# UNCERTAINTY RELATIONS

IN

QUANTUM THEORY

by RENÉ SCHWONNEK

# Uncertainty Relations in Quantum Theory

Von der Fakultät für Mathematik und Physik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades

„Doktor der Naturwissenschaften“
(Dr. rer. nat.)

genehmigte Dissertation von

M.Sc. René Schwonnek

2018

| | | |
|---|---|---|
| Referent: | Prof. Dr. Reinhard F. Werner, | Leibniz Universität Hannover |
| Korreferenten: | Prof. Dr. Paul Busch, | University of York |
| | Prof. Dr. Tobias Osborne, | Leibniz Universität Hannover |
| | Prof. Dr. Michael Wolf, | Technische Universität München |

Tag der Promotion: 11.07.2018

*Felix qui potuit rerum cognoscere causas.*
-Publius Vergilius Maro

*Wer misst, misst Mist.*
- Werner Heisenberg[1]

[1] Zitat nicht nachweisbar

# Abstract

KEYWORDS: Preparation uncertainty, measurement uncertainty, entropic uncertainty, joint measurements, multicriterial convex optimization

Uncertainty relations are commonly praised as one of the central pillars of quantum theory. Usually, they are taught in the first weeks of a beginners lecture, and introduced in the first chapters of a textbook. However, their precise operational meaning and a formulation in a general context, i.e. beyond the example of position and momentum observables, are often left out. The reasoning for this is twofold:

On one hand, an exact operational definition of uncertainty, indeterminacy and a corresponding uncertainty principle has been the content of many debates since the early days of quantum mechanics until today. From a modern perspective, we have the consent that there are at least the two notions of preparation and measurement uncertainty: the first notion prohibits the existence of dissipation free states and the latter one the existence of error free joint measurements.

On the other hand, we have that, the mathematical tools, which are needed for comprehensive treatment of uncertainty relations in a general context, are still under development and usually go far beyond the mathematical level of an introductory course.

In this thesis we will investigate these two notions of uncertainty, their corresponding uncertainty relations, such as their interplay. The aim of this thesis is to give answers to the central questions:

(1.) Which quantities should be used to formulate uncertainty?

(2.) How can we compute uncertainty relations for those?

(3.) Are there connections between the two notions of uncertainty?

We will do this, whenever possible, in a most general context and with a focus on relevant examples, otherwise. Therefore, we will consider constructions of measurement errors and deviation measures that quantify uncertainty, based on, so called, cost functions. Commonly used uncertainty measures like variances, entropies, and the Hamming distance are examples for these. We will investigate the structure of the corresponding uncertainty relations and provide several methods that enable us to compute them. The third question is addressed by a theorem that shows, for sharp observables, that measurement uncertainty relations can be lower bounded by preparation uncertainty relations, whenever the same cost function is used.

# Kurzzusammenfassung

SCHLAGWORTE: Preparationsunschärfe, Messunschärfe, entropische Unschärfe, gemeinsame Messungen, multikriterielle konvexe Optimierung

Unschärferelationen sind ein zentraler Grundpfeiler der Quantentheorie. Üblicher Weise, werden sie bereits in den ersten Wochen einer Grundvorlesung und in den ersten Kapiteln eines Lehrbuches über Quantenmechanik eingeführt. Eine präzise Erläuterung ihrer operationellen Bedeutung sowie eine Formulierung in einem allgemeinen Kontext, d.h. jenseits des Beispiels von Ort und Impuls, wird jedoch meistens ausgelassen. Hierfür sind die folgenden zwei Gründe zu vermuten:

Zum einen ist eine exakte Definition der Begrifflichkeit von Unschärfe, Unbestimmtheit, und eines diesbezüglichen Prinzips, der Inhalt vieler Debatten, seit den frühen Tagen der Quantentheorie bis heute. Aus heutiger Sicht herrscht hierzu der Konsens, dass, in jedem Fall, zwischen den Begriffen Preparations Unschärfe und Messunschäfe unterschieden werden muss. Hierbei verbietet der erste Begriff die Existenz von nicht dissipartiven Zuständen und der zweite die Existenz von fehlerfreien gemeinsamen Messungen. Zum anderen, befinden sich die, für eine zufriedenstellende Behandlung von Unschärferelationen notwendigen, mathematischen Methoden noch in Entwicklung und überschreiten sicherlich den Anforderungshorizont einer Einführungsvorlesung.

In dieser Arbeit werden die beiden obigen Begriffe von Unschärfe, die zugehörigen Relationen, sowie ihre Zusammenhänge, untersucht. Hierbei besteht das Ziel antworten auf die folgenden zentralen Fragestellungen zu geben:
(1.) Welche Größen erlauben eine quantitative Formulierung von Unschärfe?
(2.) Lassen sich die entsprechenden Unschärferelationen berechnen?
(3.) Gibt es Verbindungen zwischen den beiden Unschärfebegriffen?
Diese Fragen werden, wann immer dies möglich ist, in einem allgemeinen Kontext, und ansonsten mit einem Fokus auf relevante Beispiele, beantwortet werden. Hierzu werden Konstruktionen von Messfehlern und Streumaßen, die auf, so genannten, Kostenfunktionen fußen, betrachtet. Üblicher Weise benutzte Unschärfemaße, wie Varianzen, Entropien und Hamming-Abstände sind Beispiele hierfür. Im Detail werden die Struktur der zugehörigen Unschärferelationen und Methoden zur Berechnung dieser, betrachtet. Die dritte der obigen Fragen wird mit einem Theorem beantwortet werden das aussagt, dass, für scharfe Observable, Preparationsunschärferelationen untere Abschätzungen für Messunschärferelationen liefern, wenn beide Größen bezüglich der gleichen Kostenfunktion betrachtet werden.

# Acknowledgements

First and foremost I want to thank my advisor Reinhard F. Werner. He has taught me, by example, how a good scientist has to work and think. I appreciate all contributions of time, ideas, and funding to make my Ph.D. experience so productive.

I thank my fellow groupmates for all the stimulating discussions, for the sleepless nights we were working together in order to finish our projects, and for all the fun we have had in the last five years. In particular, I am grateful to Kais Abdelkhalek, Leander Fiedler, Lars Dammeier and David Reeb, for all the time they spend in our joint works.

I would also like to thank my other collaborators: Adrian Auer, Guido Burkard, Wissam Chemissany, Jörg Duhme, Berge Englert, Fabian Furrer, Hans Maassen, Gianpiero Mangano, Philippe Raynal and Chrisian Schoder for all the fruitful work.

Furthermore, I would like to thank all my colleagues and friends from other institutions for all the great exchanges, discussions and new insights they provided me over the years. I am especially grateful to: Silvestre Abruzzo, Adrian Auer, Ana Costa-Sprotte, Otfried Gühne, Anna L./K. Hashagen, Timo Holz, Hans Maassen, Cornelia Spee, and Giuseppe Vitagliano.

There were many people who granted me a lot of help, advice, and open ears for finishing this thesis. Thanks to Coco, Inken, Kais, Lars, Louis and Reinhard.

Last but not the least, I would like to thank my family and my girlfriend: my parents, my brother, my sister and Coco for supporting me throughout writing this thesis and my life in general. I would especially like to thank my grandmother for providing me with warm food, words of wisdom and a car.

René Schwonnek

# Contents

## Publications

1. L. Dammeier, R. Schwonnek, and R. F. Werner: *Uncertainty relations for angular momentum*, 2015, New Journal of Physics 17 (9),093046.

2. [ASM$^+$15]
   K. Abdelkhalek, R. Schwonnek, H. Maassen, F. Furrer, J. Duhme, P. Raynal, B.-G. Englert, and R. F. Werner: *Optimality of entropic uncertainty relations*, 2015, International Journal of Quantum Information 13 (06), 1550045.

3. A. Auer$^*$, R. Schwonnek$^*$, C. Schoder, L. Dammeier, R. F. Werner, and G. Burkard, *Entanglement distillation using the exchange interaction*, 2016, Applied Physics B 122 (3), 51.

4. [SRW16]
   R. Schwonnek, D. Reeb, and R. F. Werner, *Measurement uncertainty relations*, 2016 Mathematics 4 (2),38.

5. [ACF$^+$16]
   K. Abdelkhalek, W. Chemissany, L. Fiedler, G. Mangano, and R. Schwonnek, *Optimal uncertainty relations in a modified Heisenberg algebra*, 2016, Physical Review D 94 (12), 123505.

6. [SDW17]
   R. Schwonnek, L. Dammeier, and R. F. Werner: *State-independent uncertainty relations and entanglement detection*, 2017,
   Physical Review Letters 119.170404

7. [S18]
   R. Schwonnek, *Additivity of entropic uncertainty relations*, 2018,
   Quantum 2, 59

8. R. Schwonnek, and R. F. Werner: *Wigner distributions for $n$ arbitrary operators*
   arXiv:1802.08342

9. R. Schwonnek, and R. F. Werner: *Properties of Wigner distributions for $n$ arbitrary operators*
   arXiv:1802.08343

Publications labelled by author abbreviation and year (e.g. [XYZ 123]) are part of this thesis.

# CHAPTER 1

# Introduction

*'Nebenbei sind auch einige Punkte der mathematischen Struktur etwas ausführlicher dargelegt;' - E. H. Kennard*

The existence of unavoidable uncertainties and incompatibilities in a quantum measurement process is clearly one of the most characteristic implications of the quantum theory. These limitations mark one of the sharp lines that allows us to distinguish quantum from classical physics.

From a mathematical perspective, the origin of uncertainty can be traced back to the interplay of the quantum mechanical state space, which is, in contrast to classical theories, not a simplex, with the set of quantum measurements. However, from a practical perspective, this interplay can usually not be adequately captured, neither analytically or numerically. Here, uncertainty relations come into play, yielding a simplified picture that still captures the necessary information on 'goes' and 'no-goes' that are needed in applications like for example security proofs or the detection of non-local correlations.

In this thesis, we will omit considerations to the philosophical implications of uncertainty for the nature of reality, the reality of nature, and the role of an observer within these. Different perspectives to those questions can be found in [BE67, Pop89] and in [Hei44].

To avoid those considerations here, we will restrict our view on quantum theory to a minimal, operational, rather technical, perspective:

From this perspective every experiment in quantum mechanics can be abstracted to the fundamental steps of state preparation and measuring, see Fig. 1.1. Thereby, a state is handled as description for a repeatable preparation procedure and a measurement as a 'black box' that maps states to probability distributions of measurement outcomes.

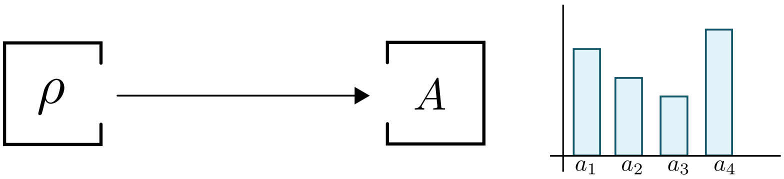


Figure 1.1: Basic setting of an experiment in quantum mechanics: A state $\rho$ is measured on a device $A$. Repetitive measuring gives rise to a probability distribution.

The only way in which we can influence an experiment is by choosing state preparations and measurements and the only observations we can make are the particular outcomes of a measurement and their respective statistics.

Within this spirit, we will aim to bind every quantity that characterises uncertainties and uncertainty relations to probability distributions of measurement outcomes. Thereby, we will have in mind that we have to take care of the individual structure a particular outcome set has. As a consequence, we can not assign a single universal quantity that meaningful quantifies uncertainty in any situation. Rather than that we will start by considering general classes of uncertainty measures and have a look at particular examples afterwards.

## 1.1 Two notions of uncertainty

In this thesis we will distinguish the following two fundamental notions of uncertainty:

### Preparation uncertainty

The first notion, the one that has been considered in the literature mostly, and the one that is introduced in any textbook or beginners course on quantum mechanics, is *preparation uncertainty*. Here, we are interested in classifying the spread/the deviation of the statistics obtained by measuring an observable $A$ on a state $\rho$.

Whenever a non-zero spread is present in a particular statistic, we will say that $\rho$ has an uncertainty with respect to the measurement $A$. Depending on the context, typically variances $\Delta_\rho^2 A$ or entropies $H(A|\rho)$ are used to quantify this spread. However, in this thesis we will also introduce a more general class of deviation measures $\nu(A|\rho)$, for which variances and entropies are examples, that is based on what we will call a cost function.

We can compare different measurement devices, say $A$ and $B$, with respect to their preparation uncertainty by the type of experiment that is sketched in Fig. 1.2. Here, we test each device separately by individual copies of the same state $\rho$. The individual shots of such a test are performed independently and no correlations between them are introduced. Hence, all measurements could either be performed at different times or locations, or also simultaneously and in the same lab.

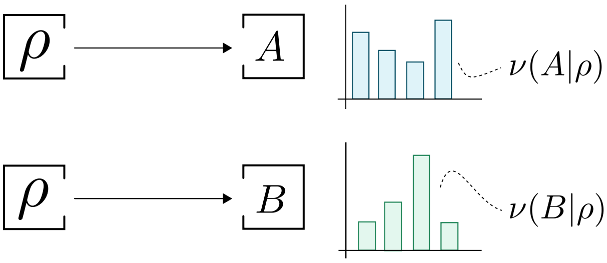


Figure 1.2: Elemental scenario for a test of preparation uncertainty. Two instances of the same state $\rho$ are measured with ideal· devices $A$ and $B$. The deviation $\nu(A|\rho)$ is compared to the deviation $\nu(B|\rho)$

At the end of a test-round we obtain a pair of deviations $(\nu(A|\rho), \nu(B|\rho))$. Here, a *preparation uncertainty relation* describes the relation between $\nu(A|\rho)$ and $\nu(B|\rho)$. More precisely, an uncertainty relation restricts the values, that $\nu(B|\rho)$ can attain, when we fix the value of $\nu(A|\rho)$, and vice versa. In practice we have to perform a *minimization over all quantum states* in order to get such a relation. The prototype for a preparation uncertainty relation is Kennard's inequality [Ken27] for position and momentum:

$$\Delta_\rho^2 Q \Delta_\rho^2 P \geq \frac{\hbar^2}{4} \tag{1.1}$$

Here, we can directly infer that for all states, that attain a $\Delta_\rho^2 Q$ close to zero, $\Delta_\rho^2 P$ has to be very big. In general the aim of a preparation uncertainty relation is to quantify statements like:

'there is no state $\rho$ such that
the deviations of $A$ and $B$ measurements
are simultaniously small'.

We note that, in the above setting, we do not have to specify the 'state after a measurement', since no particular instance of a state is measured twice. Therefore, this notion gives *no* description of folkloric, and in this context misplaced, statements like:

$$\begin{aligned} &\text{'an } A \text{ measurement, with uncertainty } \Delta^2_\rho A, \\ &\text{disturbs a } B \text{ measurement by } \Delta^2_\rho B\text{' .} \end{aligned} \tag{1.2}$$

Statements of this kind are captured as subcase of the following notion of uncertainty:

### Measurement uncertainty

The second notion, *measurement uncertainty*, is the one most discussed and debated since the early days of quantum theory until today. In this thesis we will adapt the position of [BLW14b] and treat measurement uncertainty as the operational counterpart of preparation uncertainty:

Here, we are interested in finding devices that can perform a $A$ and a $B$ measurement jointly in one shot. This is, if $A$ and $B$ are measurements with outcomes on the sets $\Omega_A$ and $\Omega_B$, we will consider measurement devices, $R$, with outcomes on the joint outcome set $\Omega_A \times \Omega_B$. For every state $\rho$, such a device will give us, in every shot, a tuple $(a_i, b_j)$, from which we will interpret the $a_i$ component as $A$-type outcome and the $b_j$ component as $B$-type outcome. Throughout this thesis we will use the convention to denote the corresponding marginal observables, i.e. the restriction of $(a_i, b_j)$ to $a_i$ or $b_j$, by $A'$ and $B'$.

In general, more precisely: if $A$ and $B$ are *incompatible*, there is no device which achieves this task without an inavoidable inprecession. Therefore, we will have to classify the proximity between measurement devices, i.e. this inprecession, by introducing error measures $\varepsilon(A|A')$.

Whenever such an error is present, we will say that a device $A'$ has a measurement uncertainty with respect to the measurement $A$.

Those error measures are constructed in two steps: Firstly we will test both devices, $A$ and $A'$, as for preparation uncertainty, on instances of a common state $\rho$ (see Fig. 1.3), and compare the resulting probability distributions. In order to avoid trivialities, we want that those probability distributions are similar for all states on which we test. Hence, we will, secondly, take either the mean value or consider a worst case in order to judge the proximity of our devices.

Note that, for measurement devices in a classical world, tests in comparable situations would be performed in a slightly different way: Here, the measured quantities, think for example of the position and the momentum of a macroscopic

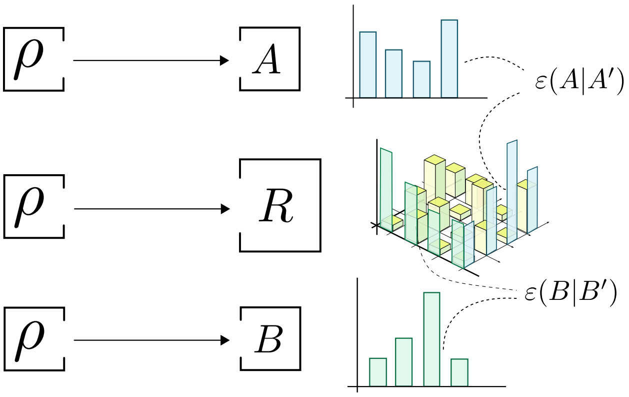


Figure 1.3: Elemental scenario for a test of measurement uncertainty. A joint measurement $R$ with marginals $A'$ and $B'$ is used to approximate the ideal observables $A$ and $B$. The proximity/error for $A$ to $A'$ is compared to the proximity/error of $B$ to $B'$

object, have a well defined value. Hence, we can compare the outcomes of an approximate measurement device with this 'correct' value and judge its quality by the corresponding deviation. This does even not change if we test our approximate device on a probabilistic ensemble of different macroscopic objects. Since we can still assign a 'correct' measurement outcome to each single shot, we are able to distinguish between the spread of the correct values within the ensemble and a spread within the outcome statistic, originated by a disturbativ measurement process.

This picture does not carry over to the quantum world: the existence of preparation uncertainty relations prohibits us to assign a corresponding 'correct' value to an arbitrary quantum state. However, when we are in the special situation that our ideal measurements $A$ and $B$ have eigenvectors corresponding to deterministic outcomes of a respective measurement we can mimic a classical test. Here, we test $A'$ only on those eigenstates and want that a 'good' approximation shows an almost 'correct' value at least in these cases.

