\mathsf{Y}B}^*\,. \end{aligned} \tag{62}$$

Here we have absorbed the transposes over A' and B into $\mathbb{1}_{A'}$ and the definition of $L_{\mathsf{Y}B}$, since this does not affect Hermiticity or the value of the objective function. Strong duality is essentially the same as before: The minimization is finite and we can choose $K = \frac{1}{2}k\mathbb{1}$ and $\varrho = k\mathbb{1}$. Then in the third constraint we have $\mathrm{Tr}_{\mathsf{Z}A}[Q_{\mathsf{Z}A'}E_{\mathsf{Y}BA}K_{\mathsf{Z}A}] = \frac{1}{2}k\mathbb{1}_{A'} \otimes E_{\mathsf{Y}B}$ since $\mathcal{Q}_Z$ is unital. Setting $L = kE_{\mathsf{Y}B}$ gives a strictly feasible set.

Finally, we come to the two preparation disturbance measures. The first is simply

$$\begin{aligned} \eta_Z(\mathcal{E}_{A\to\mathsf{Y}B}) = \underset{T,\lambda,R}{\text{minimum}} \quad & \lambda \\ \text{subject to} \quad & T_{A\mathsf{Z}} + \mathrm{Tr}_{\mathsf{Y}BA'}[R_{A\mathsf{Y}B}E_{\mathsf{Y}BA'}^{T_B}P_{A'\mathsf{Z}}^{T_{A'}}] \geq P_{A\mathsf{Z}},\ \lambda\mathbb{1}_{\mathsf{Z}} - T_{\mathsf{Z}} \geq 0,\ R_{\mathsf{Y}B} = \mathbb{1}_{\mathsf{Y}B}, \\ & \lambda, T_{\mathsf{X}A}, R_{A\mathsf{Y}B} \geq 0\,, \end{aligned} \tag{63}$$

$$\begin{aligned} = \underset{K,\varrho,L}{\text{maximum}} \quad & \mathrm{Tr}[P_{A\mathsf{Z}}K_{A\mathsf{Z}}] - \mathrm{Tr}[L_{\mathsf{Y}B}] \\ \text{subject to} \quad & K_{A\mathsf{Z}} - \mathbb{1}_A \otimes \varrho_{\mathsf{Z}} \leq 0,\ \mathrm{Tr}[\varrho_{\mathsf{Z}}] \leq 1,\ \mathrm{Tr}_{A'\mathsf{Z}}[E_{\mathsf{Y}BA'}P_{A'\mathsf{Z}}^{T_{A'}}K_{A\mathsf{Z}}] - \mathbb{1}_A \otimes L_{\mathsf{Y}B} \leq 0, \\ & \varrho_{\mathsf{Z}}, K_{A\mathsf{Z}} \geq 0, L_{\mathsf{Y}B} = L_{\mathsf{Y}B}^*\,. \end{aligned} \tag{64}$$

Here we have absorbed the transpose on B into the definition of $L_{\mathsf{Y}B}$ since this doesn't affect Hermiticity or the value of the objective function. Strong duality holds as before, and also for the demerit measure which reads

$$\begin{aligned}
\tfrac{d-1}{d} - \widehat{\eta}_Z(\mathcal{E}_{A\to \mathsf{Y}B}) = \;& \underset{T,\lambda,\sigma}{\text{minimum}} && \lambda && (65)\\
& \text{subject to} && T_{\mathsf{Y}B\mathsf{Z}} + \sigma_{\mathsf{Y}B}\otimes \mathbb{1}_\mathsf{Z} \geq \mathrm{Tr}_A[E_{\mathsf{Y}BA}P_{A\mathsf{Z}}^{T_A}],\ \lambda\mathbb{1}_\mathsf{Z} - T_\mathsf{Z} \geq 0,\ \mathrm{Tr}[\sigma_{\mathsf{Y}B}] = 1, &&\\
& && \lambda, T_{\mathsf{Y}B\mathsf{Z}}, \sigma_{\mathsf{Y}B} \geq 0\,, &&\\
= \;& \underset{K,\varrho,\mu}{\text{maximum}} && \mathrm{Tr}[E_{\mathsf{Y}BA}P_{A\mathsf{Z}}^{T_A}K_{\mathsf{Y}B\mathsf{Z}}] - \mu && (66)\\
& \text{subject to} && K_{\mathsf{Y}B\mathsf{Z}} - \mathbb{1}_{\mathsf{Y}B}\otimes \varrho_\mathsf{Z} \leq 0,\ \mathrm{Tr}[\varrho_\mathsf{Z}] \leq 1,\ K_{\mathsf{Y}B} - \mu\mathbb{1}_{\mathsf{Y}B} \leq 0, &&\\
& && \varrho_\mathsf{Z}, K_{\mathsf{Y}B\mathsf{Z}} \geq 0, \mu \in \mathbb{R}\,. &&
\end{aligned}$$

Now let us consider the particular example described in the main text, a suboptimal X measurement. Suppose we use $|\varphi_x\rangle$ from the ideal X measurement to define the Choi operator. After a bit of calculation, one finds that the Choi operator $E_{\mathsf{Y}BA}$ of $\mathcal{E}_{\mathsf{Y}B|A}$ is given by

$$E_{\mathsf{Y}BA} = |b_0\rangle\langle b_0|_\mathsf{Y} \otimes |\Psi\rangle\langle\Psi|_{BA} + |b_1\rangle\langle b_1|_\mathsf{Y} \otimes (\sigma_z\otimes\sigma_z)|\Psi\rangle\langle\Psi|_{BA}(\sigma_z\otimes\sigma_z), \tag{67}$$

where $|\Psi\rangle = \cos\frac{\theta}{2}|\varphi_0\rangle\otimes|\varphi_0\rangle + \sin\frac{\theta}{2}|\varphi_1\rangle\otimes|\varphi_1\rangle$. Tracing out B gives the Choi operator of just the measurement result Y, $E_{\mathsf{Y}A} = \sum_x |b_x\rangle\langle b_x|_\mathsf{Y} \otimes \Lambda_x$, with $\Lambda_x = \frac{1}{2}\mathbb{1} + \frac{1}{2}(-1)^x\cos\theta\ \sigma_x$.

To compute the measurement error $\varepsilon_X(\mathcal{E})$, suppose that no recovery operation is applied, i.e. the outcome Y is treated as X. Then we can work with $E_{\mathsf{X}A}$ and dispense with R so that the third constraint in (58) is satisfied. To satisfy the first constraint, choose $T_{\mathsf{X}A}$ to be the positive part of $Q_{\mathsf{X}A} - E_{\mathsf{X}A}$. This gives $T_{\mathsf{X}A} = \frac{1}{2}(1-\cos\theta)\sum_x |b_x\rangle\langle b_x| \otimes |\varphi_x\rangle\langle\varphi_x|$; consequently, $T_A = \frac{1}{2}(1-\cos\theta)\mathbb{1}_A$ and therefore $\varepsilon_X(\mathcal{E}) \leq \frac{1}{2}(1-\cos\theta)$. On the other hand, $K_{\mathsf{X}A} = \frac{1}{2}Q_{\mathsf{X}A}$ and $\varrho_A = \frac{1}{2}\mathbb{1}_A$ satisfy the first two constraints in (59). The last constraint involves the quantity $\mathrm{Tr}_A[E_{\mathsf{Y}A}K_{\mathsf{X}A}] = \frac{1}{4}\sum_{xy}|b_y\rangle\langle b_y|_\mathsf{Y} \otimes |b_x\rangle\langle b_x|_\mathsf{X}(1+(-1)^{x+y}\cos\theta)$ and can therefore be satisfied by choosing $L_\mathsf{Y} = \frac{1}{4}(1+\cos\theta)\mathbb{1}_\mathsf{Y}$. Evaluating the objective function gives $\varepsilon_X(\mathcal{E}) \geq \frac{1}{2}(1-\cos\theta)$.

Note that the choice of $K_{\mathsf{X}A}$ corresponds to the unentangled test of randomly inputting $|\varphi_x\rangle$ and checking that the result is x. We could have anticipated that unentangled tests would be sufficient in this case, since the optimal and actual measurements are both diagonal in the σ_x basis: Any input state can be freely dephased in this basis, thus removing any entanglement.

Next, consider the measurement disturbance $\nu_Z(\mathcal{E})$. Proceeding as with measurement error, suppose that no recovery operation is applied, so that the output B is just regarded as $A' \simeq A$ and the third constraint in (61) is trivially satisfied. For the first constraint we need only the operator $\mathrm{Tr}_{\mathsf{Y}A'}[Q_{\mathsf{Z}A'}E_{\mathsf{Y}A'A}^{T_{A'}}]$, and after some calculation we find that it equals $\sum_z |b_z\rangle\langle b_z| \otimes \Gamma_z$ with $\Gamma_z = \frac{1}{2}(\mathbb{1} + (-1)^z\sin\theta\sigma_z)$. Thus, the optimization is just like that of $\varepsilon_X(\mathcal{E})$, but with $\cos\theta$ replaced by $\sin\theta$. Hence $\nu_Z(\mathcal{E}) \leq \frac{1}{2}(1-\sin\theta)$. To show the other inequality from the maximization form (62) also proceeds as before, starting with $K_{\mathsf{Z}A} = \frac{1}{2}Q_{\mathsf{Z}A}$ and $\varrho_A = \frac{1}{2}\mathbb{1}_A$. For the third constraint a bit of calculation shows

$$\mathrm{Tr}_{\mathsf{Z}A}[Q_{\mathsf{Z}A'}E_{\mathsf{Y}BA}K_{\mathsf{Z}A}] = \frac{1}{4}\sum_{x,z}|b_z\rangle\langle b_z|_{A'} \otimes |b_x\rangle\langle b_x|_\mathsf{Y} \otimes (\sigma_x^z\sigma_z^x|\psi\rangle\langle\psi|\sigma_z^x\sigma_x^z)_B, \tag{68}$$

with $|\psi\rangle = \frac{1}{\sqrt{2}}(\sqrt{1+\sin\theta}\,|\theta_0\rangle + \sqrt{1-\sin\theta}\,|\theta_1\rangle$. Choosing

$$L_{\mathsf{Y}B} = \frac{1}{8}\sum_x |b_x\rangle\langle b_x|_\mathsf{Y} \otimes ((1+\sin\theta)\mathbb{1} + (-1)^x\cos\theta\,\sigma_x)_B \tag{69}$$

satisfies the constraints, and the objective function becomes $\frac{1}{2}(1-\sin\theta)$. As with $\varepsilon_X(\mathcal{E})$, entangled inputs do not increase the distinguishability in this particular case.

A trivial recovery map also optimizes $\eta_Z(\mathcal{E})$. To see this, set $K_{A\mathsf{Z}} = \frac{1}{2}P_{A\mathsf{Z}}$ and $\varrho_\mathsf{Z} = \frac{1}{2}\mathbb{1}_\mathsf{Z}$. Then in the third constraint of (64) we have $\mathrm{Tr}_{A'}[E_{\mathsf{Y}BA'}P_{A'\mathsf{Z}}^{T_{A'}}K_{A\mathsf{Z}}]$, which is precisely the same as (68) with A' replaced by A. Hence, if we choose $L_{\mathsf{Y}B}$ as in (69), we obtain the lower bound $\eta_Z(\varepsilon) \geq \frac{1}{2}(1-\sin\theta)$. To establish optimality, suppose $\mathcal{R}$ does nothing but discard the Y system. In the minimization (63) we then have $\mathrm{Tr}_{\mathsf{Y}A'}[E_{\mathsf{Y}AA'}P_{A'\mathsf{Z}}^{T_{A'}}]$, which is the same as $\mathrm{Tr}_{\mathsf{Y}A'}[Q_{\mathsf{Z}A'}E_{\mathsf{Y}A'A}^{T_{A'}}]$ from $\nu_Z(\mathcal{E})$. Proceeding as there, we find the matching upper bound.

Finally, consider $\widehat{\eta}_Z(\mathcal{E})$. Here there are two possible outputs of $\mathcal{P}_Z\mathcal{E}$, call them ξ_0 and ξ_1. It is not difficult to show that for arbitrary ξ_z the distinguishability is precisely $\widehat{\eta}_Z(\mathcal{E}) = \frac{1}{2}(1-\delta(\xi_0,\xi_1))$. On the one hand, we can simply pick the output of $\mathcal{C}$ to be $\xi = \frac{1}{2}(\xi_0+\xi_1)$. Then, with T in (65) the positive part of $\sum_z |z\rangle\langle z|_\mathsf{Z} \otimes (\xi_z - \xi)$, the objective function becomes $\frac{1}{2}\delta(\xi_0,\xi_1)$. On the other hand, in (66) we can choose

$K_{\mathsf{YBZ}} = \frac{1}{2}|0\rangle\langle 0|_{\mathsf{Z}} \otimes \Lambda_{\mathsf{Y}B} + \frac{1}{2}|1\rangle\langle 1|_{\mathsf{Z}} \otimes (\mathbb{1}-\Lambda)_{\mathsf{Y}B}$, for Λ the projector onto the nonnegative part of $\xi_0 - \xi_1$. Then $\mu = \frac{1}{2}$ and $\varrho = \frac{1}{2}\mathbb{1}$ are feasible and lead again to the same objective function. In this particular case the two states are $\xi_1 = \sigma_z \xi_0 \sigma_z$ and $\xi_0 = \frac{1}{2}\sum_x |b_x\rangle\langle b_x|_{\mathsf{Y}} \otimes \sigma_x |\psi\rangle\langle\psi|\sigma_x$, which yields $\delta(\xi_0, \xi_1) = \sin\theta$ and hence $\widehat{\eta}_Z(\mathcal{E}) = \frac{1}{2}(1 - \sin\theta)$.

C Counterexample channel

Here we present the calculations involved in §6.1 Let $|\xi_z\rangle_{BD} = \sum_y \sqrt{p_{yz}}\,|y\rangle_B\,|y\rangle_D$. Then the isometry is just

$$V = \sum_z |\xi_z\rangle\,|z\rangle_C\,\langle z|_A\,. \tag{70}$$

Observe that the action on $|\tilde{x}\rangle$ states leads to symmetric output in BC:

$$\begin{aligned} V\,|\tilde{x}\rangle &= \tfrac{1}{\sqrt{d}} \sum_z \omega^{xz} V\,|z\rangle & \text{(71a)}\\ &= \tfrac{1}{\sqrt{d}} \sum_z \omega^{xz}\,|\xi_z\rangle_{BD}\,|z\rangle_C & \text{(71b)}\\ &= Z_C^x\,\tfrac{1}{\sqrt{d}} \sum_z |\xi_z\rangle_{BD}\,|z\rangle_C\,. & \text{(71c)} \end{aligned}$$

Therefore, the probability of incorrectly identifying any particular input state is the same as any other, and we can consider the case that the input x value is chosen uniformly at random. We can further simplify the BC output by defining $p_y = \frac{1}{d}\sum_z p_{yz}$ and

$$|\eta_y\rangle = \tfrac{1}{\sqrt{d}} \sum_z \sqrt{p_{yz}/p_y}\,|z\rangle\,, \tag{72}$$

which is a normalized state on $\mathcal{H}_C$ for each y. Then we have

$$V\,|\tilde{x}\rangle = Z_C^x \sum_y \sqrt{p_y}\,|\eta_y\rangle_C\,|y\rangle_B\,|y\rangle_D\,. \tag{73}$$

Ignoring the D system will produce a classical-quantum state, with system B recording the classical value y, which occurs with probability p_y, and C the quantum state $Z^x\,|\eta_y\rangle$. The optimal measurement therefore has elements Λ_x of the form $\Lambda_x = \sum_y |y\rangle\langle y|_B \otimes (\Gamma_{x,y})_C$ for some set of POVMs $\{\Gamma_{x,y}\}_y$. In every sector of fixed y value, the measurement has to distinguish between a set of pure states occurring with equal probabilities. Therefore, by a result going back to Belavkin, the optimal measurement is the so-called “pretty good measurement” [68, 69]. This has measurement elements $\Gamma_{x,y}$ which project onto the orthonormal states $|\mu_{x,y}\rangle = S^{-1/2}Z^x\,|\eta_y\rangle$, where $S = \sum_x Z^x|\eta_y\rangle\langle\eta_y|Z^{-x}$. It is easy to work out that $S = \sum_x (p_{yz}/p_y)|z\rangle\langle z|$, and thus $|\mu_{x,y}\rangle = |\tilde{x}\rangle$ for all y. Hence, we can in fact dispense with the B system altogether, since the particular value of y does not alter the optimal measurement. The average guessing probability is thus

$$\begin{aligned} p_{\text{guess}} &= \tfrac{1}{d} \sum_{x,y} p_y \big|\,\langle\tilde{x}|\,Z^x\,|\eta_y\rangle\,\big|^2 & \text{(74a)}\\ &= \sum_y p_y \big|\,\langle\tilde{0}|\eta_y\rangle\,\big|^2 & \text{(74b)}\\ &= \tfrac{1}{d^2} \sum_y \Big(\sum_z \sqrt{p_{yz}}\Big)^2, & \text{(74c)} \end{aligned}$$

as intended.