In order to speak about *measurement uncertainty relations*, two ideal measurements are compared with a joint measurement device $R$ as sketched in Fig. 1.3. Here it is important to have in mind that the error quantities $\varepsilon(A|A')$ and $\varepsilon(B|B')$ are evaluated on different and independent states. This is, the worst case state that leads to $\varepsilon(A|A')$ does usually not coincide with the state that attains the worst case for $\varepsilon(B|B')$. Again we want to restrict the possible values, that $\varepsilon(A|A')$ can attain, when we fix $\varepsilon(B|B')$, and vice versa, in order to get statements like:

'there is no joint measurement $R$ such that
the marginals $A'$ and $B'$ are good approximations
for the measurements $A$ and $B$ simoultaniously$'$.

At the end a statement like the above demands us to perform a minimization of the error quantities with respect to all joint measurements. A prototype for a measurement uncertainty relation, again between position and momentum, is provided in [BLW14b]:

$$\Delta_M^2(Q, Q')\Delta_M^2(P, P') \geq \frac{\hbar^2}{4}. \tag{1.3}$$

Here, the quantity $\Delta_M^2(Q, Q')$, which is the worst case Wasserstein-2 distance between outcome distributions, is used as error measure. Due to a high symmetry in the interplay of the underlying observables, to which we will comment in this thesis in chapter Ch. 3 later on, the bounds of the r.h.s. of (1.3) and (1.1) coincide.

## 1.2 Previous works

In the following, we will only briefly wrap up the historical development of the aforementioned notions of uncertainty. A full review of the research history is especially difficult because of the following reasons: Uncertainty is regarded as a fundamental concept. Hence, many authors contribute with many different perceptions, definitions and conclusions. Furthermore, uncertainty as a 'principle', but not as a mathematical theorem is employed in many hand-wavy explanations and even calculations. This is especially easy, because there are many illustrative but insufficient explanations around. An example for this is the analogy to classical waves. Because uncertainty is so fundamental, there is no concept (beside the mathematical foundations of quantum theory) it can be based on, or checked against. The transition from a physicists intuition to a mathematical theorem is hard in general and the uncertainty principle is no exception to this.

For an overview of the existing literature, we will limit ourselves at this place to mention only Heisenbergs initial contribution, its common textbook formulation (1.1), and recent developments.

### Historical development

A comprehensive review of historical backgrounds are given for example in [BR18, BHL07, Bus85, BLPY16]. An analysis of Heisenbergs intention, such as a critical discussion of its recent receptions by [Oza13] and [BLW13] can be found in [HU16], see also [Wer17].

In his seminal paper [Hei27] Heisenberg introduced what was later on phrased *the uncertainty principle.* On p.3-4 he wrote:

*'also je genauer der Ort bestimmt ist, desto ungenauer ist der Impuls bekannt und umgekehrt;'*

*'That is, the more precisely the position is determined, the more imprecisely will the impulse [momentum] be known, and vice versa.'* (translation of [Hei27] taken from [Hei84]) ,

which is some lines later expressed by the heuristic relation

$$p_1 q_1 \sim \hbar.$$

Here, $p_1$ and $q_1$ are introduced as '*approximately the average error*', [Hei84], of the position $q$ and the momentum $p$, and the symbol '$\sim$' is explained as 'being on the order of'. Heisenberg explains the intuition behind this relation by introducing his famous $\gamma$-ray microscope as a gedankenexperiment, see [BR18] for a comprehensive description. However, an explicit mathematical definition of the above quantities was left out. Some months after Heisenbergs publication, a possible definition was given by Kennard [Ken27], and independently by Weyl [Wey28], who parsed $q_1$ and $p_1$ by standard deviations $\Delta_\rho P$ and $\Delta_\rho Q$ in order to derive the inequality (1.1), which is today phrased in most textbooks as Heisenberg's uncertainty principle. Although, Heisenberg seems to have accepted this as the mathematical formulation of his ideas, the debate of the precise formulation of the uncertainty principle was not settled by this and is still ongoing.

First of all, we can see that Kennard's formulation (1.1) clearly describes the setting of preparation uncertainty. By its mathematical definition the expression $\Delta_\rho Q$ refers to a full position measurement and, hence, does not coincide with the 'tuning of a wavelength' in a $\gamma$-ray microscope. This interpretation was already present in the community for a long time phrased as *minimal interpretation*, see for example Ballentine's comment [Bal69] who refers to Margenau's contribution [MC67] in the collection [BE67].

However, popular receptions of Heisenberg's statement are usually of the form (1.2), which clearly refers to a limitation on joint measurements. From our perspective, the $\gamma$-ray microscope can be regarded as prototype for this. The idea to use joint measurements, in order to mathematically formalize an uncertainty principle, goes back to the 80's, see for example [Bus85]. A discussion of the two notions and an introduction to joint measurements can be found in [BHL07].

At this point we should note that, in the past, and especially within the last decade, there have been many other attempts to formalize an uncertainty principle based on Heisenberg's original work.

In conclusion, we can not assign a unique mathematical interpretation to Heisenberg's work. However, the notion of measurement uncertainty based on joint measurements, as it is used in this work, is surely interesting by its own, as it has a clear operational interpretation and mathematical description. Furthermore, the optimizers in a corresponding measurement uncertainty yield a blueprint for useful devices in regard of many applications.

### Recent developments

The following collection of previous works should give an overview on recent contributions to this field, it does not have the aspiration of being complete or selected solely by importance:

**Preparation uncertainty:** A huge part of the literature on preparation uncertainty, that appeared in the last years, unfortunately, puts its focus on state-dependent relations, and thereby mostly on extensions of Robertson's [Rob29] and Schrödinger's [Sch30] inequality. In the next section, we will see that those relations do not fit our criteria for an uncertainty relation, since they do not give rise to state-independent statements. A convincing argumentation for state-independence was given by Deutsch's [Deu83] in the 80's.

Variance based state-independent uncertainty relations for spin measurements can be found for example in [HT03, Gü04, SR95, HPDR11, DSW15] such as in the recent works [SZ18, dGMSS18]. A special case of spin-measurements are qubits, here the most optimizations can be solved analytically, see [AAHB16] for variances and [GMR03] for entropies. Entropic uncertainty relations for spin measurements also have been considered in [SR98, RMM17b, DSW15] such as in, [GL04, Gü04, XGM$^{+}$17], with the scope of an application to entanglement detection.

A connection between entropic and variance based uncertainty relations is made in [Hua12]. A profound overview on entropic preparation uncertainty is provided by the reviews [WW10, CBTW17] and the references therein.

Uncertainty relations for continuous measurements have been considered mainly in the context of momentum and position observables. Beside Kennard's fundamental result (1.1), further directions of investigation are sums of phasespace observables [KW14, KW16] and representations of modified Heisenberg algebras [KMM95, KM97].

**Measurement uncertainty:** Due to its connection to Heisenbergs original formulation many works consider measurement uncertainty relations for position and momentum. This thesis stands in the line of [Wer04, BLW14b, BLW13]. Alternative error measures were proposed in [App98] and also in [Oza04]. An alternative error measure that can be computed efficiently for all finite observables is the so called *white noise robustness* see [HKR15].

Measurement uncertainty relations for arbitrary dichotomic measurements on qubits can be computed analytically, see [BB15, YO14, BLW14a]. An analytic result on uncertainty relations that bound the worst error of a spin measurement in arbitrary directions can be found in [DSW15]. Relations for qubit-strings, with error measures based on Hamming distance, were investigated in the bachelor thesis [Jav16] and relations, with error measures based on the discrete metric, can be found in the bachelor thesis [Fra15]. In the bachelor thesis [Kus16] uncertainty relations for position and momentum measurements with a finite operating range were investigated. An application of measurement uncertainty as security criterion for, so called, data locking protocols was established in the master thesis [Kö17].

Uncertainty relations and error-disturbance relations in terms of conditional entropies have been considered in [BHOW14] and in [AB16], with explicit results for the case of qubits. In [BGT18, BGT17] measurement uncertainty in terms of the relative entropy is considered. However, here the authors do not regard independent error measures $\varepsilon(A|A')$ and $\varepsilon(A|A')$, rather than they provide an *uncertainty index*, which is obtained by measuring all three devices from Fig. 1.3 on the same state.

Closely related to measurement uncertainty is the characterization of an error-disturbance trade-off. Here, the interest is focused on finding quantum instruments, which provide a measurement with a small error, on one hand, and a minimal disturbed state after the measurement, on the other. A framework in terms of cb-norms is provided in [RSH16]. Computable optimal bounds for the error-disturbance trade-off can be found in [HW18].

Prototypes for joint measurements are cloning devices, where the ideal observables can be measured afterwards. Optimal cloners are considered in [Wer98], optimal asymmetric cloners, i.e. those who establish a whole trade-off relation, can be found in [Has17]. The special class of phase-space covariant asymmetric cloners was considered in [CF05]. It turns out that these cloning devices coincide with the optimal joint measurements of position and momentum from [BLW14b].

## 1.3 Preliminaries

### Measurements

Measurements are the central objects of this thesis. We will encounter them as ideal measurements, i.e. as input to an optimization problem, as well as approximate joint measurement, i.e. as output of an optimization. Throughout this thesis the terms measurement, (measurement-) device, and observable are used interchangeably. Mathematically, all three terms refer to the same class of objects, namely positive operator valued measures (POVMs), which will be introduced immediately. However, we will assign a slight distinction to those terms by their connotation: We will use the term *measurement* in the general case. The term *(measurement-) device* puts an obvious reference to an actual implementation of a measurement. Hence, it is used more likely in the context of explaining an operational interpretation. The term *observable*, which is the term most commonly used in textbooks, has a reference to quantities with a counterpart in classical physics, like position-observable or phase-space observables.

- **POVMs:** In a single shot of a quantum measurement a quantum state $\rho$, given as density operator on a Hilbert space $\mathcal{H}$, is mapped to a measurement outcome $x$, originating from a set $\Omega$. In order to describe the statistics, that arise when we go beyond single shots, *positive operator valued measures* (POVMs) are used to describe the correspondence of $\rho$ to a probability measure $\mu_\rho$ defined on $(\Omega, \mathcal{A}_\Omega)$, where $\mathcal{A}_\Omega$ denotes a sigma algebra on $\Omega$.

  We interpret a linear combination $\rho = \lambda\xi + (1-\lambda)\sigma$ as statistical mixture of the states (preparation procedures) $\xi$ and $\sigma$, which, when measured, should result in a corresponding mixture $\lambda\mu_\xi + (1-\lambda)\mu_\sigma$ of outcome statistics. Hence, the mapping $\rho \mapsto \mu_\rho$ has to be linear. For a measurement $A$, we will use the notation/representation

  $$\mu_\rho(\omega) = \mathrm{tr}(A[\omega]\rho) = \int_\omega \mathrm{tr}(A[d\omega]\rho)$$

  to denote the measure of an $\omega \in \mathcal{A}_\Omega$, evaluated on $\rho$. Hereby, $A[\omega]$ is a bounded operator on $\mathcal{H}$. Hence, we understand the mapping $\omega \mapsto A[\omega]$, in this sense, as operator valued measure. In order to ensure that $\mu_\rho$ is a probability measure, for all $\rho$, we have the properties

  $$\mu_\rho(\Omega) = 1 \mapsto A[\Omega] = \mathbb{I}_\mathcal{H} \text{ and } 0 \leq \mu_\rho(\omega) \leq 1 \mapsto 0 \leq A[\Omega] \leq \mathbb{I}_\mathcal{H}.$$

In the most parts of this thesis, we will encounter measurements with finite outcome sets $\Omega = \{\omega_1, \ldots, \omega_n\}$ labelled by $i \in I = \{1, \ldots, n\}$. In this case, a measure is fully described by fixing $\{A[\omega_1], \ldots, A[\omega_n]\}$. Whenever, the outcome set $\Omega$ is clear from the context, we will use the abbreviation

$$A(i) := A[\omega_i]$$

and call $A(i)$ the *POVM element* corresponding to $\omega_i$. Here we have to respect the properties

$$0 \leq A(i) \leq \mathbb{I}_{\mathcal{H}} \quad \text{and} \quad \sum_{i \in I} A(i) = \mathbb{I}_{\mathcal{H}}$$

in order to obtain a probability measure on $\Omega$.

- **Sharp and projective measurements:** We call a measurement *projective* if the operator part of the corresponding POVM has its support only on projection operators, i.e. when we have

$$A[\omega]^2 = A[\omega] \quad \forall \omega \in \mathcal{A}_{\Omega}.$$

For finite $\Omega$, this directly implies that we have

$$A(i)^2 = A(i) \text{ and } A(i)A(j) = 0 \text{ for } i \neq j$$

In extension, we call a finite measurement a *sharp measurement* if all POVM elements $A(i)$ are one-dimensional projections, i.e.

$$\text{rank}(A(i)) = 1$$

In this case we can assign a unique eigenvector $|\phi_i\rangle$ to the outcome $i$ by

$$A(i) = |\phi_i\rangle\langle\phi_i|.$$

In this thesis we will use the common abbreviation $\phi_i := |\phi_i\rangle\langle\phi_i|$ to denote the eigenstate $\phi_i$. These eigenstates will play an important role in the following, because they correspond to an input to a measurement $A$ with the deterministic outcome $\omega_i$. Thereby we have that this outcome $i$ results with certainty *only* for the particular input $\phi_i$. Later on we will use exactly this property to define ensembles of *test states*, in order to judge the quality of an approximation $A'$ to $A$.

- **joint measurements:** Consider a collection of 'ideal measurements' $A_1, \dots, A_n$ with outcome sets $\Omega_1, \dots, \Omega_n$. Throughout this thesis, a measurement with outcomes on $\Omega_1 \times \dots \times \Omega_n$ will be denoted as *joint measurement* of $A_1, \dots, A_n$. We will use the convention to denote such a joint measurement by $R$ and the set of all joint measurements by $\mathcal{R}$. Any joint measurement can be reduced to a measurement $A'_j$ on $\Omega_j$ by reducing it to the corresponding marginal measure, i.e. for $\omega^j \in \mathcal{A}_{\Omega_i}$ we take

  $$A'[\omega^j] = R[\Omega_1 \wedge \dots \wedge \omega^j \wedge \dots \wedge \Omega_n].$$

  For finite outcome sets, the POVM elements of the $j$-th marginal measurement $A'_j$ are given by

  $$A'_j(i_j) = \sum_{i_1,\dots i_n | i_j} R(i_1, \dots, i_n).$$

  We will call a joint measurement *approximative* if the marginal POVMs differ from the ideal ones,

  $$A_j(i) \neq A'_j(i),$$

  and a *perfect* joint measurement otherwise. A collection of measurements for which a perfect joint measurement exists is also called *compatible*, and *incompatible* otherwise. A basic theorem on compatibility [BLPY16] states that: A sharp measurement $A$ is compatible with some measurement $B$ if and only if

  $$[A(i), B(j)] = 0$$

  holds for all $i, j$. Furthermore, we have that the respective perfect joint measurement admits the product form

  $$R(i,j) = A(i)B(j).$$

  The extension of this statement to general measurements fails, i.e. there are compatible measurements with non-commuting POVM elements, see for example [RRW13].

### Uncertainty relations as multicriterial optimum

One major part of this thesis is devoted to define and compute measures, like errors $\varepsilon$ for measurement uncertainty, or deviations $\nu$ for preparation uncertainty, that quantify the uncertainty of a particular state or an explicit approximative device with respect to an ideal measurement.

- **Uncertainty regions:** Assume we have given such a measure and a collection of ideal measurements $A_1, \ldots, A_n$. Evaluating $\nu$ or $\varepsilon$ on a particular state or joint measurement will give us a tuple of errors

  $$\big(\varepsilon(A_1|A_1'), \ldots, \varepsilon(A_n|A_n')\big)$$

  or a tuple of deviations

  $$\big(\nu(A_1|\rho), \ldots, \nu(A_n|\rho)\big).$$

  Here we can think, at least in principle, of evaluating $\nu$ or $\varepsilon$ on all possible states or all joint measurements, which will provide us with a set $\mathcal{U} \subset \mathbb{R}^n$, called *uncertainty region*, i.e.

  $$\mathcal{U}_\varepsilon := \{\varepsilon(A_1|A_1'), \ldots, \varepsilon(A_n|A_n')|\mathcal{R} \text{ is a joint measurement}\}$$

  or

  $$\mathcal{U}_\nu = \{\nu(A_1|\rho), \ldots, \nu(A_n|\rho)|\rho \text{ is a quantum state}\},$$

  which contains all information on the respective uncertainty between our ideal measurements.

- **Two measurements:** For a start, let us consider the uncertainty between two ideal measurement, i.e. an uncertainty region $\mathcal{U} \subset \mathbb{R}^2$. An example for such a set is sketched in Fig. 1.4, here the white space around the origin indicates the existence of an uncertainty trade-off. More precisely, we can observe:

  $$\begin{gathered}\text{for all } u = (u_1, u_2) \in \mathcal{U}: \\ \text{whenever } u_1 \text{ is small, } u_2 \text{ has to be big, e.v.v.}\end{gathered} \tag{1.4}$$

  The purpose of an *uncertainty relation* is to give a quantitative description for this. At the end, this description may come as a monotone increasing

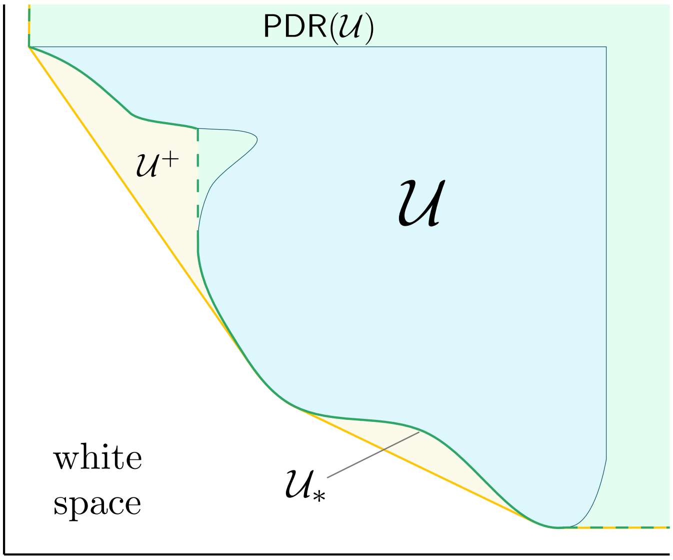


Figure 1.4: A uncertainty region $\mathcal{U}$ with its Pareto dominated region PDR($\mathcal{U}$) and its positive convex hull $\mathcal{U}^+$. An optimal tight uncertainty relation is described by all points from $\mathcal{U}$ that are placed on the strong Pareto boundary $\mathcal{U}_*$ (thick green line). Optimal linear uncertainty relations are described by the the lower boundary of the positive convex hull (thick yellow line).

function $f_2$, which allows us to impose restriction on $u_1$ based on an exclusive assumption on $u_2$. With this we would be able to conclude:

$$\forall u \in \mathcal{U} : u_1 \geq f_2(\alpha) \text{ if } u_2 \leq \alpha, \tag{1.5}$$

Here, we clearly have a pointwise biggest function $f_2$, to which we will refer as *optimal uncertainty relation.*

We could also have aimed to make the above statement the other way around, i.e. put restrictions on $u_1$ and make conclusion on $u_2$. Here, we need a corresponding function $f_1$. In order to address both perspectives at once, we will aim to formulate uncertainty relations in a symmetric way by a functional $g : \mathbb{R}^2 \mapsto \mathbb{R}$, which obeys

$$\begin{aligned} &(i) && g(u_1, u_2) \geq 0 && \forall (u_1, u_2) \in \mathcal{U} \\ &(ii) && g(u_1, u_2) \leq g(u_1', u_2') && \text{if } u_1 \leq u_1' \text{ and } u_2 \leq u_2'. \end{aligned} \tag{1.6}$$

Given such a functional, the functions $f_1$ and $f_2$ can be extracted from the implicit function $g(u_1, u_2) = 0$.

- **State-independence:** When obtained for an explicit setting, the functional $g$ gives us state-independent statements like

  $$\forall\rho : g\left(\nu(A_1|\rho), \nu(A_2|\rho)\right) \geq 0$$

  or device-independent statements like

  $$\forall R : g\left(\varepsilon(A_1|A_1'), \varepsilon(A_2|A_2')\right) \geq 0.$$

  Here the terms *state-independent* and *device-independent* refer to the fact that the respective functional $g$, i.e. our uncertainty relation, does not depend on $\rho$ or $R$. In practice, this ensures that we can put restriction on the uncertainties of an unknown quantum state or on a non-characterized joint measurement device, which is the crucial point for many applications of uncertainty relations. Important examples for this are the testing of non-local correlations [UBGP15,CSUG17], or applications in cryptography [FFB+12]. We have to note, that not all relations that can be found in the literature share this property. The most prominent example for this is the relation of Robertson and Kennard [Ken27,Rob29], which can be found in any textbook:

  $$\Delta^2 A \Delta^2 B \geq \left|\frac{1}{2}\langle[A,B]\rangle_\rho\right|^2$$

  Here, an evaluation of the r.h.s. demands us to know the state $\rho$. Hence, no function $g$ can be extracted from this directly. The only way to obtain such a $g$ is to minimize the r.h.s over all states, which, for finite dimensions, unfortunately results in the trivial functional $g \equiv 0$.

- **More than two observables:** An observation like (1.4) has more than one counterpart for the general case of more than two measurements. Here, we could have imposed restrictions on $u_1$ in order to make conclusions on all other $u_{i\neq 1}$ or we could have an assumption on $u_{i\neq 1}$ in order to make conclusions on the other $u_{i\neq 1}$ or we could have restrictions on a subset $u_1, \ldots, u_j$ in order to make conclusions on $u_{j+1}, \ldots, u_n$. Here, each case demands a different function $f$, in order to get an analogous statement to (1.5). However, this particular $f$ can be extracted in all cases from a common functional $g : \mathbb{R}^n \mapsto \mathbb{R}$, which gives us implicit functions via

  $$g(u_1, \ldots, u_n) = 0.$$

Finding such functionals, for given measures $\nu$ and $\varepsilon$ and observables $A_1, \ldots, A_n$, is the second major task in this thesis. This quest clearly demands to characterize those $u \in \mathcal{U}$ that extremize the above statements, i.e. the most 'certain'/'small' $u \in \mathcal{U}$. In order to judge this extremality, we will rely on the following notion of multicriterial optimality:

- **Pareto optimality:** A well established concept for judging multicriterial optimality is the, so called, *Pareto optimality* or *Pareto efficiency* originated in the field of operations research. Here, we use the natural half ordering on $\mathbb{R}^n$ to say that $u$ dominates $u'$ and write $u \sqsubseteq u'$, if we have
$$u_1 \leq u'_1, \ldots, u_n \leq u'_n,$$
for a pair $u, u' \in \mathcal{U}$. In terms of uncertainty, this domination means that no component of $u'$ is less uncertain than the respective component of $u$. Hence, we are in a position to say that $u'$ is more uncertain than $u$ and drop $u'$ as a candidate for the most 'certain' elements of $\mathcal{U}$.

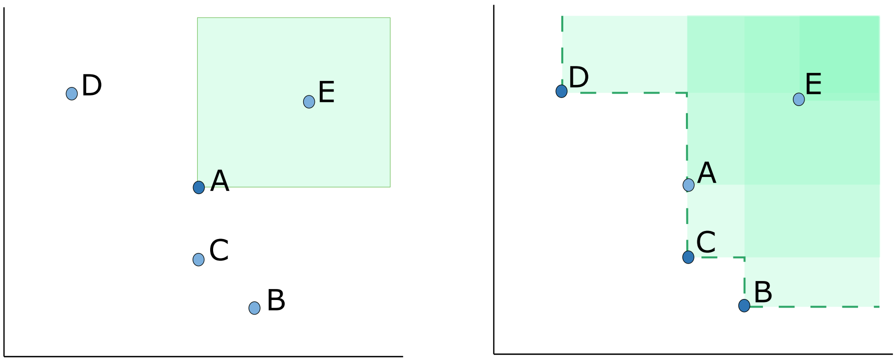


Figure 1.5: (left) Pareto dominated region $\mathsf{PDR}(\mathsf{A})$. (right) Pareto dominated region $\mathsf{PDR}(\mathsf{A},\mathsf{B},\mathsf{C},\mathsf{D},\mathsf{E})$. Here we have $\mathcal{U}_* = \{\mathsf{B},\mathsf{C},\mathsf{D}\}$.

For the quantification of an uncertainty relation we are interested in the 'most certain' points from $\mathcal{U}$, which we will define as all points from $\mathcal{U}$ that cannot be dominated by another point from $\mathcal{U}$. We will denote this set by $\mathcal{U}_*$ and use the term *minimal uncertainty state* or *optimal approximative device* for the corresponding states or joint measurements.

Note that, not all two points have a strict ordering with respect to '$\sqsubseteq$', see for example the pair of points $(\mathsf{A},\mathsf{B})$ or $(\mathsf{A},\mathsf{D})$ in Fig. 1.5. Following this observation, it makes sense to single out the set of all $x \in \mathbb{R}^n$ that are dominated by a particular $u$. We call this set the Pareto dominated region of $u$ (see Fig. 1.5(left)) , i.e.
$$\mathsf{PDR}(u) := \{x \in \mathbb{R}^n | u \sqsubseteq x\}.$$

For a whole set $\mathcal{U}$, the Pareto dominated region is constructed by taking the union over the Pareto dominated regions of all elements, i.e. by (see Fig. 1.5(right))

$$\mathsf{PDR}(\mathcal{U}) := \bigcup_{u\in\mathcal{U}} \mathsf{PDR}(u) = \{x \in \mathbb{R}^n | \exists u \in \mathcal{U} : u \sqsubset x\}.$$

This set is bigger than $\mathcal{U}$, but still contains the same information about an uncertainty relation. The (finite) boundary $\partial\mathsf{PDR}(\mathcal{U})$ of this set is called the *Pareto boundary* or *Pareto frontier*, depicted as dashed line in Fig. 1.5. All points in $\mathcal{U}$ are dominated by at least some point from this boundary. However, it could still be possible that a point from $\partial\mathsf{PDR}(\mathcal{U})$ is dominated by another point from $\partial\mathsf{PDR}(\mathcal{U})$, see for example the pair $(\mathsf{A}, \mathsf{C})$ with $C \sqsubseteq A$. If we drop those points from the boundary, the remaining set contains exactly those points that cannot be dominated by another point, i.e. we get the desired set $\mathcal{U}_*$, which is, in this context, also called the *strong Pareto boundary*. Alternatively, this set can be characterized as the smallest subset of $\mathcal{U}$ that spans the region $\mathsf{PDR}(\mathcal{U})$, i.e. as

$$\mathcal{U}_* = \inf\{\omega \subset \mathcal{U} | \mathsf{PDR}(\omega) = \mathsf{PDR}(\mathcal{U})\}.$$

A functional description of $\mathsf{PDR}(\mathcal{U})$ directly gives us the desired functional $g(u_1, \ldots, u_n)$. More precisely, any $g$ that attains $g(u_1, \ldots, u_n) = 0$ for all points on the Pareto boundary, and only there, can be easily modified to also obey the assertions (1.6). Hence, we can identify the Pareto boundary as graphical representation of an optimal uncertainty relation.

- **Convex hulls and linear bounds:** Above, we saw that a characterization of $U_*$ allows, by recovering $\mathsf{PDR}(\mathcal{U})$ and $\partial\mathsf{PDR}(\mathcal{U})$, to obtain all possible uncertainty relations. However, depending on how we choose our uncertainty measure, this characterization may turn out to be practically untreatable, see for example the entropic measures in chapter Ch.7. Hence, we should also consider simpler methods for characterizing $\mathsf{PDR}(\mathcal{U})$:

  In large parts of this thesis we will concentrate on approximating $\mathsf{PDR}(\mathcal{U})$ by its convex hull, denoted by $\mathcal{U}^+$ in the following. An example for this approximation is depicted in Fig. 1.4, here we can see that some parts of the Pareto-boundary get lost by this procedure. However, in [SRW16] Sec.3 we provide a fundamental theorem stating that this approximation gets exact for all measurement uncertainty relations based on cost-functions.

  In practice, the convex hull $\mathcal{U}^+$ can be constructed as the intersection of all linear half spaces containing it, see for example the appendix of [ACF$^+$16]

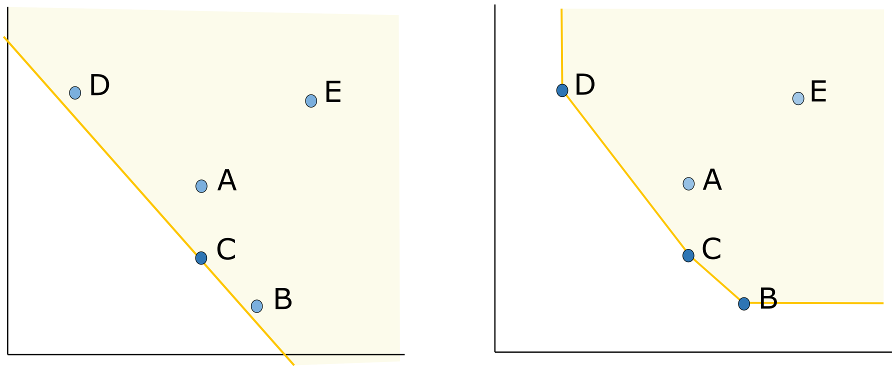


Figure 1.6: (left) Halfspace corresponding to a linear functional that gets minimized on B. (right) Positive convex hull of the points $\{\mathsf{A}, \mathsf{B}, \mathsf{C}, \mathsf{D}, \mathsf{E}\}$

for an elaboration on this. Each of those half spaces can be constructed by minimizing a linear functional

$$k(u_1, \ldots, u_n|a) = \sum_{i=1\ldots n} a_i u_i = u \cdot a$$

given by coefficients $a = (a_1, \ldots, a_n)$. In our particular case, we know that the Pareto dominated region is, by construction, unbounded for each $u_j \to \infty$, i.e. to the upper right (see Fig. 1.4). As a consequence of this, it suffices to consider only linear half space corresponding to functionals with $a \in \mathbb{R}^+$. For those functionals we have

$$c^*(a) := \min_{u \in \mathsf{PDR}(\mathcal{U})} k(u_1, \ldots, u_n|a) = \min_{u \in \mathcal{U}} k(u_1, \ldots, u_n|a) = \min_{u \in \mathcal{U}} u \cdot a. \quad (1.7)$$

Hence, we can compute $\mathcal{U}^+$ from $\mathcal{U}$ directly. Explicitly, we get a particular half space $\mathsf{H}_a(\mathcal{U})$ from (1.7) by

$$\mathsf{H}_a(\mathcal{U}) := \left\{ x \in \mathbb{R}^n \middle| x \cdot a \geq c^*(a) \right\}$$

and the convex hull by

$$\mathcal{U}^+ := \bigcap_{a \in \mathbb{R}^+} \mathsf{H}_a(\mathcal{U})$$

We note that the above also gives

$$\mathcal{U} \subseteq \mathsf{PDR}(\mathcal{U}) \subseteq \mathcal{U}^+$$

on the level of sets.

Furthermore, any $c^*(a)$ directly gives us an uncertainty relation by

$$\forall u \in \mathcal{U} : g(u_1, \ldots, u_n) := k(u_1, \ldots, u_n|a) - c^*(a) \geq 0$$

Hence, we will identify any lower bound on $c^*(a)$ as *linear uncertainty relation* and $c^*(a)$ itself as *the optimal linear uncertainty relation* with respect to weights $a$.

### Semidefinite Programming

Semidefinite programming is a subfield of convex optimization, i.e. a certain class of optimization problems, which will be introduced immediately, with a well understood theory and well performing algorithms available. A good overview can be found in [BV04].

In this thesis we will encounter *semidefinite programms* (SDPs) in the chapters Ch. 2,3, and Ch. 5. A central result of these chapters is to show that the computation of optimal measurement uncertainty relations and optimal preparation uncertainty relations, for measurements with finite outcome sets, can be formulated as semidefinite programs.

For the computation of explicit uncertainty relations we have to solve semidefinite programs by the use of a computer. Thereby, computer programs (we used SDPA and CVX) usually require a SDP to be formulated according to a common standard form:

- **Primal form:** Semidefinite programming can be understood as the extension of linear programming to the cone of positive matrices. Let $C, F_1, \ldots, F_n$ be a collection of square-matrices and let $c = (c_1, \ldots, c_n)$ be a vector with $n$ entries. These are the inputs to our optimization problem. In its primal form an SDP is given as:

$$\begin{aligned} &\text{minimize:} && s_p(X) := \mathrm{tr}(CX) \\ &\text{subjected to:} && \mathrm{tr}(F_i X) = c_i \quad \forall i \in 1, \ldots, n \\ &&& X \succeq 0 \end{aligned}$$

  Here, the optimization runs over the positive matrix $X$.

- **Dual form:** The dual form of an SDP is:

$$\begin{aligned} &\text{maximize:} && s_d(\lambda) := \sum_i c_i \lambda_i \\ &\text{subjected to:} && \sum_i \lambda_i F_i \preceq C \end{aligned}$$

Here, the optimization runs over the vector $\lambda$. Note that the ordering between primal and dual form is not fixed uniquely throughout the literature.

- **Duality:** Both, the primal and the dual form, can be transformed into each other by computing the respective Lagrange dual [BV04]. Computing the Lagrange dual also yields a basic result commonly known as *weak duality*:

  Let $s_p^*$ and $s_d^*$ denote the optimal solutions of the primal and dual formulations of an SDP. For any feasible $X$ and any feasible $\lambda$ we have

  $$\begin{aligned} s_p^* &\geq s_d(\lambda) \\ s_d^* &\leq s_p(X). \end{aligned}$$

  Typical algorithms, for example interior-point-methods [Kar84], for solving SDPs are succesively providing approximations to $s_p^*$ and $s_d^*$ by an iteration on feasible $X$'s and $\lambda$'s. Here, the duality above allows to assign a precision estimate, $\epsilon = s_p(X) - s_d(\lambda)$. For the purpose of this thesis, such a precision estimate can be translated to an estimate on the precision of an uncertainty relation.

  In the most cases, the so called Slater condition, see [BV04], is fulfilled, which gives rise to a *strong duality*

  $$s_p^* = s_d^*,$$

  i.e. the promise that any $\epsilon$ gap can be closed within a finite runtime of an algorithm. A fundamental result on the efficiency of SDPs [VB96] states that this can be done within a runtime that is polynomial in $\epsilon$, the dimension of $X$, and the number of constraints $n$.

- **Inequality constraints:** Most SDPs appearing in this thesis do not obey the standard form above. More precisely, for computing measurement uncertainty relations, (see example [SRW16] Sec. 4), we usually have inequality constraints of the form

  $$\begin{aligned} \mathrm{tr}(F_i X) &\leq c_i \\ X &\succeq 0. \end{aligned} \tag{1.8}$$

  These can be transformed to the standard form by enlarging the dimension of $X$ and introducing a new variable '$s_i \leq 1$', a so called *slack variable*: In order to implement (1.8) for a particular $i$, we replace $F_i$ by $F_i \oplus 1$, and $X$

by $X \oplus s_i$, which gives the new constraint

$$\begin{aligned} &\operatorname{tr}\left((F_i \oplus 1)(X \oplus s_i)\right) = \operatorname{tr}(F_i X) + s_i = c_i \\ &\quad X \succeq 0 \text{ and } s_i \geq 0. \end{aligned} \tag{1.9}$$

Here the constraint (1.9) is equivalent to $c_i - \operatorname{tr}(F_i X) = s$, which can only be realised by

$$\operatorname{tr}(F_i X) \leq c_i$$

since we have $s_i \geq 0$.

### Joint numerical ranges

In many parts of this thesis (see chapter Ch. 3) we will formulate the minimization of preparation uncertainty as minimization of a linear functional, i.e. of an expectation value $\operatorname{tr}(T_i\rho)$, over the set of all quantum states. A corresponding uncertainty relation then demands us to jointly minimize several functionals $\{\operatorname{tr}(T_i\rho)\}_{i=1,\ldots,n}$. All information on this is encoded in the set of all joint expectations those functionals can attain. This is the so called *joint numerical range*

$$\mathcal{C}(T_1,\ldots,T_n) = \{(\operatorname{tr}(\rho T_1),\ldots,\operatorname{tr}(\rho T_n)) \in \mathbb{R}^n | \rho \text{ is a quantum state }\},$$

which is sometimes also called *convex support* [Wei11] or *gemeinsamer Wertevorat* [Toe18]. The computation of the boundary, i.e. extremal linear functional, of joint numerical ranges can be achieved by computing maximal eigenvalues. Hence, this set can be handled numerically in an efficient way.

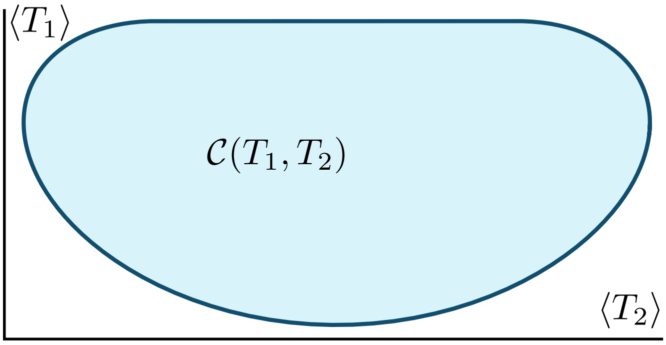


Figure 1.7: Joint numerical range for two operators. The boundary piecewise consists of algebraic curves. Flat. parts of the boundary correspond to degeneracies of the underlying eigenvalue problem, see [SW18a, SW18b].