D Englert’s complementarity relation

Here we describe Englert’s setup in our formalism and establish (49). He considers a Mach-Zehnder interferometer with a relative phase shift between the two arms and additional which-way detectors in each arm. To the two possible paths inside the interferometer we may associate the (orthogonal) eigenstates $|\vartheta_z\rangle$ of an observable Z, with $z \in \{0,1\}$. For simplicity, we assume Z has eigenvalues $(-1)^z$. The action of a relative φ phase shift is described by the unitary $U_{\text{PS}} = \sum_{z=0}^{1} e^{iz\varphi}|\vartheta_z\rangle\langle\vartheta_z|$. It will prove convenient to choose $\varphi = 0$ below, but we leave it arbitrary for now. Meanwhile, the which-way detectors can be described as producing different states of an ancilla system, depending on which path the photon takes. For pure ancilla states $|\gamma_z\rangle$, the detector corresponds to the isometry $U_{\text{WW}} = \sum_{z=0}^{1} |\vartheta_z\rangle\langle\vartheta_z|_Q \otimes |\gamma_z\rangle_A$, where A denotes the ancilla and Q the system itself, which Englert terms a "quanton".

Ignoring the phase shifts associated with reflection, the output modes of a symmetric (50/50) beamsplitter are related to the input modes by the unitary $U_{\text{BS}} = \sum_{z=0}^{1} |\vartheta_z\rangle\langle\varphi_z|$, with $|\varphi_x\rangle = \frac{1}{\sqrt{2}}\sum_z (-1)^{xz}|\vartheta_z\rangle$ for $x \in \{0,1\}$. We may associate these states with the observable X, also taking eigenvalues $(-1)^x$. Observe that all three complementarity measures are $\frac{1}{2}$. The entire Mach-Zehnder device can be described by the isometry

$$U_{\text{MZ}} = U_{\text{BS}}U_{\text{PS}}U_{\text{WW}}U_{\text{BS}} \tag{75a}$$

$$= \sum_{x,z=0}^{1} e^{ix\varphi}\,|\vartheta_z\rangle\langle\varphi_z|\vartheta_x\rangle\langle\varphi_x|_Q \otimes |\gamma_x\rangle_A \tag{75b}$$

$$= \sum_{x=0}^{1} e^{ix\varphi}|\varphi_x\rangle\langle\varphi_x|_Q \otimes |\gamma_x\rangle_A\,. \tag{75c}$$

When the ancilla is subsequently measured so as to extract information about the path, we may regard the whole operation as an apparatus $\mathcal{E}_{\text{MZ}}$ with one quantum and one classical output.

The available "which-way" information, associated with particle-like behavior of Q, is characterized by the distinguishability $D := \delta(\gamma_0, \gamma_1)$. Given the particular form of U in (75), we may set $\sin\theta = \langle\gamma_0|\gamma_1\rangle$ for $\theta \in \mathbb{R}$ without loss of generality; D is then $\cos\theta$. This amounts to defining $|\gamma_k\rangle = \cos\frac{\theta}{2}|k\rangle + \sin\frac{\theta}{2}|k+1\rangle$, where the states $\{|k\rangle\}_{k=0}^{1}$ form an orthonormal basis and arithmetic inside the ket is modulo two. Thus, $\mathcal{E}_{\text{MZ}}$ with $\varphi = 0$ is precisely the nonideal qubit X measurement $\mathcal{E}$ considered in §3. We shall see momentarily that $\varphi = 0$ can be chosen without loss of generality. Using (14) we have $\varepsilon_X(\mathcal{E}_{\text{MZ}}) = \frac{1}{2}(1-D)$ as claimed.

Meanwhile, the fringe visibility V is defined as the difference in probability (or population) in the two output modes of the interferometer, maximized over the choice of input state. Since $Z = |\vartheta_0\rangle\langle\vartheta_0| - |\vartheta_1\rangle\langle\vartheta_1|$, this is just

$$V = \max_{\varrho}\left|\text{Tr}[(Z_Q \otimes \mathbb{1}_A)U_{\text{MZ}}\varrho U_{\text{MZ}}^*]\right|. \tag{76}$$

A straightforward calculation yields $U_{\text{MZ}}^*(Z_Q \otimes \mathbb{1}_A)U_{\text{MZ}} = \sin\theta(\cos\varphi\, Z + i\sin\varphi\, XZ)$. It can be verified that $(\cos\varphi\, Z + i\sin\varphi\, XZ)$ has eigenvalues ±1, and therefore $V = \sin\theta$. Thus, $V^2 + D^2 = 1$ in this case (cf. [42, Eq. 11]). Note that φ does not appear in the visiblity itself, justifying our choice of $\varphi = 0$ above. By (15), $\nu_Z(\mathcal{E}_{\text{MZ}}) = \eta_Z(\mathcal{E}_{\text{MZ}}) = \widehat{\eta}_Z(\mathcal{E}_{\text{MZ}}) = \frac{1}{2}(1-V)$.

B.2 Jointly constrained semidefinite bilinear programming with an application to Dobrushin curves

Jointly constrained semidefinite bilinear programming with an application to Dobrushin curves

Stefan Huber, Robert König, and Marco Tomamichel

We propose a branch-and-bound algorithm for minimizing a bilinear functional of the form

$$f(X,Y)=\mathrm{tr}((X\otimes Y)Q)+\mathrm{tr}(AX)+\mathrm{tr}(BY),$$

of pairs of Hermitian matrices (X,Y) restricted by joint semidefinite programming constraints. The functional is parametrized by self-adjoint matrices Q, A and B. This problem generalizes that of a bilinear program, where X and Y belong to polyhedra. The algorithm converges to a global optimum and yields upper and lower bounds on its value in every step. Various problems in quantum information theory can be expressed in this form. As an example application, we compute Dobrushin curves of quantum channels, giving upper bounds on classical coding with energy constraints.

B.2.1 Setting

Consider the following optimization problem, which we call the *jointly constrained semidefinite bilinear program.* It is given by the minimization

$$\min_{(X,Y)\in\mathcal{S}}\mathrm{tr}\left(\left(X\otimes Y\right)Q\right)+\mathrm{tr}\left(AX\right)+\mathrm{tr}\left(BY\right)=:\min_{(X,Y)\in\mathcal{S}}F(X,Y)\ , \tag{B.1}$$

specified by a subset $\mathcal{S}\subset\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)\times\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ which is defined by a family of SDP constraints and self-adjoint operators $Q\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p\otimes\mathbb{C}^q)$, $A\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $B\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$. Optimization problems of the form (B.1) commonly appear in quantum information theory in various contexts.

A popular heuristic approach to optimization problems of the form (B.1) is the so-called seesaw algorithm, which first fixes a value of $X=X_0$ and then uses an SDP to minimize the affine-linear function $F_{X_0}(Y):=\mathrm{tr}((X_0\otimes Y)Q)+\mathrm{tr}(AX_0)+\mathrm{tr}(BY)$ over the set $\mathcal{S}_{X_0}:=\{Y\in\mathcal{B}(\mathbb{C}^q)|(X_0,Y)\in\mathcal{S}\}$. Then, the roles of X and Y are interchanged and the procedure is iterated. This algorithm can be shown to converge to a Kuhn-Tucker point of the objective function. It is in general not the case that this point will be a local (let alone global) optimum of the minimization problem. Contrary to other optimization problems such as the bilinear program for a pair of vectors (x,y) for which a wide variety of algorithms have been proposed [16,89–94], the seesaw algorithm appears to be the only procedure for (B.1) which has been used in the context of quantum information. Our work aims to provide a new approach to jointly constrained semidefinite bilinear programming in the quantum information context, which, contrary to the seesaw algorithm, provides upper and lower bounds on the optimal value of (B.1).

B.2.2 Main Results

A new branch-and-bound algorithm. Our main contribution is a branch-and-bound algorithm for the problem (B.1). The algorithm takes as input a set $\mathcal{S}\subset\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)\times\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ defined by SDP constraints, operators $Q\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p\otimes\mathbb{C}^q)$, $A\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$, $B\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ and a desired precision $\epsilon>0$. It returns a value $\overline{\alpha}$ and an element $(X^*,Y^*)\in\mathcal{S}$ such that

$$\overline{\alpha}=F(X^*,Y^*)\leq\left(\inf_{(X,Y)\in\mathcal{S}}F(X,Y)\right)+\epsilon\ . \tag{B.2}$$

The algorithm is an adaptation of a branch-and-bound algorithm given by Al-Khayyal and Falk [16] for a biconvex problem. The main difference between the latter and our algorithm is that we use SDPs to compute lower bounds on the objective function, whereas the algorithm by Al-Khayyal and Falk uses convex solvers. It can be shown that our algorithm always outputs a tuple (X^*, Y^*) which satisfies the termination condition (B.2). In fact, the algorithm produces a sequence of feasible points $(X_i, Y_i) \in \mathcal{S}$ such that $F(X_i, Y_i)$ converges to the global optimum (B.1). Furthermore, at each stage i, it provides a bound on the deviation of $F(X_i, Y_i)$ from the optimal value.

B.2.3 An application: Dobrushin curves of quantum channels

Let Φ be a channel and consider a setting where a message W is sent through a cascade of n copies of this channel, with a relay $\mathcal{E}_j$ applied before the j-th application of Φ. We are interested in the amount of information the output

$$Y_n = \Phi \circ \mathcal{E}_n \circ \cdots \circ \mathcal{E}_2 \circ \Phi \circ \mathcal{E}_1(W)$$

provides about the input W for an optimal coding strategy. Dobrushin curves in the classical setting have been introduced in [95] to study this problem in the setting where a power constraint is imposed on each of the outputs of the relays $\mathcal{E}_j$. The Dobrushin curve of a classical channel $P_{Y|X}$ with continuous variable input X (i.e., a random variable on $\mathbb{R}^m$) is defined as

$$F_E(\delta) = \sup_{\substack{P_{X_0}, P_{X_1} \in \mathcal{G}_E \\ \|P_{X_0} - P_{X_1}\|_1 \leq \delta}} \|P_{Y|X} \circ P_{X_0} - P_{Y|X} \circ P_{X_1}\|_1 \qquad \text{for } \delta \in [0,2] \ ,$$

where $\mathcal{G}_E$ is the set of distributions P_X on $\mathbb{R}^m$ which satisfy the power constraint $\mathbb{E}[\|X_j\|_2^2] \leq E$ for all $j \in \{1, \ldots m\}$ for some constant $E > 0$. The Dobrushin curve then gives an upper bound on the distance between outputs for different inputs of the channel:

$$\|P_{Y_n|W=0} - P_{Y_n|W=1}\|_1 \leq F_E^{\circ n}(2) \ ,$$

where $F_E^{\circ n} = F_E \circ \cdots \circ F_E$ is the n-fold composition of F_E. This holds independently of the choice of encoding maps $\mathcal{E}_j$.

In the quantum scenario, consider a quantum channel $\Phi : \mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{Y})$. A power constraint is introduced by fixing a Hamiltonian H on $\mathcal{X}$ and requiring that the expected energy is smaller than a constant $E \in \mathbb{R}$. The set of states satisfying this energy constraint is then given by

$$\mathcal{G}_E = \{\rho \in \mathcal{B}(\mathcal{X}) \bigm| \rho \geq 0 \ , \ \operatorname{tr} \rho = 1 \ \text{ and } \ \operatorname{tr}(\rho H) \leq E\} \ ,$$

and the Dobrushin curve for the quantum channel Φ can be defined as

$$F_E(\delta) = \sup_{\substack{\rho_0, \rho_1 \in \mathcal{G}_E \\ \|\rho_0 - \rho_1\|_1 \leq \delta}} \|\Phi(\rho_0) - \Phi(\rho_1)\|_1 \qquad \text{for } \delta \in [0,2] \ .$$

In analogy to the classical setting, the Dobrushin curve then gives upper bounds on the distinguishability at the output of the channel for different inputs:

$$\|\mathcal{F}_n(\rho_0) - \mathcal{F}_n(\rho_1)\|_1 \leq F_E^{\circ n}(2) \ ,$$

where $\mathcal{F}_n = \Phi \circ \mathcal{E}_n \circ \cdots \circ \mathcal{E}_2 \circ \Phi \circ \mathcal{E}_1$ for any choice of relays $\{\mathcal{E}_j\}_j$ with the property that the output of any state belongs to the energy-constrained set $\mathcal{G}_E$. The function $F_E(\delta)$ can be cast into an optimization problem of the form (B.1). Hence our algorithm gives us a tool to calculate Dobrushin curves of quantum channels. As an example, we numerically calculate the Dobrushin curves of dephasing channels as well as some more generic qubit channels. We numerically compute Dobrushin curves for dephasing channels of the form $\Phi\left(\alpha_0 I + \sum_{k=1}^{3} \alpha_k \sigma_k\right) = \alpha_0 I + a\left(\alpha_1 \sigma_1 + \alpha_2 \sigma_2\right) + \alpha_3 \sigma_3$. We note that for this channel, an analytical expression for the Dobrushin curve can be found with heuristic arguments. Our algorithm shows that this curve obtained by heuristic arguments is indeed the Dobrushin curve.

Jointly constrained semidefinite bilinear programming with an application to Dobrushin curves

Stefan Huber, Robert König and Marco Tomamichel

August 10, 2018

Abstract

We propose a branch-and-bound algorithm for minimizing a bilinear functional of the form

$$f(X,Y)=\operatorname{tr}((X\otimes Y)Q)+\operatorname{tr}(AX)+\operatorname{tr}(BY),$$

of pairs of Hermitian matrices (X,Y) restricted by joint semidefinite programming constraints. The functional is parametrized by self-adjoint matrices Q, A and B. This problem generalizes that of a bilinear program, where X and Y belong to polyhedra. The algorithm converges to a global optimum and yields upper and lower bounds on its value in every step. Various problems in quantum information theory can be expressed in this form. As an example application, we compute Dobrushin curves of quantum channels, giving upper bounds on classical coding with energy constraints.

1 Introduction

The bilinear program of the form

$$\min_{(x,y)\in X\times Y} x^TQy+v^Tx+w^Ty, \tag{1}$$

where $X\subset\mathbb{R}^p$ and $Y\subset\mathbb{R}^q$ are polyhedra and $v\in\mathbb{R}^p$, $w\in\mathbb{R}^q$, $Q\in\mathbb{R}^{p\times q}$, is among the most well-studied optimization problems. One of its first appearances is in the formulation of certain two nonzero-sum games studied by Nash [15]. The optimization problem (1) has various applications in operations research and information theory, including network flow problems, dynamic Markovian assignment problems, and dynamic production problems — see [11] and [7] for a discussion of a number of these. Several natural generalizations of the problem (1) exist. In particular, a *biconvex problem* is of the form $\min_{(x,y)\in\mathcal{S}} f(x,y)$ where $\mathcal{S}$ and f are biconvex, i.e., $\mathcal{S}_{x_0}=\{y\in\mathbb{R}^q \mid (x_0,y)\in\mathcal{S}\}$ and $f(x_0,\cdot)$ are convex for every $x_0\in\mathbb{R}^p$, and similarly for $y_0\in\mathbb{R}^q$. We refer to [8] for a review of biconvex problems (see also [1]).

Here we consider a different generalization of (1) which pertains to problems in quantum information theory. We refer to it as *jointly constrained semidefinite bilinear programming*. In this generalization, the vectors $x\in\mathbb{R}^p$ and $y\in\mathbb{R}^q$ are replaced by self-adjoint operators $X\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $Y\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ satisfying certain semidefinite programming (SDP) constraints. The bilinear form of the objective function is retained, leading to

$$\min_{(X,Y)\in\mathcal{S}} \operatorname{tr}((X\otimes Y)Q)+\operatorname{tr}(AX)+\operatorname{tr}(BY)\ . \tag{2}$$

The problem (2) is thus fully specified by a subset $\mathcal{S}\subset\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)\times\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ of pairs (X,Y) defined by a family of SDP constraints, as well as self-adjoint operators $Q\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p\otimes\mathbb{C}^q)$, $A\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $B\in\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$.

The problem (2) appears in various forms throughout quantum information theory [1]. For example, in entanglement distribution and quantum communication, one seeks to generate entanglement between a reference system R and a system S transmitted through a noisy channel modeled by a completely positive trace-preserving (CPTP) map $\mathcal{N} : \mathcal{B}(\mathcal{H}_A) \to \mathcal{B}(\mathcal{H}_B)$. A key figure of merit in this context is the entanglement fidelity [20]

$$\max_{(\mathcal{E},\mathcal{D})} \langle\Psi|((\mathcal{D} \circ \mathcal{N} \circ \mathcal{E} \otimes \mathsf{id}_R)(|\Psi\rangle\langle\Psi|))|\Psi\rangle \tag{3}$$

where the optimization is over all encoding CPTP maps $\mathcal{E} : \mathcal{B}(\mathcal{H}_S) \to \mathcal{B}(\mathcal{H}_A)$ and decoding CPTP maps $\mathcal{D} : \mathcal{B}(\mathcal{H}_B) \to \mathcal{B}(\mathcal{H}_S)$, and where $|\Psi\rangle \in \mathcal{H}_S \otimes \mathcal{H}_R$ is a fixed (maximally entangled) state of the joint system SR. Eq. (3) can be cast in the form (2): The set of CPTP maps $\mathcal{E} : \mathcal{B}(\mathcal{H}_S) \to \mathcal{B}(\mathcal{H}_A)$ can be described by SDP constraints via the Choi-Jamiolkowski isomorphism (and analogously for $\mathcal{D}$), and the objective function is bilinear in $(\mathcal{E}, \mathcal{D})$.

Another context in which the problem (2) appears naturally is the setting of Bell inequalities and quantum games. Consider for example the bipartite case, where a state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ is given. A Bell inequality can be expressed as a lower bound on the expectation value $\langle\Psi|B|\Psi\rangle$ of a Bell operator $B = B(\{E_j\}_j, \{F_k\}_k) \in \mathcal{B}_{\mathsf{sa}}(\mathcal{H}_A \otimes \mathcal{H}_B)$ which depends on observables $\{E_j\}_j$ on $\mathcal{H}_A$ (corresponding to A's measurement settings) and observables $\{F_k\}_k$ on $\mathcal{H}_B$ (corresponding to B's measurement settings). The Bell operator usually depends bilinearly on $(\{E_j\}_j, \{F_k\}_k)$: for example, the Bell-CHSH operator involves two measurement settings each and is given by $B(E_0, E_1, F_0, F_1) = E_0 \otimes (F_0 + F_1) + E_1 \otimes (F_0 - F_1)$. Since an observable A is a self-adjoint operator satisfying $-I \leq A \leq I$, the problem of finding the optimal value

$$\max_{\{E_j\}_j, \{F_k\}_k} \langle\Psi|B(\{E_j\}_j, \{F_k\}_k)|\Psi\rangle \tag{4}$$

optimized over all observables can directly be cast in the form (2).

We note that the form of (2) is somewhat more general than what is required in most applications to quantum information theory such as problems (3) and (4). Indeed, in the latter two problems, there are no joint constraints (i.e., the set $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2$ is a product of two sets, each of which is defined by SDP constraints), and the objective function has no linear terms. In Section 4, we discuss a problem from quantum information theory whose reformulation in terms of (2) involves linear terms.

A useful alternative but equivalent form of the optimization problem is given by

$$\min_{(X,Y)\in\mathcal{S}} \operatorname{tr}(X\mathcal{E}(Y)) + \operatorname{tr}(AX) + \operatorname{tr}(BY)$$

where $\mathcal{E}$ is a Hermiticity-preserving operation. The one-to-one correspondence of $\mathcal{E}$ and Q is a consequence of the Choi-Jamiolkowski isomorphism.

Finally we note that joint constraints allow to consider quadratic optimization problems of the form

$$\min_{X\in\mathcal{S}} \operatorname{tr}((X \otimes X)Q) + \operatorname{tr}(AX)\ , \tag{5}$$

where $\mathcal{S}$ is a set defined by SDP constraints, simply by imposing that $X = Y$. A basic example is a situation where one is interested in the maximal (or minimal) uncertainty when measuring a state $|\Psi\rangle$, using an observable $X \in \mathcal{S}$ belonging to a set $\mathcal{S}$ specified by SDP constraints. If uncertainty is quantified by the variance $\mathsf{Var}_\Psi(X) := \langle\Psi|X^2|\Psi\rangle - \langle\Psi|X|\Psi\rangle^2$, then the problem $\min_{X\in\mathcal{S}} \mathsf{Var}_\Psi(X)$ can be recast in the form (5).

[1] This problem is not to be confused with the recently introduced quantum bilinear programs [3].

The seesaw algorithm in quantum information

Given the ubiquity of optimization problems of the form (2) in quantum information theory, it is natural to seek algorithms computing its value as well as optimal solutions (X, Y). A widely used and often successful heuristic is referred to as the seesaw or mountain climbing algorithm. It is based on the observation that for every X_0 (associated with a feasible point $(X_0, Y_0) \in \mathcal{S}$), the function $f_{X_0}(Y) = \mathrm{tr}((X_0 \otimes Y)Q) + \mathrm{tr}(AX_0) + \mathrm{tr}(BY)$ is linear up to the additive constant $\mathrm{tr}(AX_0)$. Furthermore, the set $\mathcal{S}_{X_0} = \{Y \in \mathcal{B}(\mathbb{C}^q) \mid (X_0, Y) \in \mathcal{S}\}$ can be described by SDP constraints (indeed, we can augment those specifying $\mathcal{S}$ by the constraint $X = X_0$). Thus, $Y_0 := \arg\min_{Y \in \mathcal{S}_{X_0}} f_{X_0}(Y)$ can be found by solving an SDP. The role of X and Y is then interchanged: in a next step, $X_1 := \arg\min_{Y \in \mathcal{S}_{Y_0}} f_{Y_0}(X)$ is computed (where $\mathcal{S}_{Y_0}$ and f_{Y_0} are defined analogously). Iterating this produces a sequence of pairs (X_j, Y_j). It can be shown that in a finite number of iterations, this sequence converges to a Kuhn-Tucker point $(\bar{X}, \bar{Y})$ of the objective function

$$f(X, Y) = \mathrm{tr}((X \otimes Y)Q) + \mathrm{tr}(VX) + \mathrm{tr}(WY) ,$$

see [12, Prop. 2.3] for an analysis of the analogous algorithm for bilinear programs, and [14, Theorem 5] as well as [10] for conditions guaranteeing that this point is a local optimimum. Thus, while it is not generally the case that $(\bar{X}, \bar{Y})$ is a local (let alone global) optimum, this algorithm may — for a suitable choice of initial points (X_0, Y_0) indeed result in a global solution. It should be emphasized, however, that even in that case, global optimality needs to be established by other means.

Despite these limitations, the seesaw algorithm has been quite popular and has been successfully applied in quantum information theory. Its use in the context of Bell inequalities was first discussed in [25, Section 5.1]. In the context of error correction, the maximization of fidelity optimized over encoder and decoder has been investigated numerically using the seesaw algorithm, see [19] and [13]. More recently, a variant of the seesaw algorithm (involving a trilinear function) was used in [24] to optimize the value of a Bell inequality over PPT-states, yielding a counterexample to Peres' conjecture [17] that non-distillable states are local.

To date, the seesaw algorithm appears to be the only procedure for optimization problems of the form (2) which has been used in the context of quantum information. This is in sharp contrast to the bilinear program (1), for which a variety of algorithms have been proposed. This includes cutting plane algorithms [12, 23, 21], branch-and-bound algorithms [6, 2], extreme point ranking procedures [4] and methods based on polyhedral annexation [22] (see [7] for a review).

Our contribution

Our main contribution is a branch-and-bound algorithm for the jointly constrained semidefinite bilinear program (2). It is a generalization of the branch-and-bound algorithm by Al-Khayyal and Falk [2] which we review in Section 2.1. Roughly, our algorithm proceeds by iteratively solving semidefinite programs providing upper and lower bounds on the value of (2). Following standard arguments (see e.g., [2]), it can be shown to produce a sequence of feasible points $(X_i, Y_i) \in \mathcal{S}$ such that $f(X_i, Y_i)$ converges to the global optimum (2). More importantly, it provides — at each stage i — a bound on the deviation of $f(X_i, Y_i)$ from the optimum (2).

To illustrate the practical use of our algorithm, we apply it to a problem in quantum information theory: we compute so-called Dobrushin curves for quantum channels. These give upper bounds on optimal codes for classical information in a scenario where the noise acts repeatedly.

Outline of the paper

In Section 2, we briefly review branch-and-bound algorithms and discuss the algorithm by Al-Khayyal and Falk [2] for solving jointly constrained biconvex programs. In Section 3, we give our algorithm for jointly constrained semidefinite bilinear programs. Finally, in Section 4, we discuss the application to Dobrushin curves.

2 Branch-and-bound algorithms

In this section, we review the branch-and-bound algorithm of Al-Khayyal and Falk [2] to solve jointly constrained biconvex programs. We introduce the jointly constrained biconvex programming problem, and then give a description of the algorithm of [2].

2.1 Jointly constrained biconvex programming

To define jointly constrained biconvex programs, let $\mathcal{S} \subset \mathbb{R}^n \times \mathbb{R}^n$ be a non-empty, closed and convex set. For later convenience, also let $\mathcal{D} = \Omega(\ell, L, m, M) \subset \mathbb{R}^n \times \mathbb{R}^n$ be the (product of) hyperrectangle(s) defined in terms of the vectors $\ell, L, m, M \in \mathbb{R}^n$ as

$$\Omega(\ell, L, m, M) = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \big| \ \ell_i \leq x_i \leq L_i, m_i \leq y_i \leq M_i \text{ for all } i = 1, \ldots, n\} \ . \tag{6}$$

Furthermore, let $f, g : \mathcal{D} \to \mathbb{R}$ be such that their restrictions to $\mathcal{S} \cap \mathcal{D}$ are convex. The jointly constrained biconvex program is the problem

$$\min_{(x,y)\in\mathcal{S}\cap\mathcal{D}} F(x, y) \qquad \text{where} \qquad F(x, y) := f(x) + x^T y + g(y) \ . \tag{7}$$

The set $\mathcal{S}$ permits to include joint constraints on the vectors x and y. We note that although the restrictions $F(\cdot, y)$ and $F(x, \cdot)$ are convex for each $(x, y) \in \mathcal{S} \cap \mathcal{D}$ — a property referred to as biconvexity — the problem Eq. (7) itself is non-convex.

Eq. (7) is a generalization of the bilinear program (1) discussed in the introduction. Indeed, Eq. (1) can be transformed into a problem of the form (7) by replacing $x^T(Qy)$ by $x^T z$ and adding the linear constraint $z = Qy$ to the defining constraints of $\mathcal{S}$.

2.1.1 Obtaining lower bounds on the biconvex program

Being non-convex, Eq. (7) cannot directly be addressed with convex solvers. However, one can construct a convex problem whose solution gives a lower bound on the value of Eq. (7). This relies on the concept of the convex envelope $\mathsf{Vex}_{\mathcal{D}}F : \mathcal{D} \to \mathbb{R}$ of a function $F : \mathcal{D} \to \mathbb{R}$, where $\mathcal{D} \subset \mathbb{R}^n \times \mathbb{R}^n$. It is defined as the pointwise supremum of all convex functions underestimating F over $\mathcal{D}$, i.e.,

$$(\mathsf{Vex}_{\mathcal{D}}F)(x, y) := \sup_{\substack{G:\mathcal{D}\to\mathbb{R} \text{ convex} \\ G(v,w)\leq F(v,w) \text{ for all } (v,w)\in\mathcal{D}}} G(x, y) \qquad \text{for all } (x, y) \in \mathcal{D} \ ,$$

see [5] for more details.

To compute the convex envelope $\mathsf{Vex}_{\mathcal{S}\cap\mathcal{D}}F$ of the objective function in Eq. (7), where $\mathcal{D} = \Omega$ is a hyperrectangle, one uses the fact that the convex envelope of the function $(x, y) \mapsto x^T y$ over a hyperrectangle Ω (cf. (6)) is (see [2, Corollary to Theorems 2 and 3])

$$\mathsf{Vex}_{\Omega}(x^T y) = \sum_{i=1}^{n} \mathsf{Vex}_{\Omega_i}(x_i y_i) \qquad \text{for all } (x, y) \in \Omega \ ,$$

where

$$\Omega_i := \{(x, y) \in \mathbb{R} \times \mathbb{R} : \ell_i \leq x \leq L_i, m_i \leq y \leq M_i\} = [\ell_i, L_i] \times [m_i \times M_i] \subset \mathbb{R} \times \mathbb{R}$$

is the projection of the hyperrectangle Ω onto the i-th pair of coordinates for $i = 1, \ldots, n$. The convex envelope of the function $(x, y) \mapsto xy$ over $\Omega_i \subset \mathbb{R} \times \mathbb{R}$ is (see [2, Theorem 2])

$$\mathsf{Vex}_{\Omega_i}(xy) = \max\{m_i x + \ell_i y - \ell_i m_i, M_i x + L_i y_i - L_i M_i\} \ .$$

Hence the convex envelope of F in Eq. (7) over a hyperrectangle Ω has a simple expression, i.e.

$$(\mathsf{Vex}_\Omega F)(x, y) = f(x) + \sum_{i=1}^{n} \mathsf{Vex}_{\Omega_i}(x_i y_i) + g(y) \ . \tag{8}$$

In addition to being a convex underestimator for F, the function $\mathsf{Vex}_\Omega F$ has the important property that it agrees with F on the boundary

$$\partial\Omega = \Omega \setminus \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^n \bigm| \ell_i < x_i < L_i, m_i < y_i < M_i \text{ for all } i = 1, \ldots, n\}$$

of Ω. Indeed, this follows from the analogous property $\mathsf{Vex}_\Omega(x^T y) = x^T y$ for all $(x, y) \in \partial\Omega$ for the function $(x, y) \mapsto x^T y$, see [2, Theorem 3].

Note that, given (8), the problem $\min_{(x,y)\in\mathcal{S}\cap\Omega}(\mathsf{Vex}_\Omega F)(x, y)$ can be treated using a convex solver, giving a value $\underline{\alpha}$. In particular, since $\mathsf{Vex}_\Omega F$ underestimates F over Ω (and hence over $\mathcal{S} \cap \Omega$), the value $\underline{\alpha}(\Omega)$ provides a (global) lower bound on the problem (7). Trivially, any point $(x^*, y^*) \in \mathcal{S} \cap \Omega$ (including one that achieves the minimum of $\mathsf{Vex}_\Omega F$) also provides an upper bound $\overline{\alpha}(\Omega) = F(x^*, y^*)$ on (7). Thus, we obtain

$$\underline{\alpha}(\Omega) \leq \min_{(x,y)\in\mathcal{S}\cap\Omega} F(x, y) \leq \overline{\alpha}(\Omega) \ .$$

We note that the reasoning here applies to any hyperrectangle Ω.

2.1.2 The branch-and-bound algorithm for the biconvex program

We can now sketch the algorithm of [2] which solves problem (7). The algorithm takes as input the functions f and g, the set $\mathcal{S} \subset \mathbb{R}^n \times \mathbb{R}^n$ and the vectors $\ell, L, m, M \in \mathbb{R}^n$ specifying the hyperrectangle $\Omega = \Omega(\ell, L, m, M)$. In addition, it accepts a desired precision $\epsilon > 0$. The algorithm returns a value $\overline{\alpha}$ and a point $(x^*, y^*) \in \mathcal{S} \cap \Omega$ such that

$$\overline{\alpha} = F(x^*, y^*) \leq \left(\min_{(x,y)\in\mathcal{S}\cap\Omega} F(x, y)\right) + \epsilon \ .$$

Given the technique for finding lower bounds on $\min_{(x,y)\in\mathcal{S}\cap\mathcal{D}} F(x, y)$ discussed in Section 2.1.1, the main idea underlying the algorithm is to apply this strategy to increasingly smaller hyperrectangle Ω (which together form a partition of D). The respective upper and lower bounds for each hyperrectangle give (global) upper and lower bounds on the biconvex problem (7).

More precisely, the algorithm keeps track of a finite list $\mathcal{P}$ of hyperrectangles which form a partition of $\mathcal{D}$. In addition, for each $\Omega \in \mathcal{P}$, the values $(\underline{\alpha}(\Omega), \overline{\alpha}(\Omega))$ are computed and kept track of such that

$$\underline{\alpha}(\Omega) \leq \min_{(x,y)\in\mathcal{S}\cap\Omega} F(x, y) \leq \overline{\alpha}(\Omega) \ .$$

Finally, $z(\Omega) = (x, y) \in \mathcal{S} \cap \Omega$ will be an element such that $F(z) = \overline{\alpha}(\Omega)$.

As a consequence, the quantities

$$\underline{\alpha}(\mathcal{P}) = \min_{\Omega\in\mathcal{P}} \underline{\alpha}(\Omega) \qquad \text{and} \qquad \overline{\alpha}(\mathcal{P}) = \min_{\Omega\in\mathcal{P}} \overline{\alpha}(\Omega)$$

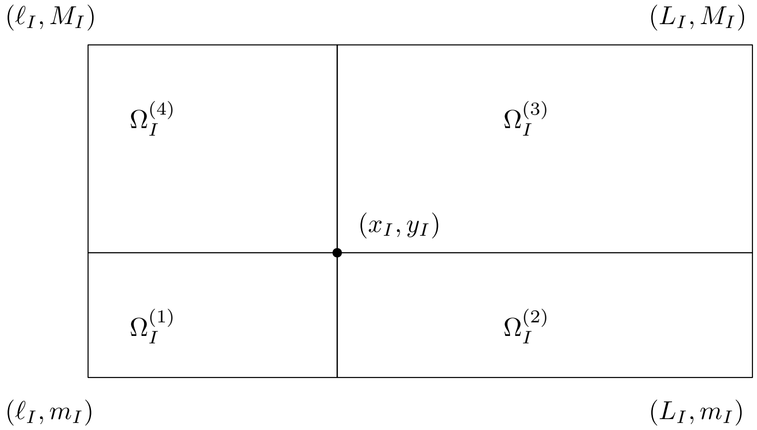


Figure 1: The creation of hyperrectangles $\Omega^{(1)}, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)}$ from Ω. This shows the projection onto the coordinates (x_I, y_I) of these hyperrectangles.

constitute global upper and lower bounds on the problem Eq. (7), that is,

$$\underline{\alpha}(\mathcal{P}) \leq \min_{(x,y)\in\mathcal{S}\cap\Omega} F(x,y) \leq \overline{\alpha}(\mathcal{P}) \ .$$

As soon as $\overline{\alpha}(\mathcal{P}) - \underline{\alpha}(\mathcal{P}) \leq \epsilon$, the algorithm returns $\overline{\alpha}(\mathcal{P})$ and $(x^*, y^*) = z(\Omega')$, where $\Omega' \in \mathcal{P}$ is such that $F(z(\Omega')) = \overline{\alpha}(\mathcal{P})$.

Recall that $\mathcal{D}$ itself is a hyperrectangle by definition of the problem. Consequently, we begin with the (trivial) partition $\mathcal{P} = \{\mathcal{D}\}$. Since an algorithm for computing bounds $(\underline{\alpha}(\Omega), \overline{\alpha}(\Omega))$ and points $z(\Omega) \in \mathcal{S}\cap\Omega$ for any hyperrectangle Ω is already constructed, it remains to specify how $\mathcal{P}$ is successively refined.