### Entropies

In the following subsection the basic entropic quantities, needed for an understanding of this thesis, are introduced. A comprehensive overview on entropies in information theory can be found in [Wil13].

- **Shannon entropy:** In his seminal work [Sha48], Claude Shannon introduced the Shannon entropy of a random variable $X$ with distribution $p$ as

  $$H(X) = -\sum_i p_i \log(p_i). \tag{1.10}$$

  Note that, by a slight abuse of notation, we will also write $H(p)$ such as $H(p_1, \ldots, p_n)$ to denote the entropy of a particular distribution $p = (p_1, \ldots, p_n)$. In this thesis we will frequently consider the Shannon entropy evaluated on the outcome statistics of quantum measurements for quantifying the corresponding preparation uncertainty. Interestingly, Shannon already used the term 'uncertainty' as intuitive paraphrase for the word 'entropy'. As postulation, the Shannon entropy is uniquely defined by the following properties (see also [Ré61], and the references therein):

  The Shannon entropy (1.10) of a random variable $X$ is a function that only depends on the underlying probability distribution, that uniquely obeys ( [Fad57] quotet in [Ré61]):

  $h(1)$ $\quad H(p) = H(\pi(p))$ for all permutations $\pi$

  $h(2)$ $\quad H(t, 1-t)$ is continous for $0 \leq t \leq 1$

  $h(3)$ $\quad H(1/2, 1/2) = 1$ the entropy of a fair coin is$1[bit]$

  $h(4)$ $\quad H(tp_1, (1-t)p_1, p_2, \ldots, p_n) = H(p_1, \ldots, p_2) + p_1 H(t, 1-t)$

  Note that, the property $(iii)$ fixes the unit of the entropy, i.e. all logarithms are to be taken to the base 2. At this point, it is also common to define the entropy to the base $e$. We will partially use this alternative convention in this thesis to simplify the mathematical presentation of some proofs.

- **Self-information:** The definition (1.10) can also be understood as an expectation value

  $$H(X) = \langle I(X, p) \rangle$$

  of the quantity

  $$I(i, p) = -\log(p_i). \tag{1.11}$$

In the context of information theory this quantity (1.11) is commonly called *self-information*. A common paraphrase, used for this quantity, is *surprisedness*. We have the following properties:

| | | |
|---|---|---|
| $(i)$ | $-\log(p_i = 1) = 0$ | 'there is no surprise about certain events' |
| $(ii)$ | $-\log(p_i = 0) = \infty$ | 'the surprise about an impossible event is infinite' |
| $(iii)$ | $-\log(p_i q_j) = -\log(p_i) - \log(q_j)$ | 'the surprise of independent events is additive' |

In this sense (1.11) measures the surprisedness of an event '$i$', with probability $p_i$, taking place. In chapter Ch. 5 we will use the self-information as a cost function in order to define an entropic measurement error.

- **Cross entropy:** In the field of estimation theory the cross entropy is used for quantifying the quality an estimative probability distribution $q$ has, with respect to an ideal but unknown distribution $p$. For a random variable $X$, distributed by $p$, the cross entropy is defined as

  $$H(X;q) = \langle I(X,q) \rangle = -\sum_i p_i \log(q_i),$$

  i.e. as the expected self-information, with respect to $q$, of events sampled by $X$. We will see in chapter Ch. 5 that entropic measurement errors can be interpreted as a minimization of cross entropies with respect to a distribution of estimators $q$.

- **Relative entropy:** The *relative entropy*, commonly also called *Kullback-Leibner divergence*, is the expected entropic difference between a distribution $p$ and its estimate $q$, i.e.

  $$\begin{aligned} D(X||q) &= \langle I(X,q) - I(X,p) \rangle = H(X;q) - H(X) \\ &= \sum_i p_i \log\left(\frac{p_i}{q_i}\right) \end{aligned}$$

  Note that the above definition requires $\text{supp}(p) \subseteq \text{supp}(q)$ in order to yield a finite quantity. For all probability distributions $p$ and $q$ with $\text{supp} \subseteq \text{supp}(q)$, the relative entropy obeys the basic properties

  | | | |
  |---|---|---|
  | $(i)$ | $D(p\|\|q) > 0$ | for $p \neq q$ |
  | $(ii)$ | $D(p\|\|q) = 0$ | for $p = p$, |

i.e. the axioms of a premetric. Hence, it can be used as distance function on the level of probability distributions. Measurement uncertainty relations in terms of relative entropies were investigated in [BGT18, BGT17]. In chapter Ch. 5, we investigate measurement uncertainty relations based on the discrete metric. Those relations give lower bounds, by Pinsker's inequality, on uncertainty relations in terms of relative entropies.

- **Joint entropy:** For two, potentially correlated, random variables $X$ and $Y$, it is common to write $XY$ to denote a random variable obeying the joint distribution

  $$p_{ij} := \mathbb{P}(X = i, Y = y).$$

  Here, the *joint entropy* of $X$ and $Y$ denotes the Shannon entropy of this joint distribution

  $$H(XY) = -\sum_{ij} p_{ij} \log(p_{ij}).$$

  The important property of the Shannon entropy is that the joint entropy of independent random variables $X$ and $Y$ is additive.

  $$H(XY) = H(X) + H(Y) \text{ for } p_{ij} = p_i p_j. \tag{1.12}$$

- **Conditional entropy:** The *conditional entropy* is used to quantify the uncertainty on an unknown random variable $X$, given information on the outcomes of a second random variable $Y$. It is defined as the difference between the joint entropy and the entropy of $Y$, i.e. as

  $$H(X|Y) = H(XY) - H(Y) = \sum_{ij} p_{ij} \log(p_{ij}).$$

  Here we have the properties:

  $$\begin{array}{lll} (i) & H(X|Y) = H(X) & \text{if } X \text{ and } Y \text{ are independent} \\ (ii) & H(X|Y) = 0 & \text{if } X \text{ is fully determined by } Y. \end{array}$$

  In chapter Ch. 5 we will use the conditional entropy to quantify a measurement uncertainty between devices with differing outcome sets. There, we will use ensembles of test states, distributed by $X$, as input to a measurement device with an outcome statistic described by $Y$. Here, the conditional entropy $H(X|Y)$ quantifies the information on the input $X$ that is obtained by observing only the output $Y$.

- **Rényi entropy:** In the seminal paper [Ré61] Alfréd Rényi suggested to drop the property $h(4)$, in the postulative definition of the Shannon entropy. In conclusion, he replaced $(iv)$ by the additivity property (1.12), and defined, for $\alpha \neq 1$, the family of *Rényi- entropies*

  $$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_i p_i^\alpha\right). \tag{1.13}$$

  which obeys this weaker property for $\alpha > 0$. In the limit of $\alpha \to 1$, the logarithm in (1.13) vanishes, whereas the pre-factor $(1-\alpha)^{-1}$ diverges. Remarkably, we recover the Shannon entropy (1.10) in this case, i.e. we have

  $$\lim_{\alpha \to 1} H_\alpha(X) = H(X). \tag{1.14}$$

  Hence, the Rényi entropy gives a generalization of the Shannon entropy. In this thesis, the Rényi entropy is used excessively in the proofs provided in the chapters Ch. 5 and Ch. 7. There, we make use of the well known technique of first proving statements on Rényi entropies and then concluding the respective statements on Shannon entropies by the limit (1.14).

## 1.4 Structure and summary

Central parts of this thesis, i.e. Ch. 2, 4, and Ch. 6-8, consist of manuscripts that have already been published in peer-reviewed journals. Each manuscript is included as a single section, embedded in a separate chapter. The chapters Ch. 4 and 6 additionally contain unpublished material and results related to the respective manuscripts. Up to modifications of the page layout, all presented manuscripts are identical to preprint versions that can be found on arXiv.org . The chapters Ch. 3 and Ch. 5 do not contain any prepublished material. These chapters contain new results, provided with the aim to build bridges between the content of the other chapters.

All the work done for this thesis was motivated by the the same set of central questions, which can be summarized as follows:

(1.) How can we quantify measurement/preparation uncertainty?

(2.) How can we compute measurement/preparation uncertainty relations?

(3.) Can we find connections between measurement and preparation uncertainty relations?

The scope of this thesis is to provide some satisfying answers to these questions, whenever possible, in a general context and with a focus on relevant examples, otherwise.

The main part of this thesis starts at chapter **chapter Ch. 2**. Here, we will begin by introducing, so called, *cost functions*. We will use such cost functions throughout this work as central tool for modelling the proximity of measurement outcomes. The main part of this chapter is **[SRW16]** 'Measurement uncertainty relations for finite quantum Observables'. This work has the aim to provide a general construction for error measures which answers the questions (1.) and (2.) for the case of measurement uncertainty. We will provide the following three types of error measures based on cost functions:

(i) Measurement error $\varepsilon_M$: This error judges the proximity of $A'$ to $A$ by a test performed on arbitrary quantum states. Here, $A'$ is a good approximation to $A$ when both devices show almost the same outcome statistics for every input state.

(ii) Calibration error $\varepsilon_C$: This error is only defined for a comparison with respect to a sharp measurement $A$. In this case, we have deterministic outcomes $x$ when we measure $A$ on its eigenstates $\{\phi_x\}$. We will construct error measures by using those eigenstates as input for testing an approximative device $A'$. Thereby, a particular eigenstate $\phi_x$ is used as a reference for a 'correct' measurement outcome $x$. Hence, a good approximation $A'$ should give outcomes close to the correct $x$ for all inputs $\phi_x$.

(iii) Entangled reference frame error $\varepsilon_E$: This error is only defined for sharp measurement $A$, too. We test $A$ and $A'$ on the local sides of a maximally entangled state $\phi^+$. Within the framework of this error, $A'$ is a good approximation to $A$ if both measurement outcomes are close to each other in any particular shot. Thereby, the advantage to the calibration error is that the state $\phi^+$ serves as an unbiased test input for arbitrary devices.

Our central contribution to the question (2.) is to show that the optimization problem corresponding to a measurement uncertainty relation, formulated in terms of the above quantities, can be formulated as SDP. Hence, we can compute those relations in the general case.

In **chapter Ch. 3** deviation measures $\nu(A|\rho)$ in terms of cost functions are introduced. We will show that the corresponding preparation uncertainty relations can be computed by solving certain eigenvalue problems, which gives a positive answer to question (2.), as well. In the second part of this chapter, we will give a partial answer to question (3.) by proving a basic theorem which states that

$$\text{preparation uncertainty} \preceq \text{ measurement uncertainty} \tag{1.15}$$

for linear uncertainty relations with respect to sharp observables.

The **chapter Ch.4** is centred on preparation uncertainty relations in terms of variances, here we will focus on question (2.). An overwhelming part of the literature on quantum physics, especially textbooks, only considers uncertainty relations formulated in terms of variances. Hence, their investigation is from a fundamental interest by itself. However, in the existing literature, explicit uncertainty relations in terms of variances were only known for some special cases. Finding a general method for computing them was an open and outstanding problem.

Within the framework established in this thesis, variances can be understood as deviation $\nu(A|\rho)$ with respect to a cost function that has infinite support. On the positive side, we have the validity of the result (1.15) for this case. However, we have the drawback that the techniques, introduced in Ch. 3 for computing uncertainty relations, fail in this case, because they result in optimization problems with infinitely many constraints. Therefore, in **[SDW17]**, the main part of this chapter, an algorithmic method is provided, which allows to compute linear preparation uncertainty relations in terms of variances for arbitrary measurements on a finite Hilbert space, and, hence, settles this open problem with a wide generality. As additional material to this work, we provide a non-algorithmic method for computing lower bounds on those relations.

In **chapter Ch.5** we will change our view on uncertainty relations to the perspective of information theory, by regarding the outcomes $i$, of a measurement, as letters from a general alphabet $I$. Here, we will again start by considering the question (1.) with respect to measurement uncertainty.

On one hand, we will investigate error measures based on the discrete metric, such as the according preparation uncertainty relations, for which we again have the basic connection (1.15), and easy computable uncertainty relations.

On the other, we will provide two types of entropic error measures: The first is based on using the self-information as cost function. This error measure allows us to compare an ideal device $A$, with outcomes on an alphabet $I$, to a device $\mathtt{A}'$ which gives us a full hypothesis, coming as probability distribution on $I$, as output. The second error measure is based on conditional entropies, rather than on a cost function. This measure allows to compare $A$ to a device $A'$ which has its output on an arbitrary alphabet $J \underline{\neq} I$. We show that, for both entropic error measures, measurement uncertainty relations are lower bounded by linear preparation uncertainty relations in terms of the Shannon entropy.

In **chapter Ch.6** we put our focus on those linear preparation uncertainty relations in terms of the Shannon entropy. Beside their connection to entropic measurement uncertainty relations, those relations are interesting by them selves. They have many applications in the fields of quantum information theory. For example: as central estimate in security proofs [FFB$^+$12], as building block for steering inequalities [CSUG17, RMM17a], and as bound in quantum metrology. Here, a satisfactory answer to question (2.), for the general case is an outstanding problem.

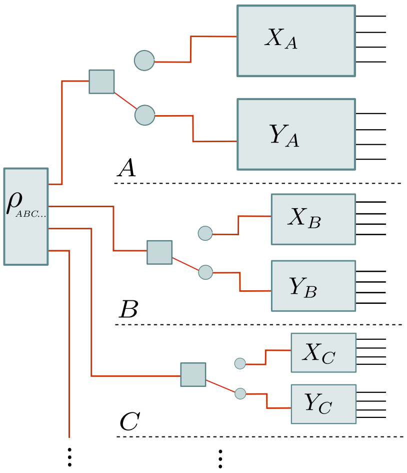


Figure 1.8: Basic setting of multipartite uncertainty: A state $\rho_{ABC\dots}$ is distributed between parties $A, B, \dots$. Depending on a coin throw, with probabilities $(\lambda, 1-\lambda)$, all parties perform measurements $X_{AB\dots} = X_A \otimes X_B \otimes \dots$ or $Y_{AB\dots} = Y_A \otimes Y_B \otimes \dots$, respectively. The probability distribution of all outcomes is given by $\lambda \mathbf{p}_{X_{AB\dots}} \oplus (1-\lambda)\mathbf{p}_{Y_{AB\dots}}$. The additivitiy shows that: for fixed $\lambda$, the entropy of this distribution is minimized by fully separable states $\rho_{AB} = \rho_A \otimes \rho_B \otimes \dots$.

However, in **[S18]** 'Additivity of entropic uncertainty relations', which is the first section and the main part of this chapter, a central result on the structure of those bounds is proven: It is shown, for pairs of local measurements on a $n$-partition (see Fig. 1.8), that linear entropic preparation uncertainty relations are additive. Interestingly, this directly implies that the minimal entropic uncertainty can always be realized by fully separable states. Furthermore, global uncertainty bounds now can be deduced from local ones by a single letter formula. As side result, the generalization of the Maassen and Uffink bound to arbitrary linear uncertainty relations is provided.

In the second section of this chapter we will provide some novel algorithmic methods, based on alternating minimization, that can be used to compute linear entropic preparation uncertainty relations quite reliably for measurements on moderate small dimensional Hilbert spaces. However, in contrast to the method provided for variances, this method does not have a promise to converge in any case, neither does it give lower bounds in finite runtime, nor a precision estimate.

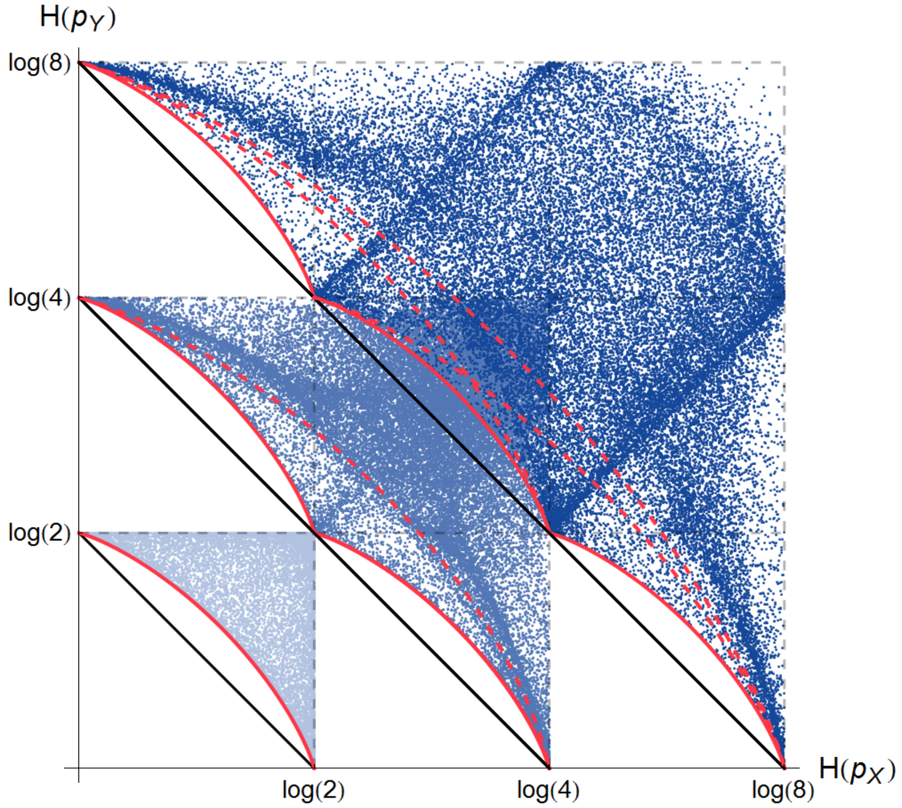


Figure 1.9: Entropic preparation uncertainty regions for measurements linked by the Fourier transformation $F_2, F_2^{\otimes 2}, F_2^{\otimes 3}$, i.e. in dimension $d = 2, 4, 8$. We can see that non-linear improvements to the best linear bounds (black line, given by the Maassen and Uffink bound) are possible. Those diagrams show an additivity structure: The uncertainty set $\mathcal{U}_{F_2^{\otimes 2}}$ is the Minkowski sum of $\mathcal{U}_{F_2}$ with itself, and $\mathcal{U}_{F_2^{\otimes 3}}$ is the Minkowski sum of $\mathcal{U}_{F_2}$ and $\mathcal{U}_{F_2^{\otimes 3}}$.

There are indications, which suggest that the underlying optimization might be an NP-hard approximation problem. We will have a brief comment on this, as well.

In **chapter Ch.7** we will consider entropic preparation uncertainty for the explicit example of two measurements given by a pair of mutually unbiased bases. Note that this example includes measurement bases linked by a finite Fourier transformation as important sub case. For mutually unbiased bases, linear preparation uncertainty relations can be computed easily, because the Maassen and Uffink bound is tight in this case. However, reliable non-linear bounds are not known for this neither.