Assume that the algorithm has not returned a solution yet, i.e., that

$$\overline{\alpha}(\mathcal{P}) - \underline{\alpha}(\mathcal{P}) \geq \epsilon \ . \tag{9}$$

The idea here is to try to improve the worst lower bound. That is, pick a hyperrectangle $\Omega \in \mathcal{P}$ such that $\underline{\alpha}(\mathcal{P}) = \underline{\alpha}(\Omega)$. We then subdivide Ω in 4 new hyperrectangles $\Omega^{(1)}, \ldots, \Omega^{(4)}$. To do so, observe that Eq. (9) implies that $\overline{\alpha}(\Omega) - \underline{\alpha}(\Omega) \geq \epsilon$. Hence, by definition of $\overline{\alpha}(\Omega)$ and $\underline{\alpha}(\Omega)$, there must exist at least one $i \in \{1, \ldots, n\}$ such that $\mathsf{Vex}_\Omega(x_i y_i) < x_i y_i$. We pick the index I which leads to the largest difference between the two sides of this inequality and split up the rectangle Ω into four subrectangles, arriving at the new hyperrectangles $\{\Omega^{(j)}\}_{j=1}^4$. For each $j \in \{1, \ldots, 4\}$, the hyperrectangle $\Omega^{(j)}$ is defined by its projections

$$\Omega_i^{(j)} = \begin{cases} \Omega_i & \text{if } i \in \{1, \ldots, n\} \backslash \{I\} \\ A^{(j)} & \text{for } i = I \ . \end{cases}$$

onto pairs of coordinates. Here $\{A^{(j)}\}_{j=1}^4$ is a certain partition of $\Omega_I \subset \mathbb{R} \times \mathbb{R}$ into four subrectangles, as shown in Fig. 1. The latter is defined by the pair of I-th coordinates (x_I, y_I) of the point $z(\Omega) = (x, y)$, as shown in Fig. 1. Hence we have constructed a partition $\{\Omega^{(j)}\}_{j=1}^4$ of Ω into smaller hyperrectangles.

These steps are iterated until Eq. (9) is no longer satisfied. This procedure can be shown to converge to a globally optimal value of the problem (7), as done in [2].

3 Jointly constrained semidefinite bilinear programming

Suppose self-adjoint operators $Q \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p \otimes \mathbb{C}^q), A \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q), B \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ are given. Here $\mathcal{B}_{\mathsf{sa}}(\mathcal{H})$ denotes the real vector space of self-adjoint operators on a Hilbert space $\mathcal{H}$ with respect to the Hilbert-Schmidt inner product $\langle A, B\rangle = \mathrm{tr}(AB)$. Define a function $F : \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q) \to \mathbb{R}$ by

$$F(X, Y) := \mathrm{tr}\left((X \otimes Y)Q\right) + \mathrm{tr}(AX) + \mathrm{tr}(BY) \ . \tag{10}$$

We consider the problem

$$\inf_{(X,Y)\in\mathcal{S}} F(X, Y) \ , \tag{11}$$

where $\mathcal{S} \subset \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ is defined by a family of semidefinite constraints, which may involve both X and Y (in particular, $\mathcal{S}$ is convex). We note that the function F is again biconvex but not convex. We refer to Eq. (11) as a jointly constrained semidefinite bilinear program.

A first step to construct an algorithm for (11) is to rephrase it in a form similar to (7). To do so, let $\{\eta_j\}_{j=1}^{p^2} \subset \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $\{\xi_k\}_{k=1}^{q^2} \subset \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ be orthonormal operator bases of the real vector spaces $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$, respectively. It will be convenient to express operators in terms of coefficients in bases that are rotated with respect to $\{\eta_j\}_{j=1}^{p^2}$ and $\{\xi_k\}_{k=1}^{q^2}$, with a rotation depending on the objective function. Consider the $p^2 \times q^2$-matrix $U_{j,k} = \mathrm{tr}(Q(\eta_j \otimes \xi_k))$ and let

$$U = S\Delta T \qquad \text{where } \Delta \in \mathbb{R}^{p^2\times q^2} \tag{12}$$

be its singular value decomposition, i.e., $S \in \mathbb{R}^{p^2\times p^2}$ and $T \in \mathbb{R}^{q^2\times q^2}$ are orthogonal, and Δ has the singular values $\{\sigma_j\}_{j=1}^K$ on the diagonal (here $K \leq \min\{p^2, q^2\}$). Define the map

$$\begin{aligned} \Gamma : \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q) &\to \mathbb{R}^{p^2} \times \mathbb{R}^{q^2} \\ (X, Y) &\mapsto \Gamma(X, Y) = (x(X), y(Y)) \ , \end{aligned}$$

where

$$x_j(X) = \sum_{k=1}^{p^2} S_{k,j}\,\mathrm{tr}(X\eta_k) \qquad \text{and} \qquad y_k(Y) = \sum_{\ell=1}^{q^2} T_{k,\ell}\,\mathrm{tr}(Y\xi_\ell) \ .$$

Let $(a, b) = (x(A), y(B)) \in \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$. Define the function $f : \mathbb{R}^{p^2} \times \mathbb{R}^{q^2} \to \mathbb{R}$ by

$$f(x, y) := \sum_{j=1}^{K} \sigma_j x_j y_j + \sum_{j=1}^{p^2} a_j x_j + \sum_{k=1}^{q^2} b_k y_k \ . \tag{13}$$

Using this construction we can now reduce the matrix problem to an equivalent vector problem:

Lemma 3.1. *Γ is one-to-one and $F(X, Y) = f(x(X), y(Y))$ for all $(X, Y) \in \mathcal{S}$. In particular,*

$$\inf_{(X,Y)\in S} F(X, Y) = \inf_{(x,y)\in\Gamma(\mathcal{S})} f(x, y) \ .$$

Proof. Observe that for (X,Y), we have

$$F(X,Y)=\sum_{j=1}^{p^2}\sum_{k=1}^{q^2}\hat{x}_j\hat{y}_kU_{j,k}+\sum_{j=1}^{p^2}\hat{x}_j\hat{a}_j+\sum_{k=1}^{q^2}\hat{y}_k\hat{b}_k\ ,$$

where $\hat{x}_j = \mathrm{tr}(X\eta_j)$, $\hat{a}_j = \mathrm{tr}(A\eta_j)$ for $j = 1,\ldots,p^2$, and similarly $\hat{y}_k = \mathrm{tr}(Y\xi_k)$, $\hat{b}_k = \mathrm{tr}(B\xi_k)$ for $k = 1,\ldots,q^2$. Using the variable substitutions

$$\begin{array}{rcl} x & = & S^T\hat{x} \\ y & = & T\hat{y} \end{array} \quad \text{and} \quad \begin{array}{rcl} a & = & S^T\hat{a} \\ b & = & T\hat{b}\ , \end{array}$$

the claim follows. □

Given Lemma 3.1, our algorithm proceeds by first finding a hyperrectangle $\mathcal{D}\subset\mathbb{R}^{p^2}\times\mathbb{R}^{q^2}$ that contains the set $\Gamma(\mathcal{S})$ (see Section 3.1). We then argue that lower bounds on the objective function restricted to hyperrectangles can be computed by solving SDPs (see Section 3.2). A branch-and-bound procedure for the problem (11) follows.

For later reference, we give pseudocode of two routines ComputeVectorRep and ComputeOperator, which compute the functions Γ respectively Γ^{-1} appearing in Lemma 3.1, see Fig. 8 in Appendix A.

3.1 Finding a bounding hyperrectangle

For $\ell, L\in\mathbb{R}^{p^2}$ and $m, M\in\mathbb{R}^{q^2}$, define the hyperrectangle

$$\Omega(\ell,L,m,M)=\{(x,y)\in\mathbb{R}^{p^2}\times\mathbb{R}^{q^2}\big|\ \ell_j\leq x_j\leq L_j \text{ for all } j=1,\ldots,p^2 \\ \text{and } m_k\leq y_k\leq M_k \text{ for all } k=1,\ldots,q^2\}\ .$$

We show the following:

Lemma 3.2. *We can efficiently find $\ell, L\in\mathbb{R}^{p^2}$ and $m, M\in\mathbb{R}^{q^2}$ such that $\Omega(\ell,L,m,M)$ has minimal volume among all hyperrectangles Ω containing the set $\Gamma(\mathcal{S})$, where Γ is defined as in Lemma 3.1. More precisely, we can find such vectors by solving $2(p^2+q^2)$ SDPs in $(X,Y)\in\mathcal{S}$.*

Proof. Clearly, we need to compute

$$\begin{array}{rclccccc} \ell_j^* & = & \inf_{(X,Y)\in\mathcal{S}} x_j(X) & \text{and} & L_j^* & = & \sup_{(X,Y)\in\mathcal{S}} x_j(Y) & \text{for} \quad j=1,\ldots,p^2 \\ & & & & \text{as well as} & & & \\ m_k^* & = & \inf_{(X,Y)\in\mathcal{S}} y_k(X) & \text{and} & M_k^* & = & \sup_{(X,Y)\in\mathcal{S}} y_k(Y) & \text{for} \quad k=1,\ldots,q^2\ . \end{array}$$

Here we write $\Gamma(X,Y)=(x(X),y(Y))$ as in Lemma 3.1. It is easy to see that each of these optimization problems is an SDP. For example, for each $j\in\{1,\ldots,p^2\}$, we have

$$\ell_j^*=\inf_{(X,Y)\in\mathcal{S}}\sum_{k=1}^{p^2}S_{k,j}\,\mathrm{tr}(X\eta_k)$$

and similar reasoning applies to the values L_j^*, m_k^* and M_k^*. □

Pseudocode for the associated procedure is given in Fig. 10 in Appendix A.

3.2 Obtaining lower bounds on the semidefinite bilinear program

As in Section 2.1.1, we next discuss how to find lower and upper bounds $\underline{\alpha}(\Omega)$, $\overline{\alpha}(\Omega)$ on the objective function $F(X,Y)$ restricted to the preimage $\Gamma^{-1}(\Omega)$ of a hyperrectangle $\Omega \subset \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$. That is, in terms of the function $f : \mathbb{R}^{p^2} \times \mathbb{R}^{q^2} \to \mathbb{R}$ defined in Lemma 3.1, these values satisfy

$$\underline{\alpha}(\Omega) \leq \inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} f(x,y) \leq \overline{\alpha}(\Omega) \ .$$

For the lower bound, recalling the definition of the convex envelope introduced in Section 2.1.2, it suffices to compute

$$\underline{\alpha}(\Omega) = \inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} (\mathsf{Vex}_\Omega f)(x,y) \ . \tag{14}$$

On the other hand, any element $(x^*, y^*) \in \Gamma(\mathcal{S}) \cap \Omega$ provides an upper bound $\overline{\alpha}(\Omega) = f(x^*, y^*)$.

To compute Eq. (14), we proceed in two steps. First, we give an explicit expression for $\mathsf{Vex}_\Omega f$.

Lemma 3.3. *Let $\Omega = \Omega(\ell, L, m, M) \subset \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$ be a hyperrectangle and $f : \mathbb{R}^{p^2} \times \mathbb{R}^{q^2} \to \mathbb{R}$ as in Lemma 3.1. Then the convex envelope of f over Ω is given by*

$$(\mathsf{Vex}_\Omega f)(x,y) = \sum_{j=1}^{K} \max\{h_j^0(x_j,y_j), h_j^1(x_j,y_j)\} + \sum_{j=1}^{p^2} a_j x_j + \sum_{k=1}^{q^2} b_k y_k \ ,$$

where

$$\begin{array}{rcll} h_j^0(x_j,y_j) &=& \sigma_j\,(m_j x_j + \ell_j y_j - \ell_j m_j) & \textit{and} \\ h_j^1(x_j,y_j) &=& \sigma_j\,(M_j x_j + L_j y_j - L_j M_j) \ . & \end{array} \tag{15}$$

for $j = 1, \ldots, K$.

Proof. By definition of f (see Lemma 3.1) and calculations analogous to those discussed in Section 2.1.1, we have

$$\begin{aligned} (\mathsf{Vex}_\Omega f)(x,y) &= \sum_{j=1}^{K} \mathsf{Vex}_{\Omega_j}(\sigma_j x_j y_j) + \sum_{j=1}^{p^2} a_j x_j + \sum_{k=1}^{q^2} b_k y_k \\ &= \sum_{j=1}^{K} \max\{\sigma_j \cdot (m_j x_j + \ell_j y_j - \ell_j m_j), \sigma_j \cdot (M_j x_j + L_j y_j - L_j M_j)\} + \sum_{j=1}^{p^2} a_j x_j + \sum_{k=1}^{q^2} b_k y_k \\ &= \sum_{j=1}^{K} \max\{h_j^0(x_j,y_j), h_j^1(x_j,y_j)\} + \sum_{j=1}^{p^2} a_j x_j + \sum_{k=1}^{q^2} b_k y_k \ , \end{aligned}$$

as claimed. □

In the following Lemma, we show that $\inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega}(\mathsf{Vex}_\Omega f)(x,y)$ can be expressed as an SDP. This provides an efficient way of computing the lower bound (14).

Lemma 3.4. *Let $\Omega \subset \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$ be a hyperrectangle and $\Gamma(\mathcal{S})$ be a set of vectors obtained from a set $\mathcal{S}$ of semidefinite constraints as described in Lemma 3.1. Furthermore, let $\Gamma(\mathcal{S}) \cap \Omega$ be nonempty. Then the problem*

$$\inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} (\mathsf{Vex}_\Omega f)(x,y) \tag{16}$$

is a semidefinite program in (X, Y, r), where $(X,Y) \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ and $r \in \mathbb{R}^K$.

Proof. Introduce the notation

$$\begin{aligned}
\hat{\ell}_j^0 &= \ell_j\ , & \hat{\ell}_j^1 &= L_j & &\text{for } j = 1,\ldots,p^2\ ,\\
\hat{m}_k^0 &= m_k\ , & \hat{m}_k^1 &= M_k & &\text{for } k = 1,\ldots,q^2\ ,
\end{aligned}$$

for the lower- and upper bounds determining the hyperrectangle $\Omega = \Omega(\ell, L, m, M)$. Then the functions h_j^0, h_j^1 introduced in Eq. (15) can be expressed as

$$h_j^b(x_j, y_j) = \sigma_j \cdot (\hat{m}_j^b x_j + \hat{\ell}_j^b y_j - \hat{\ell}_j^b \hat{m}_j^b) \qquad \text{for} \qquad b \in \{0,1\}\ .$$

We have by Lemma 3.3

$$\begin{aligned}
\inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} (\mathsf{Vex}_\Omega f)(x,y) &= \inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} \sum_{j=1}^{K} \max\{h_j^0(x_j,y_j), h_j^1(x_j,y_j)\} + \sum_{j=1}^{p^2} a_j x_j + \sum_{k=1}^{q^2} b_k y_k\\
&= \inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} \inf_{\substack{r_1,\ldots,r_K\in\mathbb{R}\\ h_j^b(x_j,y_j)\le r_j\\ \text{for } j=1,\ldots,K,\ b\in\{0,1\}}} \sum_{j=1}^{K} r_j + \sum_{j=1}^{p^2} a_j x_j + \sum_{j=k}^{q^2} b_k y_k\ . \qquad (17)
\end{aligned}$$

Here we have replaced each maximum by a semidefinite program in a scalar, that is, we have used the identity

$$\max\{h^0, h^1\} = \inf_{\substack{r\in\mathbb{R}\\ h^0\le r\\ h^1\le r}} r \qquad \text{for all } h^0, h^1 \in \mathbb{R}\ .$$

Let us first argue that in Eq. (17), we are optimizing over a set of tuples $(x, y, r_1, \ldots, r_K)$ that can be described by SDP constraints. Since $(x, y) \in \Gamma(\mathcal{S})$, Γ is linear, and $\mathcal{S}$ is given by a set of semidefinite constraints on (X, Y), it suffices to verify that the additional constraints imposed by $(x, y) \in \Omega$ and the constraints associated with the inner infimum in Eq. (17) can be expressed in semidefinite form. Indeed, with $(x, y) = \Gamma(X, Y)$ for $(X, Y) \in \Gamma^{-1}(\Omega) \cap \mathcal{S}$, the constraints take the following form. Since each function $h_j^b(\cdot,\cdot)$ is affine-linear in both arguments, the expression $h_j^b(x_j, y_j)$ is affine-linear in the operators X and Y. Explicitly, we have

$$h_j^b(x_j, y_j) = \sigma_j \left[\mathrm{tr}\left(X \hat{m}_j^b \sum_{k=1}^{p^2} S_{k,j}\eta_k\right) + \mathrm{tr}\left(Y \hat{\ell}_j^b \sum_{\ell=1}^{q^2} T_{j,\ell}\xi_\ell\right) - \hat{\ell}_j^b \hat{m}_j^b\right]$$

and the constraint

$$h_j^b(x_j, y_j) \le r_j$$

takes the form

$$\mathrm{tr}(X G_j^b) + \mathrm{tr}(Y H_j^b) - r_j \le s_j^b\ , \qquad (18)$$

where

$$G_j^b = \sigma_j \hat{m}_j^b \sum_{k=1}^{p^2} S_{k,j}\eta_k\ , \qquad H_j^b = \sigma_j \hat{\ell}_j^b \sum_{\ell=1}^{q^2} T_{j,\ell}\xi_\ell\ , \qquad s_j^b = \sigma_j \hat{\ell}_j^b \hat{m}_j^b$$

for each $j = 1, \ldots, K$. In addition, the constraints

$$\begin{aligned} \ell_j \leq x_j \leq L_j \qquad &\text{for} \qquad j = 1, \ldots, p^2 \qquad \text{and} \\ m_k \leq y_k \leq M_k \qquad &\text{for} \qquad k = 1, \ldots, q^2 \end{aligned}$$

become

$$\begin{aligned} \ell_j \leq \operatorname{tr}(X \sum_{k=1}^{p^2} S_{k,j} \eta_k) \leq L_j \qquad &\text{for} \qquad j = 1, \ldots, p^2 \qquad \text{and} \\ m_k \leq \operatorname{tr}(Y \sum_{\ell=1}^{q^2} T_{k,\ell} \xi_\ell) \leq M_k \qquad &\text{for} \qquad k = 1, \ldots, q^2 \ . \end{aligned} \tag{19}$$

In summary, we are optimizing the objective function

$$\sum_{j=1}^{K} r_j + \sum_{j=1}^{p^2} a_j x_j + \sum_{k=1}^{q^2} b_k y_k$$

over tuples (X, Y, r) satisfying the constraints given by Eqs. (18) and (19). Since this objective function is linear in X, Y, and r, respectively, the problem (16) is indeed a semidefinite program in (X, Y, r). □

We again give pseudocode giving an algorithmic realization of Lemma 3.4, see subroutine COMPUTEBOUNDSSDP in Fig. 11 of Appendix A.

A branch-and-bound algorithm for jointly constrained semidefinite bilinear programs

We are now ready to state our branch-and-bound algorithm which solves problem (11). The algorithm closely follows the algorithm of Al-Khayyal and Falk and only the subroutines need to be adapted.

Our algorithm takes as input a set $\mathcal{S} \subset \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ defined by SDP constraints, operators $Q \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p \otimes \mathbb{C}^q), A \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p), B \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ and a desired precision $\epsilon > 0$. It returns a value $\overline{\alpha}$ and an element $(X^*, Y^*) \in \mathcal{S}$ such that (for the function F defined in Eq. (10))

$$\overline{\alpha} = F(X^*, Y^*) \leq \left(\inf_{(X,Y) \in \mathcal{S}} F(X, Y) \right) + \epsilon \ . \tag{20}$$

The algorithm is given in Figs 8–12 of Appendix A. It follows exactly the same pattern as the branch-and-bound algorithm discussed in Section 2.1.2, with the only modification that lower bounds on the objective function are computed by solving SDPs (instead of general convex programs). In particular, with an identical analysis as that of a general branch-and-bound algorithm (see [2]), it follows that the iterative procedure described in the algorithm from Fig. 12 converges to a global solution of the problem (11). In other words, the terminating condition (20) will always be reached. We note, however, that (as is typical for branch-and-bound algorithms), guarantees for the rate of convergence are typically not available.

4 Application: Dobrushin curves of quantum channels

In this section, we apply our algorithm to a problem in quantum information theory. We first explain this problem in Sections 4.1 and 4.2, where we discuss the Dobrushin coefficient and the Dobrushin curve of a channel, respectively. In Section 4.3, we show that the problem of computing the Dobrushin curve is a semidefinite bilinear program. Finally, in Section 4.4, we present numerical results obtained by use of our algorithm.

4.1 Converse to unconstrained coding over cascades

Let Φ be a channel. Consider a setting where a message W is sent through a cascade consisting of n copies of this channel, with a relay $\mathcal{E}_j$ applied before the j-th application of Φ. We are interested in the amount of information the output

$$Y_n = \Phi \circ \mathcal{E}_n \circ \cdots \circ \mathcal{E}_2 \circ \Phi \circ \mathcal{E}_1(W)$$

of this cascade provides about the input W for an optimal coding strategy (defined by the choice of relay channels $\{\mathcal{E}_j\}_{j=1}^n$). Denoting the output after applying the relay $\mathcal{E}_j$ by X_j and the output of the j-th channel Φ by Y_j, we have the Markov property

$$W \to X_1 \to Y_1 \to X_2 \to Y_2 \to \cdots \to X_n \to Y_n \ . \tag{21}$$

For the case where $\Phi = P_{Y|X}$ is a classical channel from a set $\mathcal{X}$ to a set $\mathcal{W}$ and the message W is a binary random variable with uniform distribution on $\{0,1\}$, a natural information measure is the variational distance $\|P_{Y_n|W=0} - P_{Y_n|W=1}\|_1$ between the output distributions for different inputs. Accordingly, a key quantity is the Dobrushin coefficient

$$\eta(P_{Y|X}) = \sup_{P_{X_0},P_{X_1}} \frac{\|P_{Y|X} \circ P_{X_0} - P_{Y|X} \circ P_{X_1}\|_1}{\|P_{X_0} - P_{X_1}\|_1} \ , \tag{22}$$

where the optimization is over pairs (P_{X_0}, P_{X_1}) of distributions on $\mathcal{X}$ and where $P_{Y|X} \circ P_X$ is the distribution on $\mathcal{Y}$ given by the push-forward of P_X. Using the fact that $\|\cdot\|_1$ is non-increasing under application of channels, one can eliminate the choice of relays and conclude that

$$\|P_{Y_n|W=0} - P_{Y_n|W=1}\|_1 \leq \eta(P_{Y|X})^n$$

independently of the coding strategy given by $\{\mathcal{E}_j = P_{X_{j+1}|Y_j}\}$.

Similar reasoning applies to quantum channels $\Phi : \mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{Y})$ and relays $\mathcal{E}_j : \mathcal{B}(\mathcal{Y}) \to \mathcal{B}(\mathcal{Y})$ (i.e., completely positive trace-preserving maps) when a classical bit $W \in \{0,1\}$ is conveyed by encoding it into two states ρ_0, ρ_1. The so-called Dobrushin coefficient

$$\eta(\Phi) = \sup_{\rho_0,\rho_1 \in \mathcal{B}(\mathcal{X})} \frac{\|\Phi(\rho_0) - \Phi(\rho_1)\|_1}{\|\rho_0 - \rho_1\|_1} \ ,$$

provides the upper bound

$$\|\mathcal{F}_n(\rho_0) - \mathcal{F}_n(\rho_1)\|_1 \leq \eta(\Phi)^n \ , \qquad \text{where} \qquad \mathcal{F}_n = \Phi \circ \mathcal{E}_n \circ \cdots \circ \mathcal{E}_2 \circ \Phi \circ \mathcal{E}_1$$

on the trace distance (defined by $\|A\|_1 = \operatorname{tr}\sqrt{AA}$) between the output states. We refer to [9] for a detailed discussion of the Dobrushin and other information measure based contraction coefficients for quantum channels.

One may ask how the maximum output distinguishability $\|\Phi(\rho_0) - \Phi(\rho_1)\|_1$ behaves as a function of the distinguishability $\|\rho_0 - \rho_1\|_1$ of the input states. The following lemma shows that this quantity is linear in $\|\rho_0 - \rho_1\|_1$, and thus not particularly exciting.

Lemma 4.1. *Let $\delta \in [0,2]$. Then we have*

$$\sup_{\substack{\rho_0,\rho_1 \\ \|\rho_0-\rho_1\| \leq \delta}} \|\Phi(\rho_0) - \Phi(\rho_1)\|_1 = (\delta/2) \cdot \eta(\Phi) \ .$$

Proof. Consider the function $f : [0, 2] \to [0, 2]$ defined by

$$f(\delta) := \sup_{\substack{\rho_0,\rho_1 \\ \|\rho_0-\rho_1\|_1=\delta}} \|\Phi(\rho_0) - \Phi(\rho_1)\|_1 \ .$$

We first show that f is monotonically increasing. Indeed, suppose that $\delta \leq \delta'$, and let ρ_0, ρ_1 be states such that $\|\rho_0 - \rho_1\|_1 = \delta$ and $f(\delta) = \|\Phi(\rho_0) - \Phi(\rho_1)\|_1$. Let $\rho_0 - \rho_1 = A_+ - A_-$ be the decomposition of the difference into positive and negative parts (i.e., $A_+ \geq 0$ and $A_- \geq 0$). Then we have $\mathrm{tr}(A_+) = \mathrm{tr}(A_-) = \delta/2$. Accordingly, let us define the states $\sigma_\pm = \frac{2}{\delta}A_\pm$. Note that σ_+ and σ_- are orthogonal by definition. Choose an arbitrary state σ and define

$$\begin{aligned} \rho_0' &= \frac{\delta'}{2}\sigma_+ + \left(1 - \frac{\delta'}{2}\right)\sigma \\ \rho_1' &= \frac{\delta'}{2}\sigma_- + \left(1 - \frac{\delta'}{2}\right)\sigma \ . \end{aligned}$$

Then it is easy to check (using the orthogonality of σ_+ and σ_-) that

$$\|\rho_0' - \rho_1'\|_1 = \delta' \ .$$

Furthermore we have

$$\begin{aligned} \|\Phi(\rho_0') - \Phi(\rho_1')\|_1 &= \frac{\delta'}{\delta}\|\Phi(\rho_0) - \Phi(\rho_1)\|_1 \\ &\geq \|\Phi(\rho_0) - \Phi(\rho_1)\|_1 = f(\delta) \qquad \text{for } \delta' \geq \delta \ . \end{aligned}$$

This shows that $f(\delta') \geq f(\delta)$, as claimed. In particular, we also obtain $f(2) = \eta(\Phi)$.

More generally, the above proof shows that

$$\frac{f(\delta')}{\delta'} \geq \frac{f(\delta)}{\delta} \qquad \text{for all } \delta' \geq \delta \ ,$$

and thus with $\delta' = 2$

$$f(\delta) \leq \frac{\delta}{2} \cdot \eta(\Phi) \qquad \text{for all } \delta \in [0, 2] \ . \tag{23}$$

Now suppose that ρ_0, ρ_1 are states such that $\|\rho_0 - \rho_1\|_1 = 2$ and $\eta(\Phi) = \|\Phi(\rho_0) - \Phi(\rho_1)\|_1$. Then ρ_0 and ρ_1 are orthogonal, implying that (again for an arbitrary state σ) the states

$$\begin{aligned} \rho_0' &= \frac{\delta}{2}\rho_0 + \left(1 - \frac{\delta}{2}\right)\sigma \\ \rho_1' &= \frac{\delta}{2}\rho_1 + \left(1 - \frac{\delta}{2}\right)\sigma \end{aligned}$$

satisfy $\|\rho_0' - \rho_1'\|_1 = \delta$. Since we also have

$$\begin{aligned} \|\Phi(\rho_0') - \Phi(\rho_1')\|_1 &= \frac{\delta}{2}\|\Phi(\rho_0) - \Phi(\rho_1)\|_1 \\ &= \frac{\delta}{2}\eta(\Phi) \end{aligned}$$

we conclude that

$$f(\delta) \geq (\delta/2) \cdot \eta(\Phi) \ .$$

With Eq. (23) and the monotonicity of f, the claim follows. □

4.2 Converse to power-constrained coding over cascades

Consider a modified cascade coding problem, where a power constraint is introduced for each of the inputs X_j to the channel Φ in (21), for $j = 1, \ldots, n$. In other words, each relay $\mathcal{E}_j$ is required to have power-constrained outputs. In the case where $\Phi = P_{Y|X}$ is a classical channel with continuous variable input X (i.e., a random variable on $\mathbb{R}^m$), a natural power constraint is of the form

$$\mathbb{E}\left[\|X_j\|_2^2\right] \leq E \qquad \text{for all} \qquad j \in \{1, \ldots, n\} \ , \tag{24}$$

where $E > 0$ is some constant (determining the available power) and $\|x\|_2^2 = \sum_{k=1}^m x_k^2$ for $x \in \mathbb{R}^m$. Let $\mathcal{G}_E$ be the set of distributions P_X on $\mathbb{R}^m$ satisfying (24). To analyze this scenario, Polyanskiy and Wu [18] defined the function

$$F_E(\delta) = \sup_{\substack{P_{X_0}, P_{X_1} \in \mathcal{G}_E \\ \|P_{X_0} - P_{X_1}\|_1 \leq \delta}} \|P_{Y|X} \circ P_{X_0} - P_{Y|X} \circ P_{X_1}\|_1 \qquad \text{for} \qquad \delta \in [0, 2] \ , \tag{25}$$

which they call the Dobrushin curve of $P_{Y|X}$. Remarkably, Polyanskiy and Wu were able to compute (25) for the additive white Gaussian noise (AWGN) channel using a coupling argument. They then use this function to establish bounds on the distance $\|P_{Y_n|W=0} - P_{Y_n|W=1}\|_1$: inductively applying Definition (25), one obtains

$$\|P_{Y_n|W=0} - P_{Y_n|W=1}\|_1 \leq F^{\circ n}(2) \ ,$$

where $F^{\circ n} = F \circ \cdots \circ F$ is the n-fold composition of F (see Fig. 2 for an illustration). It should be noted that the Dobrushin coefficient (22) is not meaningful for the AWGN channel: it evaluates to 1 and does not provide converse bounds.

Similar concepts are naturally defined for a quantum channel $\Phi : \mathcal{B}(\mathcal{X}) \to \mathcal{B}(\mathcal{Y})$. In this case, a power constraint on states on $\mathcal{X}$ can be defined by fixing a Hamiltonian H (i.e., a self-adjoint operator) on $\mathcal{X}$ and requiring that the expected energy is less than a constant. For $E \in \mathbb{R}$, let

$$\mathcal{G}_E = \{\rho \in \mathcal{B}(\mathcal{H}) \mid \rho \geq 0, \operatorname{tr} \rho = 1 \text{ and } \operatorname{tr}(\rho H) \leq E\}$$

be the set of states satisfying this energy constraint. We can then define — in analogy with (25) — the function

$$F_E(\delta) = \sup_{\substack{\rho_0, \rho_1 \in \mathcal{G}_E \\ \|\rho_0 - \rho_1\|_1 \leq \delta}} \|\Phi(\rho_0) - \Phi(\rho_1)\|_1 \qquad \text{for} \qquad \delta \in [0, 2] \ . \tag{26}$$

Contrary to the unconstrained case discussed in Lemma 4.1, the function (26) is not linear in δ, and its evaluation appears to be challenging in general.

4.3 The Dobrushin curve as a semidefinite bilinear program

In this section, we show that the energy-constrained Dobrushin curve for finite-dimensional quantum channels can be cast as a semidefinite bilinear program of the form (11). This allows us to numerically compute the curve by applying our algorithm.

Lemma 4.2. *Consider a CPTPM $\Phi : \mathcal{B}(\mathbb{C}^d) \to \mathcal{B}(\mathbb{C}^d)$. Then we have*

$$F_E(\delta) = \delta \max_{(P,Q,R,S) \in \Gamma(E,\delta)} \operatorname{tr}(P\Phi(R - S)) \ , \tag{27}$$

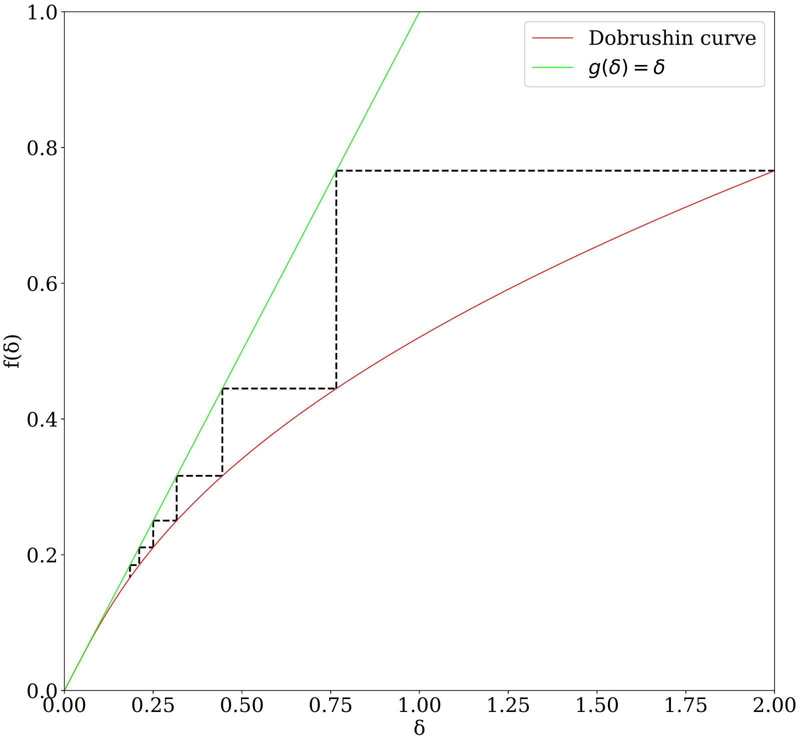


Figure 2: Using the Dobrushin curve to obtain upper bounds on the information loss of a cascade of channels.

where $\Gamma(E,\delta)$ *is the set of quadruples* $(P,Q,R,S) \in \mathcal{B}(\mathbb{C}^d)^{\times 4}$ *satisfying*

$$Q \geq 0 \ , \ \operatorname{tr}(Q) = 1 \ , \quad \textit{and} \ \operatorname{tr}(HQ) \leq E \ , \tag{28}$$

$$Q + \frac{\delta}{2}(R - S) \geq 0 \ , \tag{29}$$

$$\operatorname{tr}(H(Q + \frac{\delta}{2}(R - S)) \leq E \ , \tag{30}$$

$$0 \leq R \qquad \textit{and} \qquad \operatorname{tr}(R) = 1 \ , \tag{31}$$

$$0 \leq S \qquad \textit{and} \qquad \operatorname{tr}(S) = 1 \ , \tag{32}$$

$$0 \leq P \leq I \ . \tag{33}$$

To see more explicitly that the optimization problem (27) is a semidefinite bilinear program, define the flip operator $\mathbb{F} = \sum_{i,j=1}^{d} |ij\rangle\langle ji|$, where $\{|i\rangle\}_{i=1}^{d}$ is an orthonormal basis of $\mathbb{C}^d$. Using the identity

$\mathrm{tr}(\mathbb{F}(A \otimes B)) = \mathrm{tr}(AB)$ for all $A, B \in \mathcal{B}(\mathbb{C}^d)$, we have

$$\begin{aligned} \mathrm{tr}(P\Phi(R-S)) &= \mathrm{tr}(\mathbb{F}(P \otimes \Phi(R))) - \mathrm{tr}(\mathbb{F}(P \otimes \Phi(S))) \\ &= \mathrm{tr}\left((I \otimes \Phi^*)(\mathbb{F})(P \otimes R)\right) - \mathrm{tr}\left((I \otimes \Phi^*)(\mathbb{F})(P \otimes S)\right) \ , \end{aligned} \tag{34}$$

where Φ^* is the adjoint channel (with respect to the Hilbert-Schmidt inner product) i.e., it is defined by $\mathrm{tr}(A\Phi(B)) = \mathrm{tr}(\Phi^*(A)B)$ for all $A, B \in \mathcal{B}(\mathcal{H})$. This matches the form of (11) with $X = P$, $Y = R \oplus S$, and $Q = \hat{Q} \oplus 0 - Q_1 \oplus 0 \oplus Q_3$ where $Q_1 \oplus Q_3 = \hat{Q}$, and $\hat{Q} = (I \otimes \Phi^*)(\mathbb{F})$.