The main part of this chapter is **[ASM$^+$15]** 'Optimality of entropic uncertainty relations', here we investigate such non-linear bounds numerically. Interestingly, we observe the same additivity behaviour (see Fig. 1.9) which was proven for linear bounds in [S18], for non-linear bounds, too. As a general result we characterize the class of measurements for which the Maassen and Uffink bound gives the best linear bound, and we provide a basic theorem, which characterizes all states that achieve equality for the Maassen and Uffink inequality.

In **chapter Ch.8** we will consider applications of uncertainty relations to possible effects of quantum gravity. Many theories of quantum gravity suggest to modify the Heisenberg algebra when approaching the scale of Planck units. Depending on the individual starting point of those theories, the existence of a minimal length, which should manifest as a ultimate lower bound on the position uncertainty,

$$\Delta^2(X) \geq l_0 \quad \forall \rho, \tag{1.16}$$

is postulated, proven, or assumed.

In **[ACF$^+$15]**, the main part of this chapter, we consider a modified algebra of the form

$$[x,p] = i\hbar f(p), \tag{1.17}$$

where $f(p)$ is a symmetric convex function of the momentum operator, that allows for an expansion

$$f(p) = 1 + \alpha(p^2 + \dots),$$

with a parameter $\alpha << \hbar$, that is only relevant on very short distances or high momenta. We investigate those modifications within the framework of 'standard quantum mechanics', i.e. we start with a position operator that is represented as a multiplication operators on $L^2(\mathbb{R})$. In analogy to the Stone-von Neumann theorem, we provide a basic theorem, which shows that any representation of (1.17), obeying certain transformation rules, results in a theory that has an effective UV-cut-off and a minimal length (1.16). From a second perspective, those theories can be seen as quantum mechanics living on a subspace of band-limited functions. Here, a position coordinate is still continuous, whereas a position measurement is now described by a non-projective POVM, which, at the end, leads to a lower limit on the respective uncertainty.

We provide basic tools that allow to compute modified uncertainty relations. We also consider uncertainty relations based on the Shannon entropy and the min-entropy. Interestingly it turns out that the min-entropy is always bigger than one bit.

# CHAPTER 2

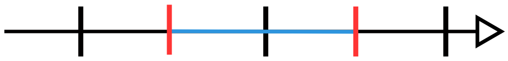
## Measurement uncertainty relations based on cost functions

In this chapter we will introduce quantitative descriptions of measurement uncertainty relations. Albeit, we have the aim to do this in a most general manner, we will restrict the investigation to observables on finite Hilbert spaces with finite outcome sets. The central part of this chapter is [SRW16], given in the following section. The central contribution of this work is to provide a couple of constructions of computable error measures between POVMs. Those will serve to quantify uncertainty and compute uncertainty relations later on. Our line of thinking is heavily inspired by classical optimal transportation theory (see for example [Vil09]), and our construction can be seen as an operator valued version of the Monge-Wasserstein-Kantorovic distance. At this point we should note that, this idea was originaly introduced in [BLW14b, BLW13, Wer04]. However, it was only applied for particular examples. In contrast, the aim of [SRW16] was to provide according results in a more general context.

The central benefit of the constructions provided in [SRW16] is to obtain an error measure that is based solely on a, so called, *cost function.* This is a function that assigns a *(transport-) cost* to pairs of measurement outcomes and, thus, gives a quantitative way to compare two measurements in a single shot. Thereby, a cost function must not necessarily obey the axioms of a metric. In general, we can regard a cost function as the input that models the nature of the underlying physical setting in a particular instance of a class of optimization problems. At the end of the construction, the resulting error measure inherits its operational meaning, physical unit, and its transformation behaviour, from the underlying cost function.

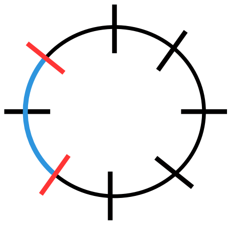
Since cost functions are kept as unspecified objects in [SRW16], we close this introduction by listing some examples of typical outcome sets and their according cost functions:

### Points on a line



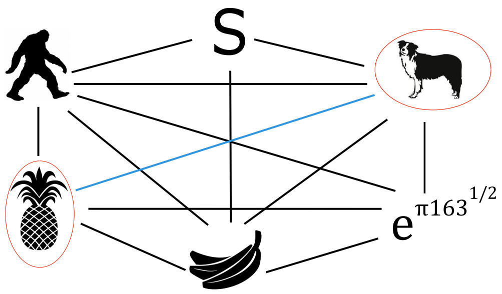
A standard situation in (quantum-) physics is to consider measurements with a discrete set of measurement outcomes that are (i) part of real numbers and (ii) the ordering according to the real numbers is a relevant part of the problem. For example, any observable that is modelled by a self adjoint operator with spectrum corresponding measurement outcomes, as done in any textbook, obeys (i). Furthermore, (ii) is a very natural assumption for physical properties that have a corresponding quantity in classical physics, like energy, position, momentum, and angular momentum. Here it is natural to take the euclidean metric $|x-y|$ as cost function. However, we will also consider cost functions of the type $|x-y|^m$. For $m \neq 1$ the triangle inequality is violated, hence this is an example for a cost function that is not a metric. The case $m=2$ is from special interest, because it is closely related to variances. Since variances have their own relevance at many points in classical statistics, they will be regarded separately in chapter Ch. 4.

### Points on a circle



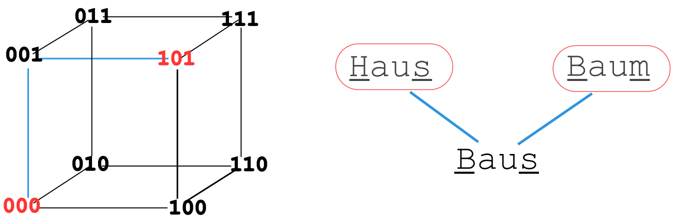
A situation, where the euclidean distance does not appropriately model the underlying setting, are outcome sets with a periodic structure. For example points on a circle. Here an appropriate cost function should respect this periodicy. Therefore, it could be based on the angle $\phi$ between two points. Furthermore, one can think of embedding the circle into $\mathbb{R}^2$ and take the euclidean distance in this space. This will lead to a cost function based on $\arccos(\phi)$. Measurement uncertainty relations for this case were investigated in [BKW16].

### The discrete metric



As most general case, one has to consider a finite outcome set with no further structure on it. Here, the only distinction that can be made is if two elements are equal or not. Hence, all reasonable non trivial cost functions are proportional to the *discrete metric*. This metric equals zero whenever an object is compared to itself and is constant otherwise. Measurement uncertainty relations with this metric were investigated in the Bachelor thesis [Fra15]. In this thesis we will have a closer look at this in chapter Ch. 5.

### Strings



A further example, with regard to applications in information theory, and more precisely coding theory, are strings of characters from some alphabet. Here, we can compare two strings within the Hamming distance. This is, we compare two strings, at first, on each position within the discrete metric and sum over this afterwards. At the end, this gives a number, which is proportional to the number of position where the corresponding symbols are different. Measurement and preparation uncertainty relations for this case were investigated in [Wer16, Jav16].

## 2.1 [SRW16]

*Measurement uncertainty relations for finite quantum Observables*

- **Authors:** René Schwonnek, David Reeb, and Reinhard F. Werner
- **Published in:** Mathematics 4 (2),38.
- **DOI:** 10.3390/math4020038
- **Presented version:** arXiv: 1604.00382
- **Contributions:** All authors contributed equally.
- **Main results:**
  - Definition of the measurement error quantities $\varepsilon_M(A|A')$, $\varepsilon_C(A|A')$, and $\varepsilon_E(A|A')$ for arbitrary cost functions, based on the ansatz of Wasserstein metrics.
  - Uncertainty regions of measurement uncertainties are always convex.
  - The computation of optimal measurement uncertainty relations can be formulated as SDP. The main ingredient for this is the use of the Kantrorovic duality for cost functions on finite domains. Here, the full list of optimal pricing schemes can be computed.

# Measurement Uncertainty for Finite Quantum Observables

René Schwonnek[*], David Reeb[†], and Reinhard F. Werner[‡]

Quantum Information Group, Institute for Theoretical Physics,
Leibniz Universität Hannover

April 1, 2016

**Abstract**

Measurement uncertainty relations are lower bounds on the errors of any approximate joint measurement of two or more quantum observables. The aim of this paper is to provide methods to compute optimal bounds of this type. The basic method is semidefinite programming, which we apply to arbitrary finite collections of projective observables on a finite dimensional Hilbert space. The quantification of errors is based on an arbitrary cost function, which assigns a penalty to getting result $x$ rather than $y$, for any pair $(x, y)$. This induces a notion of optimal transport cost for a pair of probability distributions, and we include an appendix with a short summary of optimal transport theory as needed in our context. There are then different ways to form an overall figure of merit from the comparison of distributions. We consider three, which are related to different physical testing scenarios. The most thorough test compares the transport distances between the marginals of a joint measurement and the reference observables for every input state. Less demanding is a test just on the states for which a "true value" is known in the sense that the reference observable yields a definite outcome. Finally, we can measure a deviation as a single expectation value by comparing the two observables on the two parts of a maximally entangled state. All three error quantities have the property that they vanish if and only if the tested observable is equal to the reference. The theory is illustrated with some characteristic examples.

## 1 Introduction

Measurement uncertainty relations are quantitative expressions of complementarity. As Bohr often emphasized, the predictions of quantum theory are always relative to some definite experimental arrangement, and these settings often exclude each other. In particular, one has to make a choice of measuring devices, and typically quantum observables cannot be measured simultaneously. This often used term is actually misleading, because time has nothing to do with it. For a better formulation recall that quantum experiments are always statistical, so the predictions refer to the frequency with which one will see certain outcomes when the whole experiment is repeated very often. So the issue is not *simultaneous* measurement of two observables, but *joint* measurement in the same shot. That is, a device $R$ is a joint measurement of observable $A$ with outcomes $x \in X$ and observable $B$ with outcomes $y \in Y$, if it produces outcomes of the form $(x, y)$ in such a way that if we ignore outcome $y$, the statistics of the $x$ outcomes is always (i.e., for every input state) the same as obtained with a

[*] rene.schwonnek@itp.uni-hannover.de

[†] david.reeb@itp.uni-hannover.de

[‡] reinhard.werner@itp.uni-hannover.de

measurement of $A$, and symmetrically for ignoring $x$ and comparing with $B$. It is in this sense that non-commuting projection valued observables fail to be jointly measurable.

However, this is not the end of the story. One is often interested in *approximate* joint measurements. One such instance is Heisenberg's famous $\gamma$-ray microscope [11], in which a particle's position is measured by probing it with light of some wavelength $\lambda$, which from the outset sets a scale for the accuracy of this position measurement. Naturally, the particle's momentum is changed by the Compton scattering, so if we make a momentum measurement on the particles after the interaction, we will find a different distribution from what would have been obtained directly. Note that in this experiment we get from every particle a position value and momentum value. Moreover, errors can be quantified by comparing the respective distributions with some ideal reference: The *accuracy* of the microscope position measurement is judged by the degree of agreement between the distribution obtained and the one an ideal position measurement would give. Similarly, the *disturbance* of momentum is judged by comparing a directly measured distribution with the one after the interaction. The same is true for the *uncontrollable disturbance* of momentum. This refers to a scenario, where we do not just measure momentum after the interaction, but try to build a device that recovers the momentum in an optimal way, by making an arbitrary measurement on the particle after the interaction, utilizing everything that is known about the microscope, correcting all known systematic errors, and even using the outcome of the position measurement. The only requirement is that at the end of the experiment, for each individual shot, some value of momentum must come out. Even then it is impossible to always reproduce the pre-microscope distribution of momentum. The tradeoff between accuracy and disturbance is quantified by a measurement uncertainty relation. Since it simply quantifies the impossibility of a joint exact measurement, it simultaneously gives bounds on how an approximate momentum measurement irretrievably disturbs position. The basic setup is shown in Fig. 1.

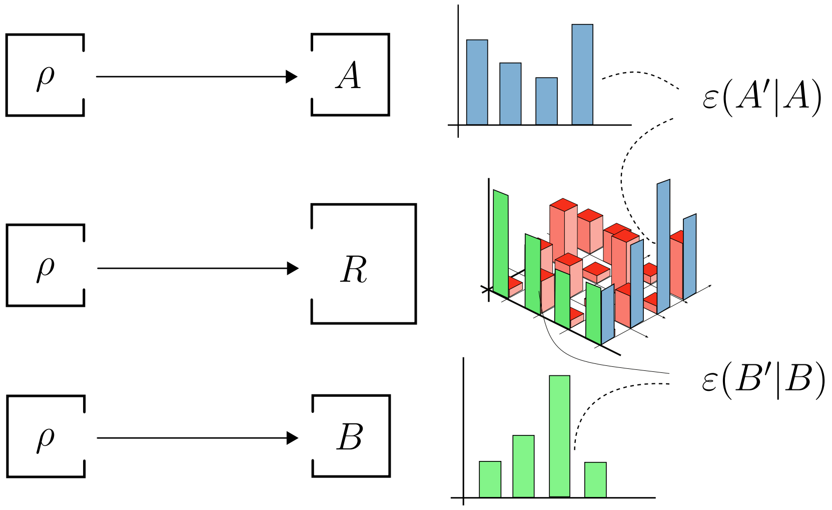


Figure 1: Basic setup of measurement uncertainty relations. The approximate joint measurement $R$ is shown in the middle, with its array of output probabilities. The marginals $A'$ and $B'$ of this array are compared with the output probabilities of the reference observables $A$ and $B$, shown at the top and at the bottom. The uncertainties $\varepsilon(A'|A)$ and $\varepsilon(B'|B)$ are quantitative measures for the difference between these distributions.

Note that in this description of errors we did not ever bring in a comparison with some hypothetical "true value". Indeed it was noted already by Kennard [13] that such comparisons are problematic in quantum mechanics. Even if one is willing to feign hypotheses about the true value of position, as some hidden variable theorists will, an operational criterion for agreement will always have to be based on statistical criteria, i.e., the comparison of distributions. Another fundamental feature of this view of errors is that it provides a figure of merit for the comparison of two devices, typically some ideal reference observable and and an approximate version of it. An “accuracy” $\varepsilon$ in this sense is a promise that no matter which input state is chosen, the distributions will not deviate by more than $\varepsilon$. Such a promise does not involve a particular state. This is in contrast to *preparation uncertainty* relations, which quantify the impossibility to find a state for which the distributions of two given observables (e.g., position and momentum) are both sharp.

Measurement uncertainty relations in the sense described here were first introduced for position and momentum in [23], and were initially largely ignored. A bit earlier, an attempt by Ozawa [15] to quantify error-disturbance tradeoffs with state dependent and somewhat unfortunately chosen [7] quantities had failed, partly for reasons already pointed out in [1]. When experiments confirmed some predictions of the Ozawa approach (including the failure of the error-disturbance tradeoff), a debate ensued [4, 16, 6, 2]. Its unresolved part is whether a meaningful role for Ozawa’s definitions can be found. Technically, the computation of measurement uncertainty for position and momentum in [6] carries over immediately to more general phase spaces [24, 3]. Apart from some special further computed instances [8, 5], this remained the only case in which sharp measurement uncertainty relations could be obtained. This was in stark contrast with preparation uncertainty, for which an algorithm based on solving ground state problems [8] efficiently provides the optimal relations for generic sets of observables. The main aim of the current paper is to provide efficient algorithms also for sharp measurement uncertainty relations.

In order to do that we restrict the setting in some ways, but allow maximal generality in others. We will restrict to finite dimensional systems, and reference observables which are projection valued and non-degenerate. Thus, each of the ideal observables will basically be given by an orthonormal basis in the same $d$-dimensional Hilbert space. The labels of this basis are the outcomes $x \in X$ of the measurement, where $X$ is a set of $d$ elements. We could choose all $X = \{1, \ldots, d\}$, but it will help to keep track of things using a separate set for each observable. Moreover, this includes the choice $X \subset \mathbb{R}$, the set of eigenvalues of some hermitian operator. We allow not just two observables but any finite number $n \geq 2$ of them. This is makes some expressions easier to write down, since the sum of an expression involving observable $A$ and analogous one for observable $B$ becomes an indexed sum. We also allow much generality in the way errors are quantified. In earlier works, we relied on two elements to be chosen for each observable, namely a metric $D$ on the outcome set, and an error exponent $\alpha$, distinguishing, say absolute ($\alpha = 1$), root-mean-square ($\alpha = 2$), and maximal ($\alpha = \infty$) deviations. Deviations were then averages of $D(x, y)^\alpha$. Here we generalize further to an arbitrary *cost function* $c : X \times X \to \mathbb{R}$, which we take to be positive, and zero exactly on the diagonal (e.g., $c(x, y) = D(x, y)^\alpha$), but not necessarily symmetric. Again this generality comes mostly as a simplification of notation. For a reference observable $A$ with outcome set $X$ and an approximate version $A'$ with the same outcome set, this defines an error $\varepsilon(A'|A)$. Our aim is to provide algorithms for computing the *uncertainty diagram* associated with such data, of which Fig. 2 gives an example. The given data for such a diagram are $n$ projection valued observables $A_1, \ldots, A_n$, with outcome sets $X_i$, for each of which we are given also a cost function $c_i : X_i \times X_i \to \mathbb{R}$ for quantifying errors. An approximate joint measurement is then an observable $R$ with outcome set $\bigtimes_i X_i$, and hence with POVM elements $R(x_1, \ldots, x_n)$, where $x_i \in X_i$. By ignoring every output but one we get the $n$ marginal observables

$$A'_i(x_i) = \sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} R(x_1, \ldots, x_n) \tag{1}$$

and a corresponding tuple

$$\vec{\varepsilon}(R) = \big(\varepsilon(A'_1|A_1), \ldots, \varepsilon(A'_n|A_n)\big) \tag{2}$$

of errors. The set $\mathcal{U}_L$ of such tuples, as $R$ runs over all joint measurements, is the *uncertainty region*. The surface bounding this set from below describes the uncertainty tradeoffs. For $n = 2$ we call it the tradeoff curve. Measurement uncertainty is the phenomenon that, for general reference observables $A_i$, the uncertainty region is bounded away from the origin. In principle there are many ways to express this mathematically, from a complete characterization of the exact *tradeoff* curve, which is usually hard to get, to bounds which are simpler to state, but suboptimal. Linear bounds will play a special role in this paper.