Proof. For convenience, let $\Theta(E, \delta)$ denote the set of pairs (ρ_0, ρ_1) of states satisfying

$$\mathrm{tr}(H\rho_j) \leq E \qquad \text{for } j = 1, 2 \qquad \text{and} \qquad \|\rho_0 - \rho_1\|_1 \leq \delta \ . \tag{35}$$

Suppose $(P, Q, R, S) \in \Gamma(E, \delta)$. Set

$$\rho_1 = Q \qquad \text{and} \qquad \rho_0 = Q + \frac{\delta}{2}(R - S) \ .$$

Because of (28), ρ_1 is a state and satisfies the energy constraint, i.e., $\rho_1 \in \mathcal{G}_E$. Similarly, ρ_0 is a state since it has unit trace because of (28), (31), and (32), and because it is non-negative by (29). By Eq. (30), it also belongs to $\mathcal{G}_E$. Now observe that

$$\|\rho_0 - \rho_1\|_1 = \frac{\delta}{2}\|R - S\|_1 \leq \delta \ ,$$

since both R and S are states (cf. Eqs. (31) and (32)) and hence $\|R - S\|_1 \leq 2$. This shows that $(\rho_0, \rho_1) \in \Theta(E, \delta)$. Furthermore, we have

$$\delta \max_{0 \leq P \leq I} \mathrm{tr}(P\Phi(R - S)) = \left\| \Phi\left(\frac{\delta}{2}(R - S)\right) \right\|_1 = \|\Phi(\rho_0) - \Phi(\rho_1)\|_1 \ . \tag{36}$$

We conclude that $F_E(\delta) \geq \delta \max_{(P,Q,R,S) \in \Gamma(E,\delta)} \mathrm{tr}(P\Phi(R - S))$.

To show the converse inequality, assume that $(\rho_0, \rho_1) \in \Theta(E, \delta)$. Then

$$\rho_0 - \rho_1 = A_+ - A_- \tag{37}$$

for two orthogonal nonnegative operators A_+, A_- satisfying

$$\mathrm{tr}(A_+) = \mathrm{tr}(A_-) = \frac{\delta}{2} \ .$$

Set $Q = \rho_1$, $R = \frac{2}{\delta}A_+$, $S = \frac{2}{\delta}A_-$ and

$$P = \arg\max\nolimits_{0 \leq P \leq I} \mathrm{tr}(P(\Phi(\rho_0) - \Phi(\rho_1))) \ .$$

Clearly, the quadruple (P, Q, R, S) satisfies Eqs. (28), (31), (32) and (33). It remains to check (29) and (30). Observe that by definition, we have

$$Q + \frac{\delta}{2}(R - S) = \rho_1 + A_+ - A_- = \rho_0 \ .$$

This implies that (29) and (30) are also satisfied.

Since the identity (36) also holds by definition of (P, Q, R, S), we find

$$F_E(\delta) \leq \delta \max_{(P,Q,R,S) \in \Gamma(E,\delta)} \mathrm{tr}(P\Phi(R - S)) \ .$$

This concludes the proof. □

We note that the statement of Lemma 4.2 simplifies somewhat in the case where the map $\Phi : \mathcal{B}(\mathbb{C}^2) \to \mathcal{B}(\mathbb{C}^2)$ is a qubit channel. This is because the operators $A_\pm$ in Eq. (37) are orthogonal, and hence proportional to rank-1-projections $|\varphi_\pm\rangle\langle\varphi_\pm|$ which satisfy $|\varphi_+\rangle\langle\varphi_+| + |\varphi_-\rangle\langle\varphi_-| = I$. Here I is the identity operator on $\mathbb{C}^2$. In particular, this means that we can eliminate $S = I - R$. Retracing the proof of Lemma 4.2, we obtain the following.

Corollary 4.3. *Consider a qubit channel $\Phi : \mathcal{B}(\mathbb{C}^2) \to \mathcal{B}(\mathbb{C}^2)$. Then*

$$F_E(\delta) = \delta \max_{(P,Q,R)\in\Gamma(E,\delta)} \operatorname{tr}(P\Phi(2R - I)) \ ,$$

where $\Gamma(E, \delta)$ is the set of triples $(P, Q, R) \in \mathcal{B}(\mathbb{C}^2)^{\times 3}$ satisfying

$$\begin{aligned} \operatorname{tr}(Q) = 1 \ , \ Q \geq 0 \ , \ \ &\textit{and} \ \operatorname{tr}(HQ) \leq E \ , \\ Q + \frac{\delta}{2}(2R - I) &\geq 0 \ , \\ \operatorname{tr}(H(Q + \frac{\delta}{2}(2R - I)) &\leq E \ , \\ \operatorname{tr}(R) &= 1 \ , \\ 0 &\leq R \ , \\ 0 &\leq P \leq I \ . \end{aligned}$$

One can furthermore add the condition $\operatorname{tr}(P) = 1$ as we know that the optimal P is rank 1 and satisfies this condition. Again we may recast this in the form of (11) using the fact that (in analogy to (34))

$$\operatorname{tr}(P\Phi(2R - I)) = 2 \operatorname{tr}\left((I \otimes \Phi^*)(\mathbb{F})(P \otimes R)\right) - \operatorname{tr}\left(P\Phi(I)\right) \ .$$

In this case, we obtain both a bilinear term as well as a term which is linear in P.

4.4 Numerical computation of Dobrushin curves

According to Lemma 4.2 (respectively Corollary 4.3), we can use our biconvex programming algorithm to calculate Dobrushin curves for quantum channels.

For concreteness, we consider qubit channels. Let $\{\sigma_j\}_{j=1}^3$ be the Pauli matrices

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ .$$

A state ρ (i.e. a non-negative operator with unit trace) can be represented as

$$\rho = \frac{1}{2}\left(I + \sum_{k=1}^{3} w_k \sigma_k\right) \tag{38}$$

with $w \in \mathbb{R}^3$ satisfying $\|w\|_2 \leq 1$. The vector w is called the Bloch vector of the state ρ. We remark that Eq. (38) provides an isometric identification of the set of states on $\mathbb{C}^2$ with trace-norm, and the unit ball (with respect to the Euclidean norm) in $\mathbb{R}^3$, see e.g., [16, Chapter 9].

Without loss of generality, we will assume that the Hamiltonian under consideration is

$$H = \sigma_3 \ . \tag{39}$$

In other words, we will be interested in states ρ having Bloch vectors $w = (w_1, w_2, w_3)$ satisfying an inequality of the form $w_3 \leq E$.

4.4.1 Example: the dephasing channel

As a first example consider a dephasing channel $\Phi : \mathcal{B}(\mathbb{C}^2) \to \mathcal{B}(\mathbb{C}^2)$. For $a \in [0,1]$, this acts as

$$\Phi\left(\alpha_0 I + \sum_{k=1}^{3} \alpha_k \sigma_k\right) = \alpha_0 I + a\left(\alpha_1\sigma_1 + \alpha_2\sigma_2\right) + \alpha_3\sigma_3 \qquad \text{for all } \alpha \in \mathbb{R}^4 \ . \tag{40}$$

The dephasing channel (40) has the invariance property

$$\Phi\left(e^{i\theta\sigma_3}\rho e^{-i\theta\sigma_3}\right) = e^{i\theta\sigma_3}\Phi(\rho)e^{-i\theta\sigma_3} \qquad \text{for all } \theta \in [0,2\pi) \text{ and } \rho \in \mathcal{B}(\mathbb{C}^2) \ . \tag{41}$$

The Hamiltonian (39) is also invariant under rotations around the σ_3-axis, i.e., we have

$$H = e^{-i\theta\sigma_3} H e^{i\theta\sigma_3} \qquad \text{for all } \theta \in [0,2\pi) \ . \tag{42}$$

Eq. (42) implies that the set $\mathcal{G}_E$ of energy-constrained states is closed under the family of maps $\rho \mapsto e^{i\theta\sigma_3}\rho e^{-i\theta\sigma_3}$. By the invariance property (41) and the unitary invariance of the trace norm, we conclude that applying the joint rotations

$$(\rho_0, \rho_1) \mapsto (e^{i\theta\sigma_3}\rho_0 e^{-i\theta\sigma_3}, e^{i\theta\sigma_3}\rho_1 e^{-i\theta\sigma_3}) \qquad \text{for any } \theta \in [0,2\pi)$$

to a pair (ρ_0, ρ_1) of states leaves their energies as well as the distances $\|\rho_0 - \rho_1\|_1$ and $\|\Phi(\rho_0) - \Phi(\rho_1)\|_1$ invariant. Because $\rho \mapsto e^{i\theta}\rho e^{-i\theta}$ amounts to the map

$$(w_1, w_2, w_3) \mapsto ((\cos 2\theta)w_1 - (\sin 2\theta)w_2, (\sin 2\theta)w_1 + (\cos 2\theta)w_2, w_3)$$

on the level of Bloch vectors $w = (w_1, w_2, w_3) \in \mathbb{R}^3$, we conclude the following: for any fixed energy E, there is a pair of states $(\rho_0, \rho_1) \in \Theta(E, \delta)$ (see Eq. (35)) such that

$$F_E(\delta) = \|\Phi(\rho_0) - \Phi(\rho_1)\|_1 \ ,$$

(i.e., the states achieve the optimum in the definition of the Dobrushin curve), and such that the Bloch vector w of ρ_1 lies in the (σ_1, σ_3)-plane, i.e.,

$$w_2 = 0 \ . \tag{43}$$

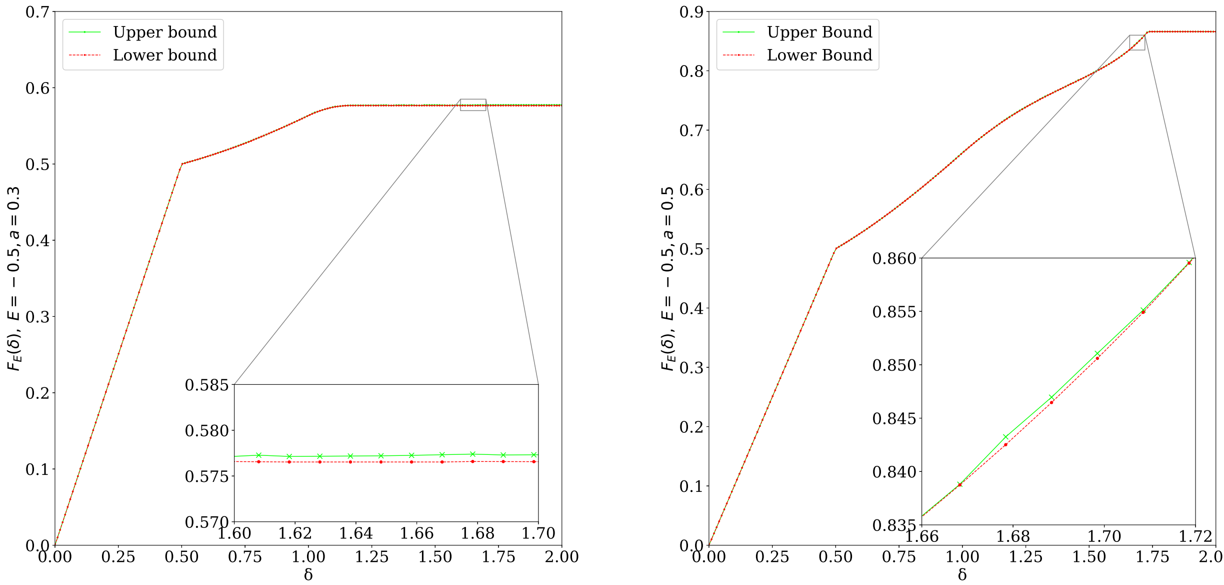
In the semidefinite bilinear program introduced in Corollary 4.3 (where Q corresponds to ρ_1), this means that we may add the constraint

$$\mathrm{tr}(\sigma_2 Q) = 0 \tag{44}$$

without changing the value of the optimization problem.

In Figs. 3 and 4, we show numerically computed Dobrushin curves for dephasing channels. These are applied by using the formulation as a semidefinite bilinear program (see Corollary 4.3), and imposing the constraint (44).

In Fig. 5, we present numerical data illustrating the importance of exploiting continuous symmetries by a constraint as in Eq. (44). It is well-known that branch-and-bound algorithms perform badly in the context of such symmetries, hence it is important to include such constraints. Note that more generic channels as discussed in Section 4.4.3 typically do not exhibit such continuous symmetries.



(a) The Dobrushin curve for the dephasing channel Φ for $a = 0.3$. The algorithm required an average of 88 elements (worst case: 178 elements) in the partition $\mathcal{P}$ to reach the desired precision.

(b) The Dobrushin curve for the dephasing channel Φ for $a = 0.5$. The algorithm required an average of 92 elements (worst case: 142 elements) in the partition $\mathcal{P}$ to reach the desired precision.

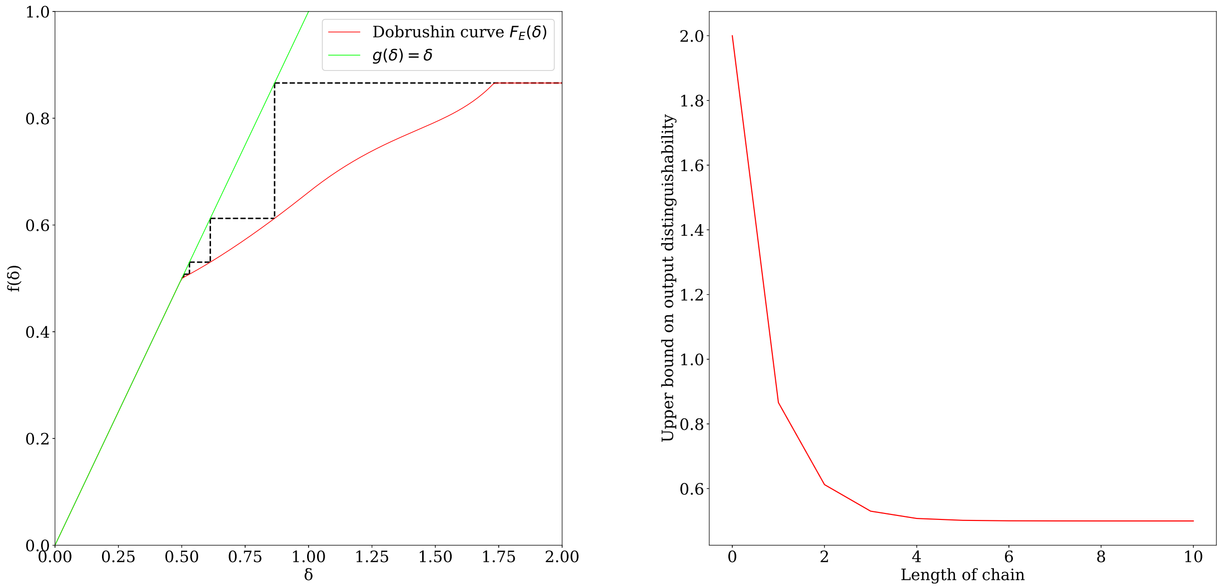
Figure 3: The Dobrushin curves of the dephasing channels Φ for two different values of a, for $E = -0.5$. For 200 values of $\delta \in [0, 2]$, the algorithm was run with a desired precision of $\epsilon = 10^{-3}$.

4.4.2 Discussion of the dephasing channel

We note that our algorithm also provides — in addition to the value $F_E(\delta)$ of the Dobrushin curve — a pair of states $(\rho_0, \rho_1) \in \Theta(E, \delta)$ satisfying $F_E(\delta) = \|\Phi(\rho_0) - \Phi(\rho_1)\|_1$. We call such a pair of states optimal for the Dobrushin curve. In the special case of the dephasing channel defined in Eq. (40) for some $a \in [0, 1]$, we can provide the following description of such pairs (valid for instance for Fig. 3b, i.e., for $a = 0.5$ and $E = -0.5$). We note that this description is based on a heuristic geometric analysis of the problem. However, our numerical data shows that the following pairs of states are indeed optimal. We also note that — while a full analytical proof of optimality may in principle be constructed for the dephasing channel, such a brute-force calculation is unlikely to be achievable e.g., for generic qubit or qutrit channels, where symmetry arguments are not applicable and positivity constraints are particularly difficult to deal with.