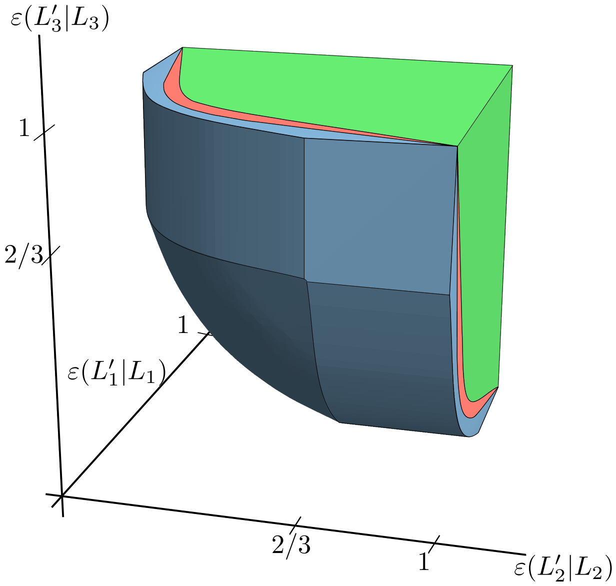


Figure 2: Uncertainty regions for three reference observables, namely the angular momentum components $L_1, L_2, L_3$ for spin 1, each with outcome set $X = \{-1, 0, +1\}$ and the choice $c(x, y) = (x - y)^2$ for the cost function. The three regions indicated correspond to the different overall figures of merit $\varepsilon_M(A'|A) \geq \varepsilon_C(A'|A) \geq \varepsilon_E(A'|A)$ described in Sect. 2.

We will consider three ways to build a single error quantity out of the comparison of distributions, denoted by $\varepsilon_M(A'|A)$, $\varepsilon_C(A'|A)$, and $\varepsilon_E(A'|A)$. These will be defined in Sect. 2. For every choice of observables and cost functions, each will give an uncertainty region, denoted by $\mathcal{U}_M$, $\mathcal{U}_C$, and $\mathcal{U}_E$, respectively. Since the errors are all based on the same cost function $c$, they are directly comparable (see Fig. 2). We show in Sect. 3 that the three regions are convex, and hence characterized completely by linear bounds. In Sect. 4 we show how to calculate the optimal linear lower bounds by semidefinite programs. Finally, an Appendix collects the basic information on the beautiful theory of optimal transport, which is needed in Sects. 2.1 and 4.1.

## 2 Deviation measures for observables

Here we define the measures we use to quantify how well an observable $A'$ approximates a desired observable $A$. In this section we do not use the marginal condition (1), so $A'$ is an arbitrary observable with the same outcome set $X$ as $A$, i.e., we drop all indices $i$ identifying the different observables. Our error quantities are operational in the sense that each is motivated by an experimental setup, which will in particular provide a natural way to measure them. All error definitions are based on the same cost function $c : X \times X \to \mathbb{R}$, where $c(x, y)$ is the “cost” of getting a result $x \in X$, when $y \in X$ would have been correct. The only assumptions are that $c(x, y) \geq 0$ with $c(x, y) = 0$ iff $x = y$.

As described above, we consider a quantum system with Hilbert space $\mathbb{C}^d$. As a reference observable $A$ we allow any complete von Neumann measurement on this system, that is, any observable whose the set $X$ of possible measurement outcomes has size $|X| = d$ and whose POVM elements $A(y) \in \mathcal{B}(\mathbb{C}^d)$ $(y \in X)$ are mutually orthogonal projectors of rank 1; we can then also write $A(y) = |\phi_y\rangle\langle\phi_y|$ with an orthonormal basis $\{\phi_y\}$ of $\mathbb{C}^d$. For the approximating observable $A'$ the

POVM elements $A'(x)$ (with $x \in X$) are arbitrary with $A'(x) \geq 0$ and $\sum_{x \in X} A(x) = \mathbb{1}$.

The comparison will be based on a comparison of output distributions, for which we use the following notations: Given a quantum state $\rho$ on this system, i.e., a density operator with $\rho \geq 0$ and $\operatorname{tr} \rho = 1$, and an observable such as $A$, we will denote the outcome distribution by $\rho A$, so $(\rho A)(y) := \operatorname{tr}(\rho A_y)$. This is a probability distribution on the outcome set $X$ and can be determined physically as the empirical outcome distribution after many experiments.

For comparing just two probability distributions $p : X \to \mathbb{R}_+$ and $q : X \to \mathbb{R}_+$, a canonical choice is the "minimum transport cost"

$$\check{c}(p, q) := \inf_{\gamma} \Big\{ \sum_{xy} c(x, y)\gamma(x, y) \bigm| \gamma \text{ couples } p \text{ to } q \Big\}, \tag{3}$$

where the infimum runs over the set of all *couplings*, or "transport plans" $\gamma : X \times X \to \mathbb{R}_+$ of $p$ to $q$, i.e., the set of all probability distributions $\gamma$ satisfying the marginal conditions $\sum_y \gamma(x, y) = p(x)$ and $\sum_x \gamma(x, y) = q(y)$. The motivations for this notion, and the methods to compute it efficiently are described in the Appendix. Since $X$ is finite, the infimum is over a compact set, so it is always attained. Moreover, since we assumed $c \geq 0$ and $c(x, y) = 0 \Leftrightarrow x = y$, we also have $\check{c}(p, q) \geq 0$ with equality iff $p = q$. If one of the distributions, say $q$, is concentrated on a point $\widetilde{y}$, only one coupling exists, namely $\gamma(x, y) = p(x)\delta_{y\widetilde{y}}$. In this case we abbreviate $\check{c}(p, q) = \check{c}(p, \widetilde{y})$, and get

$$\check{c}(p, \widetilde{y}) = \sum_x p(x) c(x, \widetilde{y}), \tag{4}$$

i.e., the average cost of moving all the points $x$ distributed according to $p$ to $\widetilde{y}$.

## 2.1 Maximal measurement error $\varepsilon_M(A'|A)$.

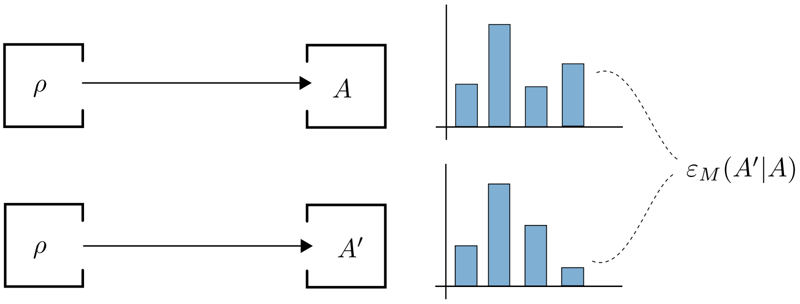


Figure 3: For the maximal measurement error $\varepsilon_M(A'|A)$ the transport distance of output distributions is maximized over all input states $\rho$.

The worst case error over *all* input states is

$$\varepsilon_M(A'|A) := \sup_{\rho} \big\{ \check{c}(\rho A', \rho A) \bigm| \rho \text{ quantum state on } \mathbb{C}^d \big\}, \tag{5}$$

which we call the *maximal measurement error*. Note that, like the cost function $c$ and the transport costs $\check{c}$, the measure $\varepsilon_M(A'|A)$ need not be symmetric in its arguments, which is sensible as the reference and approximating observables have distinct roles. Similar definitions for the deviation of an approximating measurement from an ideal one have been made, for specific cost functions, in [4, 6] and [8] before.

The definition (5) makes sense even if the reference observable $A$ is not a von Neumann measurement. Instead, the only requirement is that $A$ and $A'$ be general observables with the same (finite) outcome set $X$, not necessarily of size $d$. All our results below that involve only the maximal measurement error immediately generalize to this case as well.

One can see that it is expensive to determine the quantity $\varepsilon_M(A'|A)$ experimentally according to the definition: one would have to measure and compare the outcome statistics $\rho A'$ and $\rho A$ for all possible input states $\rho$, which form a continuous set. The following definition of observable deviation alleviates this burden.

## 2.2 Calibration error $\varepsilon_C(A'|A)$.

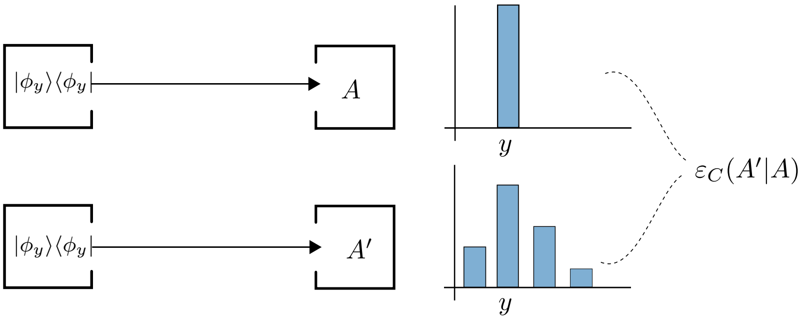

$|\phi_y\rangle\langle\phi_y|$
$A$
$y$
$\varepsilon_C(A'|A)$
$|\phi_y\rangle\langle\phi_y|$
$A'$
$y$

Figure 4: For the calibration error $\varepsilon_C(A'|A)$, the input state is constrained to the eigenstates of $A$, say with sharp $A$-value $y$, and the cost of moving the $A'$-distribution to $y$ is maximized over $y$.

Calibration is a process by which one tests a measuring device on inputs (or measured objects) for which the "true value" is known. Even in quantum mechanics we can set this up by demanding that the measurement of the reference observable on the input state gives a sharp value $y$. In a general scenario with continuous outcomes this can only be asked with a finite error $\delta$, which goes to zero at the end [4], but in the present finite scenario we can just demand $(\rho A)(y) = 1$. Since, for every outcome $y$ of a von Neumann measurement, there is only one state with this property (namely $\rho = |\phi_y\rangle\langle\phi_y|$) we can simplify even further, and define the *calibration error* by

$$\varepsilon_C(A'|A) := \sup_{y,\rho}\{\check{c}(\rho A', y) \bigm| \operatorname{tr}(\rho A(y)) = 1\} = \max_y \sum_x \langle\phi_y|A'(x)|\phi_y\rangle\; c(x, y). \tag{6}$$

Note that the calibration idea only makes sense when there are sufficiently many states for which the reference observable has deterministic outcomes, i.e., for projective observables $A$.

A closely related quantity has recently been proposed by Appleby [2]. It is formulated for real valued quantities with cost function $c(x, y) = (x - y)^2$, and has the virtue that it can be expressed entirely in terms of first and second moments of the probability distributions involved. So for any $\rho$, let $m$ and $v$ be the mean and variance of $\rho A$, and $v'$ the mean quadratic deviation of $\rho A'$ from $m$. Then Appleby defines

$$\varepsilon_D(A'|A) = \sup_\rho(\sqrt{v'} - \sqrt{v})^2. \tag{7}$$

Here we added the square to make Appleby's quantity comparable to our variance-like (rather than standard deviation-like) quantities, and chose the letter $D$, because Appleby calls this the $D$-error. Since in the supremum we have also the states for which $A$ has a sharp distribution (i.e. $v = 0$), we clearly have $\varepsilon_D(A'|A) \geq \varepsilon_C(A'|A)$. On the other hand, let $\Phi(x) = t(x - m)^2$ and $\Psi(y) = t/(1-t)(y-m)^2$ with some parameter $t \in (-\infty, 1)$. Then one easily checks that $\Phi(x) - \Psi(y) \leq (x-y)^2$, so $(\Phi, \Psi)$ is a pricing scheme in the sense defined in the Appendix. Therefore

$$\check{c}(\rho A', \rho A) \geq \sum_x (\rho A')(x)\Phi(x) - \sum_y (\rho A)(y)\Psi(y) = t\, v' - \frac{t}{1-t}\, v. \tag{8}$$

Maximizing this expression over $t$ gives exactly (7). Therefore $\varepsilon_C(A'|A) \leq \varepsilon_D(A'|A) \leq \varepsilon_M(A'|A)$.

## 2.3 Entangled reference error $\varepsilon_E(A'|A)$.

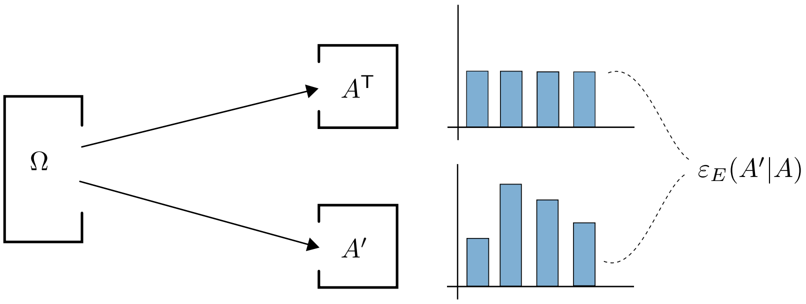


Figure 5: The entangled reference error $\varepsilon_E(A'|A)$ is a single expectation value, namely of the cost $c(x, y)$, where $y$ is the output of $A^\mathsf{T}$ and $x$ the output of $A'$. Like the other error quantities this expectation vanishes iff $A' = A$.

In quantum information theory a standard way of providing a reference state for later comparison is by applying a channel or observable to one half of a maximally entangled system. Two observables would be compared by measuring them (or suitable modifications) on the two parts of a maximally entangled system. Let us denote the entangled vector by $\Omega = d^{-1/2} \sum_k |kk\rangle$. Since later we will look at several distinct reference observables, the basis kets $|k\rangle$ in this expression have no special relation to $A$ or its eigenbasis $\phi_y$. We denote by $X^\mathsf{T}$ the transpose of an operator $X$ in the $|k\rangle$ basis, and by $A^\mathsf{T}$ the observable with POVM elements $A(y)^\mathsf{T} = |\overline{\phi_y}\rangle\langle\overline{\phi_y}|$, where $\overline{\phi_y}$ is the complex conjugate of $\phi_y$ in $|k\rangle$-basis. These transposes are needed due to the well-known relation $(X \otimes \mathbb{1})\Omega = (\mathbb{1} \otimes X^\mathsf{T})\Omega$. We now consider an experiment, in which $A'$ is measured on the first part and $A^\mathsf{T}$ on the second part of the entangled system, so we get the outcome pair $(x, y)$ with probability

$$p(x, y) = \langle\Omega|A'(x) \otimes A(y)^\mathsf{T}|\Omega\rangle = \langle\Omega|A'(x)A(y) \otimes \mathbb{1}|\Omega\rangle = \frac{1}{d}\operatorname{tr}\big(A'(x)A(y)\big). \tag{9}$$

As $A$ is a complete von Neumann measurement, this probability distribution is concentrated on the diagonal ($x = y$) iff $A' = A$, i.e., there are no errors of $A'$ relative to $A$. Averaging with the error costs we get a quantity we call the *entangled reference error*

$$\varepsilon_E(A'|A) := \sum_{xy} \frac{1}{d}\operatorname{tr}\big(A'(x)A(y)\big)\ c(x, y). \tag{10}$$

Note that this quantity is measured as a single expectation value in the experiment with source $\Omega$. Moreover, when we later want to measure different such deviations for the various marginals, the source and the tested joint measurement device can be kept fixed, and only the various reference observables $A_i^\mathsf{T}$ acting on the second part need to be adapted suitably.

## 2.4 Summary and comparison

The quantities $\varepsilon_M(A'|A)$, $\varepsilon_C(A'|A)$ and $\varepsilon_E(A'|A)$ constitute three different ways to quantify the deviation of an observable $A'$ from a projective reference observable $A$. Nevertheless, they are all based on the same distance-like measure, the cost function $c$ on the outcome set $X$. Therefore it makes sense to compare them quantitatively. Indeed, they are ordered as follows:

$$\varepsilon_M(A'|A) \geq \varepsilon_C(A'|A) \geq \varepsilon_E(A'|A). \tag{11}$$

Here the first inequality follows by restricting the supremum (5) to states which are sharp for $A$, and the second by noting the (6) is the maximum of a function of $y$, of which (10) is the average.

Moreover, as we argued before Eq. (10), $\varepsilon_E(A'|A) = 0$ if and only if $A = A'$, which is hence equivalent also to $\varepsilon_M(A'|A) = 0$ and $\varepsilon_C(A'|A) = 0$.

## 3 Convexity of uncertainty diagrams

In this section we will consider tuples $(A_1, \ldots, A_n)$ of projection valued non-degenerated observables, as described in the introduction. We will collect some basic properties of the uncertainty regions $\mathcal{U}_L$, where $L \in \{M, C, E\}$, that is,

$$\mathcal{U}_L := \Big\{ \big(\varepsilon_L(A'_1|A_1), \ldots, \varepsilon_L(A'_n|A_n)\big) \;\Big|\; A'_i \text{ marginals of a joint measurement} \Big\}. \tag{12}$$

For two observables $B_1$ and $B_2$ with the same outcome set we can easily realize their mixture, or convex combination $B = tB_1 + (1-t)B_2$ by flipping a coin with probability $t$ for heads in each instance and then apply $B_1$ when heads is up and $B_2$ otherwise. In terms of POVM elements this reads $B(x) = tB_1(x) + (1-t)B_2(x)$. We show first that this mixing operation does not increase the error quantities from Sect. 2.

**Lemma 1.** *For $L \in \{M, D, C, E\}$ the error quantity $\varepsilon_L(B|A)$, is a convex function of $B$ , i. e. for $B = tB_1 + (1-t)B_2$ and $t \in [0,1]$:*

$$\varepsilon_L(B|A) \leq t\, \varepsilon_L(B_1|A) + (1-t)\, \varepsilon_L(B_1|A). \tag{13}$$

*Proof.* The basic fact used here is that the pointwise supremum of affine functions (i.e., those for which equality holds in the definition of a convex function) is convex. This is geometrically obvious, and easily verified from the definitions. Hence we only have to check that each of the error quantities is indeed represented as a supremum of functions, which are affine in the observable $B$.

For $L = E$ we even get an affine function, because (10) is linear in $A'$. For $L = C$ equation (6) has the required form. For $L = M$ the definition (5) is as a supremum, but the function $\check{c}$ is defined as an infimum. However, we can use the duality theory described in the Appendix (e.g. in (49)) to write it instead as a supremum over pricing schemes, of an expression which is just the expectation of $\Phi(x)$ plus a constant, and therefore an affine function. Finally, for Appleby's case (7), we get the same supremum, but over a subset of pricing schemes (the quadratic ones, see below (7)). □

The convexity of the error quantities distinguishes measurement from preparation uncertainty. Indeed, the variances appearing in preparation uncertainty relations are typically concave functions, because they arise from minimizing the expectation of $(x-m)^2$ over $m$. Consequently, the preparation uncertainty regions may have gaps, and non-trivial behaviour on the side of large variances. The following proposition will show that measurement uncertainty regions are better behaved.