Recall from Eq. (43) that we can assume without loss of generality that one of the states (ρ_0, ρ_1) — say ρ_0 for concreteness — has Bloch vector lying in the plane orthogonal to $(0, 1, 0)$. It turns out that ρ_1 can also be chosen to lie in this plane. In Fig. 6, we show the projection of the Bloch sphere onto this plane, and illustrate a choice of optimal code states (in terms of their Bloch vectors). More precisely, we identify three regimes:

Regime I: for $\delta \in [0, 1 - |E|]$, an optimal pair $(\rho_0, \rho_1) = \Theta(E, \delta)$ is given by the pure state $\rho_1 = |1\rangle\langle 1|$ with Bloch vector $\vec{r}_1 = (0, 0, 1)$ and a mixed state ρ_0 whose Bloch vector $\vec{r} = (0, 0, 1 - \delta)$ also lies on the e_3-axis.see Fig. 6b



(a) The procedure to obtain upper bounds on the output distinguishability of a cascade of channels.

(b) The maximal output distinguishability for a cascade of phase-damping channels ($E = -0.5$), as a function of the length of the cascade.

Figure 4: Using the Dobrushin curve to obtain upper bounds on the information loss of a cascade of dephasing channels Φ, for $E = -0.5$ and $a = 0.5$.

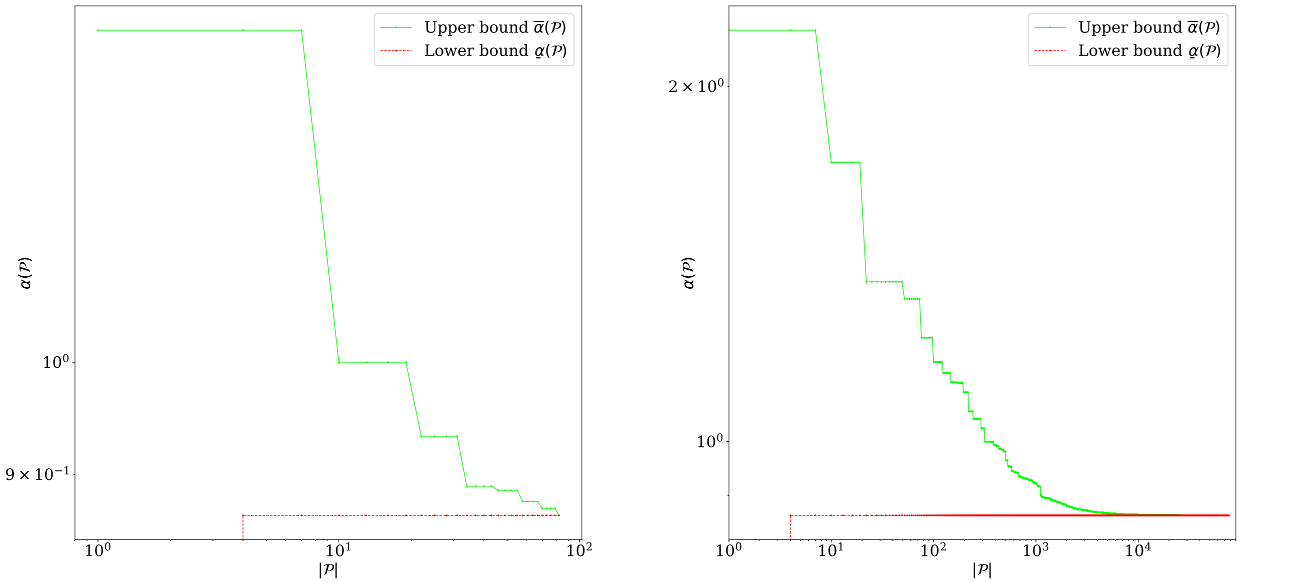
Regime II: for $1 - |E| \leq \sqrt{2(1-|E|)}$, an optimal pair is given by $\rho_1 = |1\rangle\langle 1|$ as in Regime II and ρ_0 having Bloch vector $(x, 0, -|E|)$, with x chosen such that $\|\rho_0 - \rho_1\|_1 = \delta$, see Fig. 6c.

Regime III: for $\sqrt{2(1-|E|)} \leq \delta \leq 2\sqrt{1-|E|^2}$, we can choose the state ρ_0 to have Bloch vector given by the "eastern" point of intersection of the projection of the Bloch sphere onto the plane orthogonal to $(0, 1, 0)$, and the plane $(x, y, -|E|)$. On the other hand, ρ_1 a pure state at distance δ, see Fig. 6.

This completes the description, as there are no pairs of states at distance $\delta > 2\sqrt{1-|E|^2}$ which belong to the energy-constrained subset $\mathcal{G}_E$. We now analyze this "coding" strategy for the dephasing channel and show the following:

Lemma 4.4. *We have $(\rho_0, \rho_1) \in \Theta(E, \delta)$ in all three regimes. In particular, these pairs of states give the following lower bound on the Dobrushin curve of the dephasing channel:*

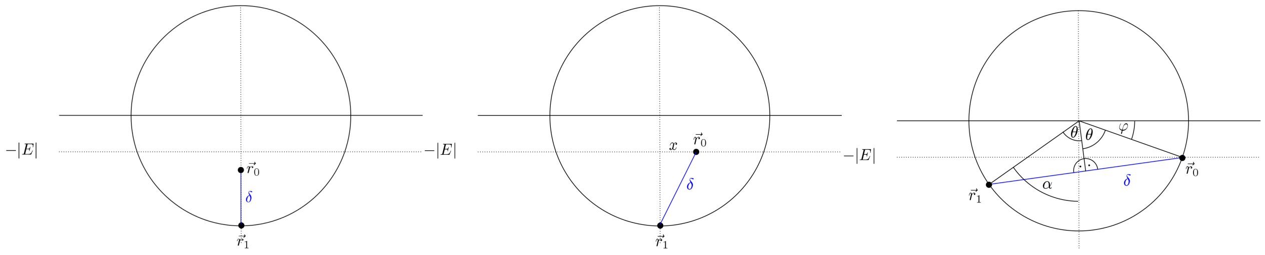
$$f_E(\delta) \geq \begin{cases} \delta & \text{if } \delta < 1 - |E| \\ g_E(\delta) & \text{if } 1 - |E| \leq \delta \leq \sqrt{2(1-|E|)} \\ h_E(\delta) & \text{if } \sqrt{2(1-|E|)} \leq \delta \leq 2\sqrt{1-|E|^2} \\ 2a\sqrt{1-|E|^2} & \text{if } \delta > 2\sqrt{1-|E|^2} \end{cases} \tag{45}$$



(a) Convergence of the algorithm with the symmetry constraint (44)

(b) Convergence of the algorithm without the symmetry constraint (44)

Figure 5: Upper and lower bounds $\overline{\alpha}(\mathcal{P})$, $\underline{\alpha}(\mathcal{P})$ as a function of the size $|\mathcal{P}|$ of the partition. This provides a measure for the speed of convergence of the algorithm. The figures are for $E = -0.5$, $a = 0.5$, $\delta = 2$, and $\epsilon = 10^{-5}$.



(a) The regime $0 \leq \delta \leq 1 - |E|$.

(b) The regime $1 - |E| \leq \delta \leq \sqrt{2(1-|E|)}$.

(c) The regime $\sqrt{2(1-|E|)} \leq \delta \leq 2\sqrt{1-|E|^2}$.

Figure 6: Bloch vectors $\vec{r}_0$ and $\vec{r}_1$ of a pair of optimal states for the Dobrushin curve of the dephasing channel.

where

$$\begin{aligned}
g_E(\delta) &:= \sqrt{a^2(\delta^2 - (1-|E|)^2) + (1-|E|)^2} \\
h_E(\delta) &:= \Big[(|E| + \cos\left(2\arccos(\delta/2) + \arccos(|E|)\right))^2 \\
&\quad + a^2 \left(\sqrt{1-|E|^2} + \sin\left(2\arccos(\delta/2) + \arccos(|E|)\right)\right)^2 \Big]^{1/2} \ .
\end{aligned}$$

The proof relies on elementary geometry. The curve given by the rhs. of Eq. (45) matches the numerically observed Dobrushin curve shown in Fig. 3b; this shows that the pairs of states considered above are indeed optimal.

Proof. We consider each regime separately.

Regime I: Consider $\delta \in [0, 1-|E|]$. In this case, the choice

$$\begin{aligned}
\vec{r}_1 &= (0, 0, -1+|E|) \ , \\
\vec{r}_2 &= (0, 0, -1)
\end{aligned}$$

is optimal and leads to $\|\vec{r}_1' - \vec{r}_2'\|_1 = \|\vec{r}_1 - \vec{r}_2\|_1$ for the output Bloch vectors $\vec{r}_1'$ and $\vec{r}_2'$.

Regime II: Now consider $\delta \in [1-|E|, \sqrt{2(1-|E|)}]$ see Fig. 6. The initial Bloch vectors are

$$\begin{aligned}
\vec{r}_1 &= (x, 0, -|E|) \ , \\
\vec{r}_2 &= (0, 0, -1) \ ,
\end{aligned}$$

where $x^2 + (1-|E|)^2 = \delta^2$, i.e., $x = \sqrt{\delta^2 - (1-|E|)^2}$. The Bloch vectors after application of the channel are

$$\begin{aligned}
\vec{r}_1' &= (ax, 0, -|E|) \ , \\
\vec{r}_2' &= (0, 0, -1) \ ,
\end{aligned}$$

such that

$$\begin{aligned}
\|\vec{r}_1' - \vec{r}_2'\|_1 &= \sqrt{a^2x^2 + (1-|E|)^2} \\
&= \sqrt{a^2(\delta^2 - (1-|E|)^2) + (1-|E|)^2} \ ,
\end{aligned}$$

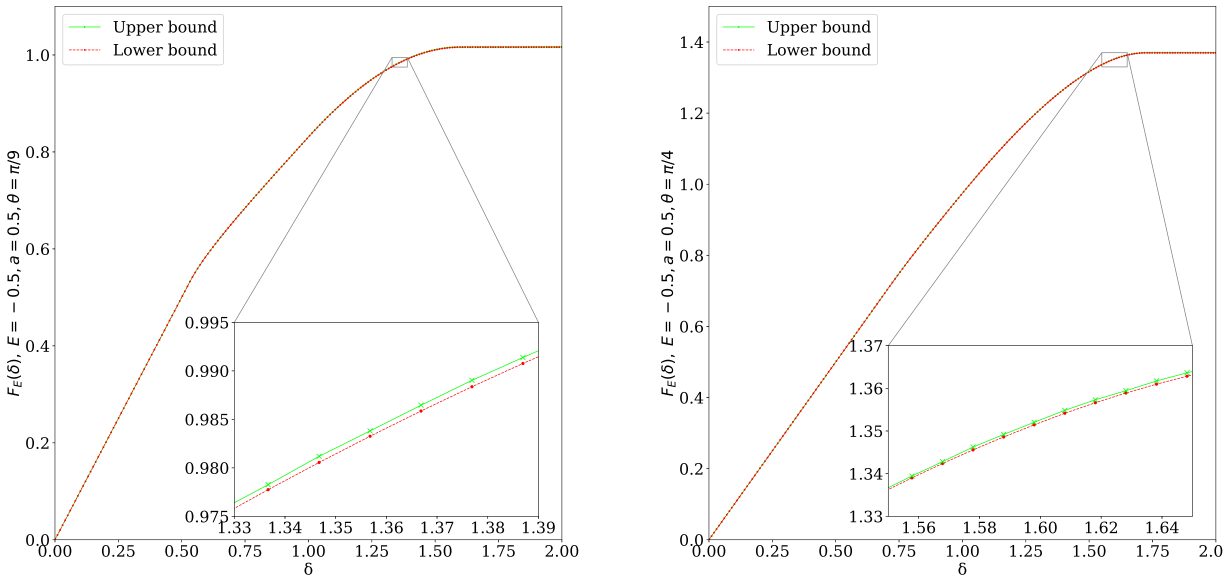
for $\delta \in [1-|E|, \sqrt{(1-|E|)^2 + 1}]$.

Regime III: Let us now look at $\delta \in [\sqrt{2(1-|E|)}, 2\sqrt{1-|E|^2}]$. Consider Fig. 6. Assume that ρ_1 has Bloch vector given by

$$\vec{r}_1 = (\cos\varphi, 0, -\sin\varphi) \ ,$$

where $\sin\varphi = |E|$, and that ρ_2 has Bloch vector $\vec{r}_2$ specified by an angle θ as in Fig. 6c. The figure shows that $\delta = 2\sin\theta$ and $\alpha = 2\theta + \varphi - \pi/2$. Thus, the Bloch vector $\vec{r}_2$ is

$$\begin{aligned}
\vec{r}_2 &= (-\sin\alpha, 0, -\cos\alpha) \\
&= (-\sin(2\theta + \varphi - \pi/2), 0, -\cos(2\theta + \varphi - \pi/2)) \ .
\end{aligned}$$



(a) The Dobrushin curve for the dephasing channel $\Phi_{\pi/9}$ with rotated principal axes. The algorithm required an average of 898 elements (worst case: 2056 elements) in the partition $\mathcal{P}$ to reach the desired precision.

(b) The Dobrushin curve for the dephasing channel $\Phi_{\pi/4}$ with rotated principal axes. The algorithm required an average of 598 elements (worst case: 730 elements) in the partition $\mathcal{P}$ to reach the desired precision.

Figure 7: The Dobrushin curves of the dephasing channels $\Phi_{\pi/9}$ and $\Phi_{\pi/4}$ with rotated principal axes for two different values of the rotation angle θ, for $a = \frac{1}{2}$ and $E = -0.5$. For 200 values of $\delta \in [0, 2]$, the algorithm was run with a desired precision of $\epsilon = 10^{-3}$.

These get mapped to

$$\begin{aligned}\vec{r}_1' &= (a\cos\varphi, 0, -\sin\varphi)\ ,\\ \vec{r}_2' &= (-a\sin(2\theta+\varphi-\pi/2), 0, -\cos(2\theta+\varphi-\pi/2))\ ,\end{aligned}$$

such that

$$\begin{aligned}\|\vec{r}_1' - \vec{r}_2'\|_1 = \big(&a^2\left(\cos\varphi + \sin(2\theta+\varphi-\pi/2)\right)^2\\ &+ \left(\sin\varphi - \cos(2\theta+\varphi-\pi/2)\right)^2\big)^{1/2}\ ,\end{aligned}$$

where $\theta = \arcsin(\delta/2)$ and $\varphi = \arcsin|E|$.

Finally, for $\delta > 2\sqrt{1-|E|^2}$ the two inputs

$$\begin{aligned}\vec{r}_1 &= (2\sqrt{1-|E|^2}, 0, -1+|E|)\ ,\\ \vec{r}_2 &= (-2\sqrt{1-|E|^2}, 0, -1+|E|)\end{aligned}$$

are optimal and lead to $\|\vec{r}_1' - \vec{r}_2'\|_1 = a\cdot\|\vec{r}_1 - \vec{r}_2\|_1$. □

4.4.3 Dobrushin-curves of generic qubit channels

In Fig. 7, we consider more general channels which no do not obey the symmetry condition (41).

Consider a dephasing channel whose principal axes are not the $\{\sigma_1, \sigma_2, \sigma_3\}$-axes. To achieve this, we rotate the Kraus operators of the channel around the σ_1-axis by an angle $\theta_1 \in [0, 2\pi]$. This means that we conjugate by the unitary $e^{\frac{i}{2}\theta_1\sigma_1}$, obtaining the channel

$$\Phi_{\theta_1}(\rho) = e^{-\frac{i}{2}\theta_1\sigma_1}\Phi\left(e^{\frac{i}{2}\theta_1\sigma_1}\rho e^{-\frac{i}{2}\theta_1\sigma_1}\right)e^{\frac{i}{2}\theta_1\sigma_1} ,$$

where Φ is the dephasing channel (see Eq. (40)). The Dobrushin curve of such a channel for a fixed θ_1 can be calculated by our algorithm and results in the curve depicted in Fig. 7.

Program code

Python program code for the algorithm constructed here is available together with the TeX-source code on the ArXiv.

Acknowledgments

SH and RK acknowledge support by DFG project no. K05430/1-1. RK is supported by the Technische Universität München - Institute of Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement no. 291763. MT acknowledges an Australian Research Council Discovery Early Career Researcher Award, project no. DE160100821. The authors thank Volkher Scholz and David Reeb for discussions on Dobrushin curves.

References

[1] F. A. Al-Khayyal. *"Generalized bilinear programming: Part I. Models, applications and linear programming relaxation"*. European Journal of Operational Research **60**(3): 306 – 314 (1992).

[2] F. A. Al-Khayyal and J. E. Falk. *"Jointly Constrained Biconvex Programming"*. Mathematics of Operations Research **8**(2): 273–286 (1983).

[3] M. Berta, O. Fawzi, and V. B. Scholz. *"Quantum Bilinear Optimization"*. SIAM Journal on Optimization **26**(3): 1529–1564 (2016).

[4] A. V. Cabot and R. L. Francis. *"Solving Certain Nonconvex Quadratic Minimization Problems by Ranking the Extreme Points"*. Operations Research **18**(1): 82–86 (1970).

[5] J. Falk. *"Lagrange Multipliers and Nonconvex Programs"*. SIAM Journal on Control **7**(4): 534–545 (1969).

[6] J. E. Falk. *"A linear max—min problem"*. Mathematical Programming **5**(1): 169–188 (1973).

[7] C. A. Floudas and V. Visweswaran. *Quadratic Optimization*, pages 217–269. Springer US (1995).

[8] J. Gorski, F. Pfeuffer, and K. Klamroth. *"Biconvex sets and optimization with biconvex functions: a survey and extensions"*. Mathematical Methods of Operations Research **66**(3): 373–407 (2007).

[9] F. Hiai and M. B. Ruskai. *"Contraction coefficients for noisy quantum channels"*. Journal of Mathematical Physics **57**(1): 015211 (2016).

[10] R. Horst, P. Pardalos, and N. Van Thoai. *Introduction to Global Optimization.* Springer US (2000).

[11] H. Konno. *"Bilinear Programming: Part II. Applications of Bilinear Programming"*. Technical Report, Operations Research, Stanford University **10**, (1971).

[12] H. Konno. *"A cutting plane algorithm for solving bilinear programs"*. Mathematical Programming **11**(1): 14–27 (1976).

[13] R. L. Kosut and D. A. Lidar. *"Quantum error correction via convex optimization"*. Quantum Information Processing **8**(5): 443–459 (2009).

[14] A. G. Nahapetyan. *Bilinear programming*, pages 279–282. Springer US (2009).

[15] J. Nash. *"Non-Cooperative Games"*. Annals of Mathematics **54**(2): 286–295, (1951).

[16] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition.* Cambridge University Press (2011).

[17] A. Peres. *"Collective tests for quantum nonlocality"*. Phys. Rev. A **54**: 2685–2689 (1996).

[18] Y. Polyanskiy and Y. Wu. *"Dissipation of Information in Channels With Input Constraints"*. IEEE Transactions on Information Theory **62**(1): 35–55 (2016).

[19] M. Reimpell and R. F. Werner. *"Iterative Optimization of Quantum Error Correcting Codes"*. Phys. Rev. Lett. **94**: 080501 (2005).

[20] B. Schumacher. *"Sending entanglement through noisy quantum channels"*. Physical Review A **54**(4): 2614–2628 (1996).

[21] H. D. Sherali and C. M. Shetty. *"A finitely convergent algorithm for bilinear programming problems using polar cuts and disjunctive face cuts"*. Mathematical Programming **19**(1): 14–31 (1980).

[22] H. Vaish and C. M. Shetty. *"The bilinear programming problem"*. Naval Research Logistics Quarterly **23**(2): 303–309 (1976).

[23] H. Vaish and C. M. Shetty. *"A cutting plane algorithm for the bilinear programming problem"*. Naval Research Logistics Quarterly **24**(1): 83–94 (1977).

[24] T. Vertesi and N. Brunner. *"Disproving the Peres conjecture by showing Bell nonlocality from bound entanglement"*. Nature Communications **5**(5297) (2014).

[25] R. F. Werner and M. M. Wolf. *"Bell Inequalities and Entanglement"*. Quantum Info. Comput. **1**(3): 1–25, (2001).

A Pseudocode for jointly constrained semidefinite bilinear programs

In this appendix we give pseudocode for the full algorithm.

```
function ComputeVectorRep($Q, X, Y, \{\eta_j\}_{j=1}^{p^2}, \{\xi_j\}_{j=1}^{q^2}$)

Input: Operators $Q \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p \otimes \mathbb{C}^q), X \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p), Y \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$
orthonormal bases $\{\eta_j\}_{j=1}^{p^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $\{\xi_k\}_{k=1}^{q^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$
Output: Vectors $(x, y) \in \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$ such that $(x, y) = \Gamma(X, Y)$

    Find $U_{j,k} = \mathrm{tr}((\eta_j \otimes \xi_k)Q)$ for $j = 1, \ldots, p^2, k = 1, \ldots, q^2$
    Find singular value decomposition $U = S\Delta T$
    for each $j = 1, \ldots, p^2$ do
        Set $x_j = \sum_{k=1}^{p^2} S_{k,j} \,\mathrm{tr}(X\eta_k)$
    for each $k = 1, \ldots, q^2$ do
        Set $y_k = \sum_{\ell=1}^{q^2} T_{\ell,k} \,\mathrm{tr}(Y\xi_\ell)$
    return $(x, y)$
```

```
function ComputeOperator($Q, x, y, \{\eta_j\}_{j=1}^{p^2}, \{\xi_j\}_{j=1}^{q^2}$)

Input: Operators $Q \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p \otimes \mathbb{C}^q), (x, y) \in \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$
orthonormal bases $\{\eta_j\}_{j=1}^{p^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $\{\xi_k\}_{k=1}^{q^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$
Output: Operators $(X, Y) \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ such that $(x, y) = \Gamma(X, Y)$

    Find $U_{j,k} = \mathrm{tr}((\eta_j \otimes \xi_k)Q)$ for $j = 1, \ldots, p^2, k = 1, \ldots, q^2$
    Find singular value decomposition $U = S\Delta T$
    Set $X = \sum_{j=1}^{p^2} \sum_{k=1}^{p^2} S_{k,j} x_k \eta_j$
    Set $Y = \sum_{k=1}^{q^2} \sum_{\ell=1}^{q^2} T_{\ell,k} y_\ell \xi_k$
    return $(X, Y)$
```

Figure 8: The algorithm ComputeVectorRep computes the image $\Gamma(X, Y)$ of two operators (X, Y), cf. Lemma 3.1. Conversely, the procedure ComputeOperator computes $\Gamma^{-1}(x, y)$, i.e, converts a pair of vectors into a pair of operators. We note that the required singular value decomposition can be computed once and stored. It can then be reused whenever the function is invoked.

function BranchHyperrectangle($\Omega, (v,w)$)

Input: Hyperrectangle $\Omega = \Omega(\ell, L, m, M) \subset \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$, branching point $(v,w) \in \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$
Output: Hyperrectangles $\Omega^{(1)}, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)} \subset \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$ such that $\cup_{i=1}^{4} \Omega^i = \Omega$

```
    Pick index $I \in \{1, \dots, K\}$ which produces the largest difference between the two sides in the inequality
    $\max\{m_i v_i + \ell_i w_i - \ell_i m_i, M_i v_i + L_i w_i - L_i M_i\} < v_i w_i$
    for $i \leftarrow 1, \dots, K$ do
        if $i \neq I$ then
            for $j \leftarrow 1, 2, 3, 4$ do
                $(\ell_i^{(j)}, L_i^{(j)}, m_i^{(j)}, M_i^{(j)}) \leftarrow (\ell_i, L_i, m_i, M_i)$
        $(\ell_I^{(1)}, L_I^{(1)}, m_I^{(1)}, M_I^{(1)}) \leftarrow (\ell_I, v_I, m_I, w_I)$
        $(\ell_I^{(2)}, L_I^{(2)}, m_I^{(2)}, M_I^{(2)}) \leftarrow (v_I, L_I, m_I, w_I)$
        $(\ell_I^{(3)}, L_I^{(3)}, m_I^{(3)}, M_I^{(3)}) \leftarrow (v_I, L_I, w_I, M_I)$
        $(\ell_I^{(4)}, L_I^{(4)}, m_I^{(4)}, M_I^{(4)}) \leftarrow (\ell_I, v_I, w_I, M_I)$
    for $i \leftarrow K+1, \dots, p^2$ do
        for $j \leftarrow 1, 2, 3, 4$ do
            $(\ell_i^{(j)}, L_i^{(j)}) \leftarrow (\ell_i, L_i)$
    for $i \leftarrow K+1, \dots, q^2$ do
        for $j \leftarrow 1, 2, 3, 4$ do
            $(m_i^{(j)}, M_i^{(j)}) \leftarrow (m_i, M_i)$
    return $\Omega(\ell^{(1)}, L^{(1)}, m^{(1)}, M^{(1)}), \Omega(\ell^{(2)}, L^{(2)}, m^{(2)}, M^{(2)}), \Omega(\ell^{(3)}, L^{(3)}, m^{(3)}, M^{(3)}), \Omega(\ell^{(4)}, L^{(4)}, m^{(4)}, M^{(4)})$
```

Figure 9: This algorithm splits a hyperrectangle Ω into four subrectangles, cf. Fig. 1. The choice of coordinates (x_I, y_I) in the first step is restricted to $I \leq K$ (where K is the number of non-zero singular values of U as in Eq. (12)). The hyperrectangles do not need to be refined along the remaining coordinates because the objective function is linear in these. In particular, the dependence of the convex envelope $\mathsf{Vex}_\Omega F$ on the parameters $\{v_j\}_{j=K+1}^{p^2}$ and $\{w_j\}_{j=K+1}^{q^2}$ coincides with that of the function F irrespective of which hyperrectangle Ω is considered.

```
function BOUNDINGRECTANGLE($Q, \mathcal{S}, \{\eta_j\}_{j=1}^{p^2}, \{\xi_k\}_{k=1}^{q^2}$)
```

Input: Operators $Q \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p \otimes \mathbb{C}^q)$, set $\mathcal{S} \subset \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ defined by SDP constraints orthonormal bases $\{\eta_j\}_{j=1}^{p^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $\{\xi_k\}_{k=1}^{q^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$
Output: $\ell^*, L^* \in \mathbb{R}^{p^2}$ and $m^*, M^* \in \mathbb{R}^{q^2}$ such that $\Gamma(\mathcal{S}) \subset \Omega(\ell^*, L^*, m^*, M^*)$ and Ω is minimal as discussed in Lemma 3.2

```
Find $U_{j,k} = \mathrm{tr}((\eta_j \otimes \xi_k)Q)$  for  $j = 1, \dots, p^2, k = 1, \dots, q^2$
Find singular value decomposition $U = S\Delta T$
for  each $j = 1, \dots, p^2$ do
    Use the SDP solver to compute $\ell_j^* = \inf_{(X,Y)\in\mathcal{S}} \sum_{k=1}^{p^2} S_{k,j}\,\mathrm{tr}(X\eta_k)$
    Use the SDP solver to compute $L_j^* = \sup_{(X,Y)\in\mathcal{S}} \sum_{k=1}^{p^2} S_{k,j}\,\mathrm{tr}(X\eta_k)$
for  each $k = 1, \dots, q^2$ do
    Use the SDP solver to compute $m_k^* = \inf_{(X,Y)\in\mathcal{S}} \sum_{\ell=1}^{q^2} T_{\ell,k}\,\mathrm{tr}(Y\xi_\ell)$
    Use the SDP solver to compute $M_k^* = \sup_{(X,Y)\in\mathcal{S}} \sum_{\ell=1}^{q^2} T_{\ell,k}\,\mathrm{tr}(Y\xi_\ell)$
return $(\ell^*, L^*, m^*, M^*)$
```

Figure 10: The procedure BOUNDINGRECTANGLE which finds a minimal hyperrectangle Ω containing the vectors $\Gamma(\mathcal{S})$. It invokes an SDP solver. Again, the singular value decomposition of Q can be precomputed.

```
function COMPUTEBOUNDSSDP($f, \Omega, \mathcal{S}, \{\eta_j\}_{j=1}^{p^2}, \{\xi_k\}_{k=1}^{q^2}$)
--------------------------------------------------------------------------------
Input: a function $f : \mathbb{R}^{p^2} \times \mathbb{R}^{q^2} \to \mathbb{R}$ of the form (13)
hyperrectangle $\Omega \subset \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$ and
set $\mathcal{S} \subset \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ defined by SDP constraints
orthonormal bases $\{\eta_j\}_{j=1}^{p^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $\{\xi_k\}_{k=1}^{q^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$
Output: $(\underline{\alpha}(\Omega), \overline{\alpha}(\Omega)) \in \mathbb{R}^2$ and $z(\Omega) = (x^*, y^*) \in \Gamma(\mathcal{S}) \cap \Omega$ such that
$\underline{\alpha}(\Omega) \le \min_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} f(x,y) \le \overline{\alpha}(\Omega) = f(x^*, y^*)$
--------------------------------------------------------------------------------
    Define $\hat{\ell}_j^0 = \ell_j$ and $\hat{\ell}_j^1 = L_j$ for $j = 1, \dots, p^2$
    Define $\hat{m}_k^0 = m_k$ and $\hat{m}_k^1 = M_k$ for $k = 1, \dots, q^2$
    Define the operators and scalars

        $G_j^b = \sigma_j \hat{m}_j^b \sum_{k=1}^{p^2} S_{k,j}\eta_k \qquad H_j^b = \sigma_j \hat{\ell}_j^b \sum_{\ell=1}^{q^2} T_{j,\ell}\xi_\ell \qquad s_j^b = \sigma_j \hat{\ell}_j^b \hat{m}_j^b$

for all $b \in \{0,1\}$ and $j = 1, \dots, K$.
    Define the function $G : \mathcal{B}(\mathbb{C}^p) \times \mathcal{B}(\mathbb{C}^q) \times \mathbb{R}^K \to \mathbb{R}$ by

        $G(X,Y,r) = \sum_{j=1}^{K} r_j + \sum_{j=1}^{p^2} a_j \sum_{k=1}^{p^2} S_{k,j} \operatorname{tr}(X\eta_k) + \sum_{\ell=1}^{q^2} b_j T_{\ell,j} \operatorname{tr}(Y\xi_\ell)$

    Invoke the SDP solver to compute

        $(X^*, Y^*, r^*) = \arg\min_{(X,Y,r)} G(X,Y,r)$

subject to the constraints

        $(X,Y) \in \mathcal{S}$
        $\ell_j \le \operatorname{tr}(X \sum_{k=1}^{p^2} S_{k,j}\eta_k) \le L_j$   for   $j = 1, \dots, p^2$   and
        $m_k \le \operatorname{tr}(Y \sum_{\ell=1}^{q^2} T_{k,\ell}\xi_\ell) \le M_k$   for   $k = 1, \dots, q^2$
        $\operatorname{tr}(XG_j^b) + \operatorname{tr}(YH_j^b) - r_j \le s_j^b$   for all $b \in \{0,1\}$   and   $j = 1, \dots, K$ .

    Set $\underline{\alpha}(\Omega) = G(X^*, Y^*, r^*)$
    Set $\overline{\alpha}(\Omega) = f(\Gamma(X^*, Y^*))$
    return $\underline{\alpha}(\Omega)$, $\overline{\alpha}(\Omega)$ and $(x(X^*), y(Y^*))$
```

Figure 11: The subroutine COMPUTEBOUNDSSDP takes as input a set $\Omega \subset \mathbb{R}^{p^2} \times \mathbb{R}^{q^2}$ and a function $f : \Omega \to \mathbb{R}$ as in Lemma 3.1. It further accepts a set $\mathcal{S} \subset \mathcal{B}(\mathbb{C}^p) \times \mathcal{B}(\mathbb{C}^q)$ defined by SDP constraints. It computes a pair $(\underline{\alpha}(\Omega), \overline{\alpha}(\Omega))$ of lower and upper bounds on $\inf_{(x,y)\in\Gamma(\mathcal{S})\cap\Omega} f(x,y)$, where Γ is defined as in Lemma 3.1.

Algorithm 1 Jointly constrained semidefinite bilinear programming

```
**Input:** $\mathcal{S} \subset \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ specified by SDP constraints. $Q \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p \otimes \mathbb{C}^q), A \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p), B \in \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$ determining a function $F : \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p) \times \mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q) \to \mathbb{R}$ as in Eq. (11)
desired precision $\epsilon > 0$.
**Output:** $(X^*, Y^*) \in \mathcal{S}$ and $\overline{\alpha}$ such that Eq. (20) holds

Fix bases $\{\eta_j\}_{j=1}^{p^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^p)$ and $\{\xi_k\}_{k=1}^{q^2}$ of $\mathcal{B}_{\mathsf{sa}}(\mathbb{C}^q)$
Set $(\ell^*, L^*, m^*, M^*) = $ BoundingRectangle$(Q, \mathcal{S}, \{\eta_j\}_{j=1}^{p^2}, \{\xi_k\}_{k=1}^{q^2})$
Define $D$ as the hyperrectangle $D = \Omega(\ell^*, L^*, m^*, M^*)$
Set $(\underline{\alpha}(D), \overline{\alpha}(D), z(D)) = $ ComputeBoundsSDP$(f, D, \mathcal{S}, \{\eta_j\}_{j=1}^{p^2}, \{\xi_k\}_{k=1}^{q^2})$
Set $\mathcal{P} = \{D\}$, $\underline{\alpha}(\mathcal{P}) = \underline{\alpha}(D)$ and $\overline{\alpha}(\mathcal{P}) = \overline{\alpha}(D)$
**while** $\overline{\alpha}(\mathcal{P}) - \overline{\alpha}(\mathcal{P}) > \epsilon$ **do**
    Find (any) hyperrectangle $\Omega \in \mathcal{P}$ such that $\underline{\alpha}(\mathcal{P}) = \underline{\alpha}(\Omega)$
    Set $(x, y) = z(\Omega)$
    $(\Omega^{(1)}, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)}) = $ BranchHyperrectangle$(\Omega, (x, y))$
    **for** $j \leftarrow 1, 2, 3, 4$ **do**
        Set $(\underline{\alpha}(\Omega^{(j)}), \overline{\alpha}(\Omega^{(j)}), z(\Omega^{(j)})) = $ ComputeBoundsSDP$(f, \mathcal{S}, (\Omega^{(j)}))$
    Update $\mathcal{P} \leftarrow (\mathcal{P} \backslash \{\Omega\}) \cup \{\Omega^{(1)}, \Omega^{(2)}, \Omega^{(3)}, \Omega^{(4)}\}$
    Compute $\underline{\alpha}(P) = \min_{\Omega \in \mathcal{P}} \underline{\alpha}(\Omega)$ and $\overline{\alpha}(P) = \min_{\Omega \in \mathcal{P}} \overline{\alpha}(\Omega)$
Find (any) hyperrectangle $\Omega \in \mathcal{P}$ such that $\overline{\alpha}(\Omega) = \overline{\alpha}(\mathcal{P})$
Terminate and **return** ComputeOperator$(z(\Omega))$, $\overline{\alpha}(\mathcal{P})$
```

Figure 12: The algorithm for jointly constrained biconvex programming. The key modification compared to the biconvex programming algorithm from Section 2.1.2 is the use of the subroutine ComputeBoundsSDP, which uses an SDP solver to establish bounds on the objective function.