For every cost function $c$ on a set $X$ we can define a "radius" $\overline{c}^*$, the largest transportation cost from the uniform distribution (the "center" of the set of probability distributions) and a "diameter" $c^*$, the largest transportation cost between any two distributions:

$$\overline{c}^* = \max_y \sum_x c(x,y)/d \qquad\qquad c^* = \max_{xy} c(x,y). \tag{14}$$

**Proposition 1.** *Let $n$ observables $A_i$ and cost functions $c_i$ be given, and define $c_i^M = c_i^C = c_i^*$ and $c_i^E = \overline{c_i}^*$. Then, for $L \in \{M, C, E\}$, the uncertainty regions $\mathcal{U}_L$ is a convex set and has the* *following (monotonicity) property: When $\vec{x} = (x_1, \dots, x_n) \in \mathcal{U}_L$ and $\vec{y} = (y_i, \dots, y_n) \in \mathbb{R}^n$ such that $x_i \leq y_i \leq c_i^L$, then $\vec{y} \in \mathcal{U}_L$.*

*Proof.* Let us first clarify how to make the worst possible measurement $B$, according to the various error criteria, for which we go back to the setting of Sect. 2, with just one observable $A$, and cost function $c$. In all cases, the worst measurement is one with constant and deterministic output, i.e., $B(x) = \delta_{x^*,x}\mathbb{1}$. For $L = C$ and $L = M$ such a measurement will have $\varepsilon_L(B|A) = \max_y c(x^*, y)$, and we can choose $x^*$ to make this equal to $c^* = c^L$. For $L = E$ we get instead the average, which is maximized by $\overline{c}^*$.

We can now make a given joint measurement $R$ worse by replacing it partly by a bad one, say for the first observable $A_1$. That is, we set, for $\lambda \in [0, 1]$,

$$\widetilde{R}(x_1, x_2, \dots, x_n) = \lambda B_1(x_1) \sum_{y_1} R(y_1, x_2, \dots, x_n) + (1 - \lambda) R(x_1, x_2, \dots, x_n). \tag{15}$$

Then all marginals $\widetilde{A}_i'$ for $i \neq 1$ are unchanged, but $\widetilde{A}_1'(x_1) = \lambda B_1(x_1) + (1 - \lambda) A'(x_1)$. Now as $\lambda$ changes from 0 to 1, the point in the uncertainty diagram will move continuously in the first coordinate direction from $\vec{x}$ to the point in which the first coordinate is replaced by its maximum value (see Fig. 6(left)). Obviously, the same holds for every other coordinate direction, which proves the monotonicity statement of the proposition.

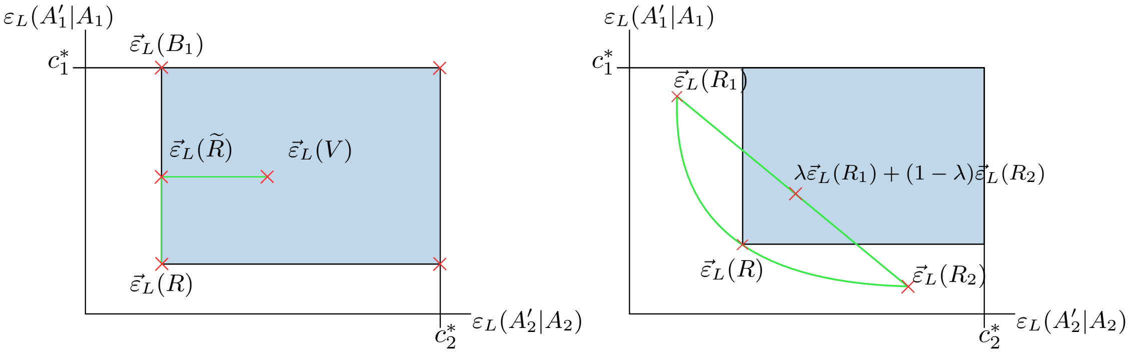

$\varepsilon_L(A_1'|A_1)$
$\vec{\varepsilon}_L(B_1)$
$c_1^*$
$\vec{\varepsilon}_L(\widetilde{R})$
$\vec{\varepsilon}_L(V)$
$\vec{\varepsilon}_L(R)$
$c_2^*$
$\varepsilon_L(A_2'|A_2)$
$\varepsilon_L(A_1'|A_1)$
$c_1^*$
$\vec{\varepsilon}_L(R_1)$
$\lambda\vec{\varepsilon}_L(R_1) + (1-\lambda)\vec{\varepsilon}_L(R_2)$
$\vec{\varepsilon}(R)$
$\vec{\varepsilon}_L(R_2)$
$c_2^*$
$\varepsilon_L(A_2'|A_2)$

Figure 6: The blue shaded region corresponds to the monotonicity statement for $\vec{\varepsilon}_L(R)$. (left) $\widetilde{R}$ is a mixture of $R$ and $B_1$. We can also get an observable $V$ by mixing the second marginal of $\widetilde{R}$ with $B_2$ and thus reach every point in the blue shaded region. (right) $\vec{\varepsilon}_L(R)$ is componentwise convex. So the mixture of the points $\vec{\varepsilon}_L(R_1)$ and $\vec{\varepsilon}_L(R_2)$ is always in the monotonicity region corresponding to $\vec{\varepsilon}_L(R)$.

Let $R_1$ and $R_2$ be two observables, and let $R = \lambda R_1 + (1 - \lambda) R_2$ be their mixture. For proving the convexity of $\mathcal{U}_L$ we will have to show that every point on the line between $\vec{\varepsilon}_L(R_1)$ and $\vec{\varepsilon}_L(R_2)$ can be attained by a tuple of errors corresponding to some allowed observable (see Fig. 6 (right)). Now lemma 1 tells us that every component of $\vec{\varepsilon}_L(R)$ is convex, which implies that $\vec{\varepsilon}_L(R) \leq \lambda\vec{\varepsilon}_L(R_1) + (1 - \lambda)\vec{\varepsilon}_L(R_2)$. But, by monotonicity, this also means that $\lambda\vec{\varepsilon}_L(R_1) + (1 - \lambda)\vec{\varepsilon}(R_2)$ is in $\mathcal{U}_L$ again, which shows the convexity of $\mathcal{U}_L$. □

## 3.1 Example: Phase space pairs

As is plainly visible from Fig. 2, the three error criteria considered here usually give different results. However, under suitable circumstances they all coincide. This is the case for conjugate pairs related by Fourier transform [24]. The techniques needed to show this are the same as for the standard position/momentum case [6, 22], and in addition imply that the region for preparation uncertainty is also the same.

In the finite case there is not much to choose: We have to start from a finite abelian group, which we think of as position space, and its dual group, which is then the analogue of momentum space. The unitary connecting the two observables is the finite Fourier associated with the group. The cost function needs to be translation invariant, i.e., $c(x, y) = c(x - y)$. Then, by an averaging argument, we find for all error measures that a covariant phase space observable minimizes measurement uncertainty (all three versions). The marginals of such an observable can be simulated by first doing the corresponding reference measurement, and then adding some random noise. This implies [8] that $\varepsilon_M(A'|A) = \varepsilon_C(A'|A)$. But we know more about this noise: It is independent of the input state so that the average and the maximum of the noise (as a function of the input) coincide, i.e., $\varepsilon_C(A'|A) = \varepsilon_E(A'|A)$. Finally, we know that the noise of the position marginal is distributed according to the position distribution of a certain quantum state which is, up to normalization and a unitary parity inversion, the POVM element of the covariant phase space observable at the origin. The same holds for the momentum noise. But then the two noise quantities are exactly related like the position and momentum distributions of a state, and the tradeoff curve for that problem is exactly preparation uncertainty, with variance criteria based on the same cost function.

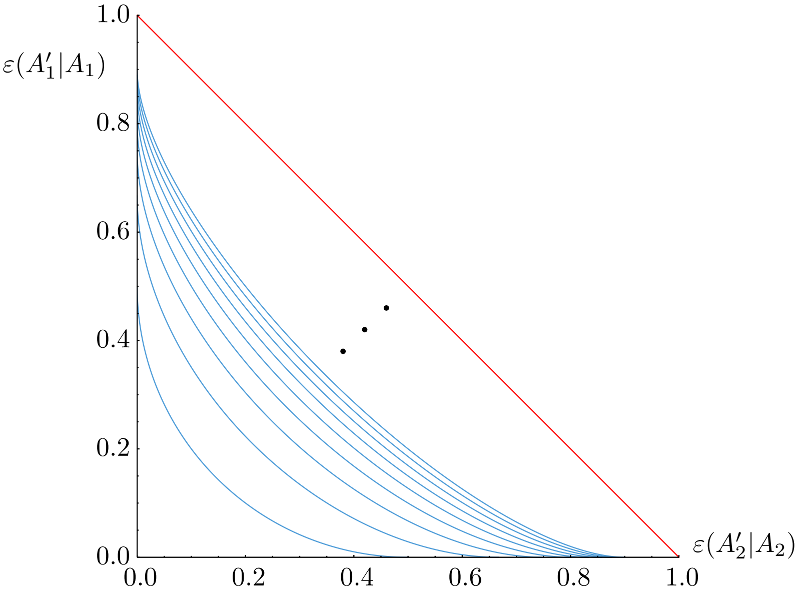


Figure 7: The uncertainty tradeoff curves for discrete position/momentum pairs, with discrete metric. In this case all uncertainty regions, also the one for preparation uncertainty, coincide. The parameter of the above tradeoff curves is the order $d = 2, 3, \ldots, 10, \cdots, \infty$ of the underlying abelian group.

If we choose the discrete metric for $c$, the uncertainty region depends only on the number $d$ of elements in the group we started from [24]. The largest $\varepsilon$ for all quantities is the distance from a maximally mixed state to any pure state, which is $\Delta = (1 - 1/d)$. The exact tradeoff curve is then an ellipse, touching the axes at the points $(0, \Delta)$ and $(\Delta, 0)$. The resulting family of curves, parameterized by $d$, is shown in Fig. 7. In general, however, the tradeoff curve requires the solution of a non-trivial family of ground state problems, and cannot be given in closed form. For bit strings of length $n$, and the cost some convex function of Hamming distance there is an expression for large $n$ [24].

# 4 Computing uncertainty regions via semidefinite programming

We show here how the uncertainty regions – and therefore optimal uncertainty relations – corresponding to each of the three error measures can actually be computed, for any given set of projective observables $A_1, \ldots, A_n$ and cost functions $c_1, \ldots, c_n$. Our algorithms will come in the form of semidefinite programs (SDPs) [20, 19], facilitating efficient numerical computation of the uncertainty regions via the many existing program packages to solve SDPs. Moreover, the accuracy of such numerical results can be rigorously certified via the duality theory of SDPs. To obtain the illustrations in this paper we used the CVX package [10, 9] under MATLAB.

As all our uncertainty regions $\mathcal{U}_L \subset \mathbb{R}^n$ (for $L = M, C, E$) are convex and closed (Sect. 3), they are completely characterized by their supporting hyperplanes (for a reference to convex geometry see [17]). Due to the monotonicity property stated in Prop. 1 some of these hyperplanes just cut off the set parallel along the planes $x_i = c_i^L$. The only hyperplanes of interest are thus those with nonnegative normal vectors $\vec{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n_+$ (see Fig. 8). Each hyperplane is completely specified by its "offset" $b_L(\vec{w})$ away from the origin, and this function determines $\mathcal{U}_L$:

$$b_L(\vec{w}) \quad := \quad \inf\left\{ \vec{w}\cdot\vec{\varepsilon} \;\middle|\; \vec{\varepsilon}\in\mathcal{U}_L \right\}, \tag{16}$$

$$\mathcal{U}_L \quad = \quad \left\{ \vec{\varepsilon}\in\mathbb{R}^n \;\middle|\; \forall\vec{w}\in\mathbb{R}^n_+ :\ \vec{w}\cdot\vec{\varepsilon} \geq b_L(\vec{w}) \text{ and } \forall i : \varepsilon_i \leq c_i^L \right\}. \tag{17}$$

In fact, due to homogeneity $b_L(t\vec{w}) = t\, b_L(\vec{w})$ we can restrict everywhere to the subset of vectors $\vec{w} \in \mathbb{R}^n_+$ that, for example, satisfy $\sum_i w_i = 1$, suggesting an interpretation of the $w_i$ as weights of the different uncertainties $\varepsilon_i$. Our algorithms will, besides evaluating $b_L(\vec{w})$, also allow to compute an (approximate) minimizer $\vec{\varepsilon}$, so that one can plot the boundary of the uncertainty region $\mathcal{U}_L$ by sampling over $\vec{w}$, which is how the figures in this paper were obtained.

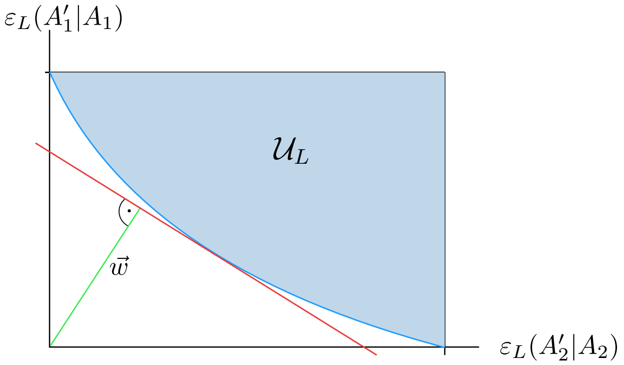


Figure 8: The lower bound of the uncertainty region $\mathcal{U}_L$ can be described by its supporting hyperplanes (red line) with a normal vector $\vec{w} \in \mathbb{R}^n_+$ .

Let us further note that knowledge of $b_L(\vec{w})$ for some $\vec{w} \in \mathbb{R}^n_+$ immediately yields a quantitative uncertainty relation: every error tuple $\vec{\varepsilon} \in \mathcal{U}_L$ attainable via a joint measurement is constrained by the affine inequality $\vec{w}\cdot\vec{\varepsilon} \geq b_L(\vec{w})$, meaning that some weighted average of the attainable error quantities $\varepsilon_i$ cannot become too small. When $b_L(\vec{w}) > 0$ is strictly positive, this excludes in particular the zero error point $\vec{\varepsilon} = \vec{0}$. The obtained uncertainty relations are *optimal* in the sense that there exists $\vec{\varepsilon} \in \mathcal{U}_L$ which attains strict equality $\vec{w}\cdot\vec{\varepsilon} = b_L(\vec{w})$.

Having reduced the computation of an uncertainty region essentially to determining $b_L(\vec{w})$ (possibly along with an optimizer $\vec{\varepsilon}$), we now treat each case $L = M, C, E$ in turn.

## 4.1 Computing the uncertainty region $\mathcal{U}_M$

On the face of it, the computation of the offset $b_M(\vec{w})$ looks daunting: expanding the definitions we obtain

$$b_M(\vec{w}) = \inf_R \sum_{i=1}^{n} w_i \sup_\rho \check{c}_i(\rho A_i', \rho A_i), \tag{18}$$

where the infimum runs over all joint measurements $R$ with outcome set $X_1 \times \ldots \times X_n$, inducing the marginal observables $A_i' = A_i'(R)$ according to (1), and the supremum over all sets of $n$ quantum states $\rho_1, \ldots, \rho_n$, and where the transport costs $\check{c}_i(p, q)$ are given as a further infimum (3) over the couplings $\gamma_i$ of $p = \rho A_i'$ and $q = \rho A_i$.

The first simplification is to replace the infimum over each coupling $\gamma_i$, via a dual representation of the transport costs, by a maximum over *optimal pricing schemes* $(\Phi_\alpha, \Psi_\alpha)$, which are certain pairs of functions $\Phi_\alpha, \Psi_\alpha : X_i \to \mathbb{R}$, where $\alpha$ runs over some finite label set $\mathcal{S}_i$. The characterization and computation of the pairs $(\Phi_\alpha, \Psi_\alpha)$, which depend only on the chosen cost function $c_i$ on $X_i$, is described in the Appendix. The simplified expression for the optimal transport costs is then

$$\check{c}_i(p, q) = \max_{\alpha \in \mathcal{S}_i} \sum_x \Phi_\alpha(x)\, p(x) - \sum_y \Psi_\alpha(y) q(y). \tag{19}$$

We can then continue our computation of $b_M(\vec{w})$:

$$\begin{aligned} b_M(\vec{w}) &= \inf_R \sum_i w_i \sup_\rho \max_{\alpha \in \mathcal{S}_i} \Big( \sum_x \Phi_\alpha(x) \operatorname{tr}[\rho A_i'(x)] - \sum_y \Psi_\alpha(y) \operatorname{tr}[\rho A_i(y)] \Big) && (20) \\ &= \inf_R \sum_i w_i \max_{\alpha \in \mathcal{S}_i} \sup_\rho \operatorname{tr}\Big[ \rho \Big( \sum_x \Phi_\alpha(x) A_i'(x) - \sum_y \Psi_\alpha(y) A_i(y) \Big) \Big] && (21) \\ &= \inf_R \sum_i w_i \max_{\alpha \in \mathcal{S}_i} \lambda_{\max} \Big( \sum_x \Phi_\alpha(x) A_i'(x) - \sum_y \Psi_\alpha(y) A_i(y) \Big), && (22) \end{aligned}$$

where $\lambda_{\max}(B_{i,\alpha})$ denotes the maximum eigenvalue of a Hermitian operator $B_{i,\alpha}$. Note that $\lambda_{\max}(B_{i,\alpha}) = \inf\{\mu_i \mid B_{i,\alpha} \leq \mu_i \mathbb{1}\}$, which one can also recognize as the dual formulation of the convex optimization $\sup_\rho \operatorname{tr}(\rho B_{i,\alpha})$ over density matrices, so that

$$\max_{\alpha \in \mathcal{S}_i} \lambda_{\max}(B_{i,\alpha}) = \inf\{\mu_i \mid \forall \alpha \in \mathcal{S}_i : \, B_{i,\alpha} \leq \mu_i \mathbb{1}\} \tag{23}$$

We obtain thus a single constrained minimization:

$$b_M(\vec{w}) = \inf_{R, \{\mu_i\}} \Big\{ \sum_i w_i \mu_i \;\Big|\; \forall i \forall \alpha \in \mathcal{S}_i : \sum_x \Phi_\alpha(x) A_i'(x) - \sum_y \Psi_\alpha(y) A_i(y) \leq \mu_i \mathbb{1} \Big\}. \tag{24}$$

Making the constraints on the POVM elements $R(x_1, \ldots, x_n)$ of the joint observable $R$ explicit and expressing the maginal observables $A_i' = A_i'(R)$ directly in terms of them by (1), we finally obtain the following SDP representation for the quantity $b_M(\vec{w})$:

$$\boxed{\begin{array}{l} b_M(\vec{w}) = \inf \;\; \sum_i w_i \mu_i \\ \text{with real variables } \mu_i \text{ and } d \times d\text{-matrix variables } R(x_1, \ldots, x_n) \text{ subject to} \\ \begin{array}{rl} \mu_i \mathbb{1} & \geq \sum_{x_1, \ldots, x_n} \Phi_\alpha(x_i)\, R(x_1, \ldots, x_n) - \sum_y \Psi_\alpha(y) A(y) \quad \forall i \, \forall \alpha \in \mathcal{S}_i \\ R(x_1, \ldots, x_n) & \geq 0 \quad \forall x_1, \ldots, x_n \\ \sum_{x_1, \ldots, x_n} R(x_1, \ldots, x_n) & = \mathbb{1}. \end{array} \end{array}} \tag{25}$$

The derivation above shows further that, when $w_i > 0$, the $\mu_i$ attaining the infimum equals $\mu_i = \sup_\rho \check{c}_i(\rho A_i', \rho A_i) = \varepsilon_M(A_i'|A_i)$, where $A_i'$ is the marginal coming from a corresponding optimal joint measurement $R(x_i, \dots, x_n)$. Since numerical SDP solvers usually output an (approximate) optimal variable assignment, one obtains in this way directly a boundary point $\vec{\varepsilon} = (\mu_1, \dots, \mu_n)$ of $\mathcal{U}_M$ when all $w_i$ are strictly positive. If $w_i = 0$ vanishes, a corresponding boundary point $\vec{\varepsilon}$ can be computed via $\varepsilon_i = \varepsilon_M(A'_i|A_i) = \max_{\alpha \in \mathcal{S}_i} \lambda_{\max}(\sum_{x_1,\dots,x_n} \Phi_\alpha(x_i)\, R(x_1, \dots, x_n) - \sum_y \Psi_\alpha(y) A(y))$ from an optimal assignment for the POVM elements $R(x_1, \dots, x_n)$.

For completeness we also display the corresponding dual program [20] (note that strong duality holds, and the optima of both the primal and the dual problem are attained):

$$\boxed{\begin{array}{rl} \multicolumn{2}{l}{b_M(\vec{w}) = \sup \ \operatorname{tr}[C] - \sum_{i,\alpha} \operatorname{tr}[D_{i,\alpha} \sum_y \Psi_\alpha(y) A_i(y)]} \\ \multicolumn{2}{l}{\text{with } d \times d\text{-matrix variables } C \text{ and } D_{i,\alpha} \text{ subject to}} \\ C & \leq \sum_{i,\alpha} \Phi_\alpha(x_i) D_{i,\alpha} \quad \forall x_1, \dots, x_n \\ 0 & \leq D_{i,\alpha} \quad \forall i \, \forall \alpha \in \mathcal{S}_i \\ w_i & = \sum_\alpha \operatorname{tr}[D_{i,\alpha}] \quad \forall i. \end{array}} \tag{26}$$

## 4.2 Computing the uncertainty region $\mathcal{U}_C$

To compute the offset function $b_C(\vec{w})$ for the calibration uncertainty region $\mathcal{U}_C$ we use the last form in (6) and recall that the projectors onto the sharp eigenstates of $A_i$ (see Sect. 2.2) are exactly the POVM elements $A_i(x)$ for $x \in X_i$:

$$b_C(\vec{w}) = \inf_R \sum_i w_i \max_y \sum_x \operatorname{tr}[A'_i(x) A_i(y)] c_i(x, y) \tag{27}$$

$$= \inf_R \sum_i w_i \sup_{\{\lambda_{i,y}\}} \sum_y \lambda_{i,y} \sum_x \operatorname{tr}[A'_i(x) A_i(y)] c_i(x, y) \tag{28}$$

$$= \inf_R \sup_{\{\lambda_{i,y}\}} \sum_{x_1,\dots,x_n} \operatorname{tr}\Big[R(x_1, \dots, x_n) \sum_{i,y} w_i \lambda_{i,y} c_i(x_i, y) A_i(y)\Big] \tag{29}$$

where again the infimum runs over all joint measurements $R$, inducing the marginals $A'_i$, and we have turned, for each $i = 1, \dots, n$, the maximum over $y$ into a linear optimization over probabilities $\lambda_{i,y} \geq 0$ $(y = 1, \dots, d)$ subject to the normalization constraint $\sum_y \lambda_{i,y} = 1$. In the last step, we have made the $A'_i$ explicit via (1).

The first main step towards a tractable form is von Neumann's minimax theorem [14, 18]: As the sets of joint measurements $R$ and of probabilities $\{\lambda_{i,y}\}$ are both convex and the optimization function is an affine function of $R$ and, separately, also an affine function of the $\{\lambda_{i,y}\}$, we can interchange the infimum and the supremum:

$$b_C(\vec{w}) = \sup_{\{\lambda_{i,y}\}} \inf_R \sum_{x_1,\dots,x_n} \operatorname{tr}\Big[R(x_1, \dots, x_n) \sum_{i,y} w_i \lambda_{i,y} c_i(x_i, y) A_i(y)\Big]. \tag{30}$$

The second main step is to use SDP duality [19] to turn the constrained infimum over $R$ into a supremum, abbreviating the POVM elements as $R(x_1, \dots, x_n) = R_\xi$:

$$\inf_{\{R_\xi\}} \Big\{ \sum_\xi R_\xi B_\xi \;\Big|\; R_\xi \geq 0 \ \forall \xi, \ \sum_\xi R_\xi = \mathbb{1} \Big\} \;=\; \sup_Y \Big\{ \operatorname{tr}[Y] \;\Big|\; Y \leq B_\xi \ \forall \xi \Big\}, \tag{31}$$

which is very similar to a dual formulation often employed in optimal ambiguous state discrimination [12, 25].

Putting everything together, we arrive at the following SDP representation for the offset quantity

$b_C(\vec{w})$:

$$\boxed{\begin{array}{l} b_C(\vec{w}) = \sup \ \ \mathrm{tr}[Y] \\ \text{with real variables } \lambda_{i,y} \text{ and a } d \times d\text{-matrix variable } Y \text{ subject to} \\ \begin{array}{rll} \qquad Y & \leq \sum_{i,y} w_i \lambda_{i,y} c_i(x_i, y) A_i(y) & \forall x_1, \ldots, x_n \\ \lambda_{i,y} & \geq 0 \quad \forall i \, \forall y & \\ \sum_y \lambda_{i,y} & = 1 \quad \forall i. & \end{array} \end{array}} \tag{32}$$

The dual SDP program reads (again, strong duality holds, and both optima are attained):

$$\boxed{\begin{array}{l} b_C(\vec{w}) = \inf \ \ \sum_i w_i m_i \\ \text{with real variables } m_i \text{ and } d \times d\text{-matrix variables } R(x_1, \ldots, x_n) \text{ subject to} \\ \begin{array}{rl} m_i & \geq \sum_{x_1,\ldots,x_n} \mathrm{tr}\big[R(x_1, \ldots, x_n) A_i(y)\big] c_i(x_i, y) \quad \forall i \, \forall y \\ R(x_1, \ldots, x_n) & \geq 0 \quad \forall x_1, \ldots, x_n \\ \sum_{x_1,\ldots,x_n} R(x_1, \ldots, x_n) & = \mathbb{1}. \end{array} \end{array}} \tag{33}$$

This dual version can immediately be recognized as a translation of Eq. (27) into SDP form, via an alternative way of expressing the maximum over $y$ (or via the linear programming dual of $\sup_{\{\lambda_{i,y}\}}$ from Eq. (29)).

To compute a boundary point $\vec{\varepsilon}$ of $\mathcal{U}_C$ lying on the supporting hyperplane with normal vector $\vec{w}$, it is best to solve the dual SDP (33) and obtain $\vec{\varepsilon} = (m_1, \ldots, m_n)$ from an (approximate) optimal assignment of the $m_i$. Again, this works when $w_i > 0$, whereas otherwise one can compute $\varepsilon_i = \max_y \sum_{x_1,\ldots,x_n} \mathrm{tr}[R(x_1, \ldots, x_n) A_i(y)] c_i(x_i, y)$ from an optimal assingment of the $R(x_1, \ldots, x_n)$. From many primal-dual numerical SDP solvers (such as CVX [10, 9]), one can alternatively obtain optimal POVM elements $R(x_1, \ldots, x_n)$ also from solving the primal SDP (32) as optimal dual variables corresponding to the constraints $Y \leq \ldots$, and compute $\vec{\varepsilon}$ from there.

### 4.3 Computing the uncertainty region $\mathcal{U}_E$

As one can see by comparing the last expressions in the defining equations (6) and (10), respectively, the evaluation of $b_E(\vec{w})$ is quite similar to (27), except that the maximum over $y$ is replaced by a uniform average over $y$. This simply corresponds to fixing $\lambda_{i,y} = 1/d$ for all $i, y$ in Eq. (29), instead of taking the supremum. Therefore, the primal and dual SDPs for the offset $b_E(\vec{w})$ are

$$\boxed{\begin{array}{l} b_E(\vec{w}) = \sup \ \ \frac{1}{d} \mathrm{tr}[Y] \\ \text{with a } d \times d\text{-matrix variable } Y \text{ subject to} \\ \quad Y \quad \leq \sum_{i,y} w_i c_i(x_i, y) A_i(y) \quad \forall x_1, \ldots, x_n. \end{array}} \tag{34}$$

and

$$\boxed{\begin{array}{l} b_E(\vec{w}) = \inf \ \ \frac{1}{d} \sum_i \sum_y \sum_{x_1,\ldots,x_n} w_i \, \mathrm{tr}\big[R(x_1, \ldots, x_n) A_i(y)\big] c_i(x_i, y) \\ \text{with } d \times d\text{-matrix variables } R(x_1, \ldots, x_n) \text{ subject to} \\ \begin{array}{rl} R(x_1, \ldots, x_n) & \geq 0 \quad \forall x_1, \ldots, x_n \\ \sum_{x_1,\ldots,x_n} R(x_1, \ldots, x_n) & = \mathbb{1}. \end{array} \end{array}} \tag{35}$$

The computation of a corresponding boundary point $\vec{\varepsilon} \in \mathcal{U}_E$ is similar as above.

## Acknowledgements

We thank Oliver Sachse and Kais Abdelkhalek for valuable discussions. The authors acknowledge financial support from the BMBF project Q.com-Q, the DFG project WE1240/20 and the ERC grant DQSIM.

# CHAPTER 3

# Preparation uncertainty relations based on cost functions

In the previous chapter we used transport theory, i.e. Wasserstein distances, in order to construct error measures from general cost functions, this enabled us to formulate measurement uncertainty relations. A meaningful aspect of this formulation is that the error between POVMs inherits the unit and the scaling behaviour from the underlying cost function. In this chapter we will introduce deviation measures based on cost function. With those measures in hand we then formulate preparation uncertainty relations, which again come with the unit and the scaling behaviour of the underlying cost function. Since both notions of uncertainty now can be formulated with this property, it is operationally meaningful to compare them. This comparison is placed in the second part of this chapter. We will do this by a basic theorem, which states that

$$\textbf{preparation uncertainty} \prec \textbf{measurement uncertainty}$$

whenever we consider linear uncertainty relations and sharp observables. At the end of this chapter we will comment on cases where the lower bounds for both notions of uncertainty coincide and provide examples of unsharp measurements where the above fails.

## 3.1 Preparation uncertainty for cost functions

For the formulation of preparation uncertainty relations with respect to a cost function a deviation measure on the level of probability distributions is needed. In classical statistics, the standard example of such a measure is the variance. For a real valued random variable $X$, it is typically introduced as

$$\Delta^2 X = \langle |X - \langle X \rangle|^2 \rangle \ . \tag{3.1}$$

However, (3.1) can be formulated alternatively by the variational expression

$$\Delta^2 X = \min_{x_0 \in \mathbb{R}} \langle |X - x_0|^2 \rangle \ . \tag{3.2}$$

It is straight forward to check that the minimum in (3.2) will be attained on $x_0 = \langle X \rangle$, such that both definitions, (3.1) and (3.2), coincide. From a more general point of view, the squared euclidean distance, in (3.2), can be interpreted as cost function which is employed to measure the expected deviation of $X$ from a constant estimate $x_0$.

For general cost functions, this concept is generalized by the following definition:

**Definition 3.1.** Let $X$ be a random variable with outcomes on the set $\Omega_X$, which is distributed by the probability distribution $p : \Omega_X \mapsto [0,1]$. Furthermore, let $c(x,y)$ be a cost function on a pair of outcome sets $(\Omega_X, \Omega_Y)$. For $X$ we define the deviation w.r.t. to $c$ as the minimal expected transport cost from $X$ to a constant estimate from $\Omega_Y$

$$\nu(X) := \min_{y \in \Omega_Y} \langle c(X,y) \rangle = \min_{y \in \Omega_Y} \sum_x p(x) c(x,y). \tag{3.3}$$

**Definition 3.2.** For a quantum state $\rho$ and a measurement $A$, with POVM elements $\{A(x)\}$, we define the deviation with respect to $c$ on the level of outcome distributions:

$$\nu(A|\rho) := \min_{y \in \Omega_Y} \operatorname{tr} \left( \rho \sum_x A(x) c(x,y) \right) .$$

In regard of the above definition and its operational interpretation as transport cost to an optimal constant estimate it is reasonable to consider probabilistic estimators as well. The following corollary shows that the above definition does not change under this more general consideration.

**Corollary 3.3.** *Let $\mathcal{Y}$ denote the set of all random variables that are independent of $X$ and have values in $\Omega_Y$. The minimal expected transport cost between $X$ and $Y \in \mathcal{Y}$ is realized by a point measure, i.e. we have*

$$\inf_{Y \in \mathcal{Y}} \langle c(X,Y) \rangle = \nu(X).$$

*Proof.* This corollary immediately follows from the convex structure of the underlying minimization problem. Let $p$ be the distribution of $X$ and let $q$ denote the distribution of a $Y \in \mathcal{Y}$. Then we know from independence that the joint distribution of $X$ and $Y$ can be written as $p(x)q(y)$ for any $(x, y) \in \Omega_X \times \Omega_Y$. Within this notation we have

$$\inf_{Y \in \mathcal{Y}} \langle c(X, Y) \rangle = \min_q \sum_{x,y} p(x)q(y)c(x, y) \; .$$

We substitute $k_X(y) = \sum_x p(x)c(x, y)$ and get

$$\begin{aligned}\min_q \sum_y q(y)k_X(y) &\geq \min_q \sum_y q(y) \min_{y_0 \in \Omega_y} k(y_0) \\ &= \min_{y_0 \in \Omega_Y} k(y_0) = \min_{y_0 \in \Omega_Y} \sum_x p(x)c(x, y) = \nu(X)\end{aligned}$$

□

For all cost functions, $\nu(A|\rho)$ is a concave functional on the set of quantum states, i.e., for $\lambda \in [0, 1]$ and states $\rho$ and $\sigma$, we have

$$\nu(A|\lambda\rho + (1 - \lambda)\sigma) \geq \lambda\, \nu(A|\rho) + (1 - \lambda)\, \nu(A|\sigma).$$

This essential property directly follows from the concavity of the minimum and the linearity of taking the expectation value with respect to a quantum state.

However, if $\nu(A|\rho)$ is used as a measure of uncertainty, it is often meaningful to demand the cost function to be a premetric. In this case the axioms of a premetric can be directly translated into the following properties that make $\nu(A, \rho)$ a proper measure of uncertainty. If $c(x, y)$ is a premetric we have:

1. $c(x, y) \geq 0 \qquad \Rightarrow \qquad \nu(A|\rho) \geq 0 \; \forall \rho$
2. $c(x, y) = 0 \Leftrightarrow x = y \qquad \Rightarrow \qquad \nu(A|\phi_x) = 0$ iff $\phi_x$ is an eigenstate of $A$.

We note that, for the above to be well defined, we have to assume that $c$ is a function on $\Omega_X \times \Omega_X$, i.e. we compare two random variables on the same set of outcomes. On one hand, we can motivate this assumption by the purpose of quantifying the spread of a probability distribution on its outcome set. On the other, we need this assumption in our current definition of measurement uncertainty, more precisely in the definition of a joint measurement with approximative marginals. However, in this thesis we will also encounter cases where we compare distributions on different outcome sets $\Omega_X \neq \Omega_Y$. This will be considered in chapter Ch. 5 in the context of entropic quantities.

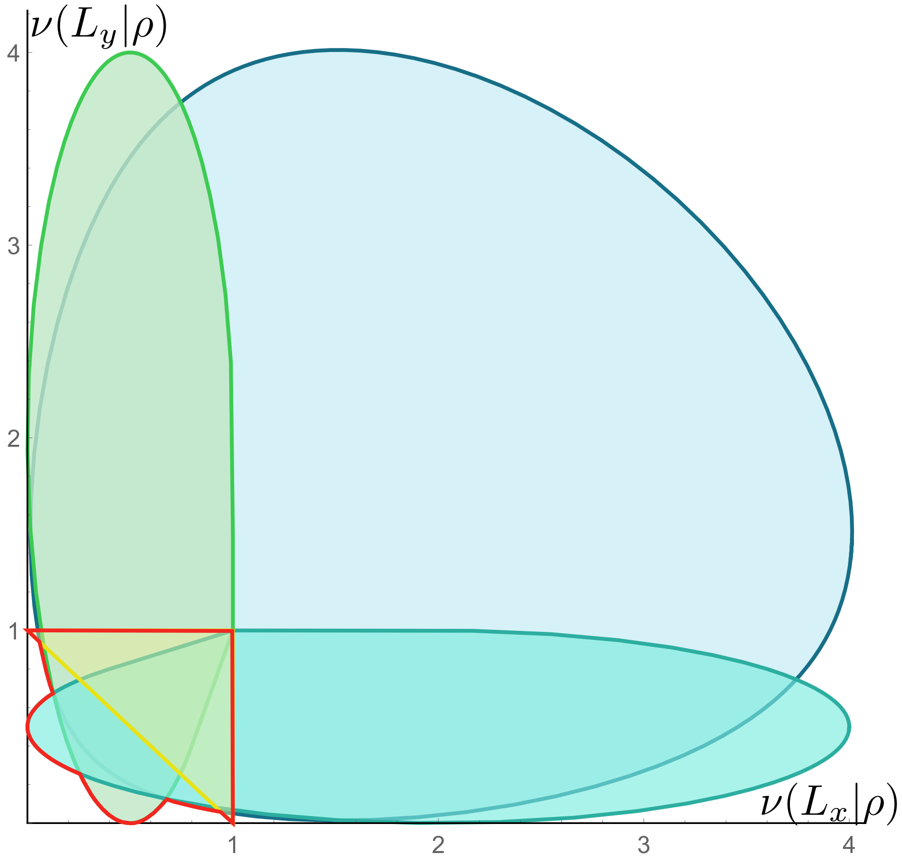


Figure 3.1: The uncertainty region for spin-1 angular momentum measurements $L_x$ and $L_y$, with cost function $c(x,y) = |x-y|^2$. All attainable points are placed in the joint numerical range of $F(x,y)$ for some $(x,y) \in \{-1,0,1\}^2$. We have $F(0,1), F(0,-1) =$ *green*, $F(0,0) =$ *yellow*, $F(-1,1), F(1,-1), F(1,1), F(-1,-1) =$ *blue*, and $F(1,0), F(-1,0) =$ *cyan*. The boundary of the uncertainty region $\mathcal{U}_\nu$ is marked in red.

**Computing uncertainty diagrams:** Computing uncertainty relations for observables with respect to continuous outcome sets is a presumably very hard problem. The special case of variances is considered in the next chapter, but beside this, a general method for the continuous case is outstanding. In contrast, measurements with finite outcome sets can be handled straight forwardly. Here, we can compute uncertainty relations as follows:

Let $A_1 \dots A_n$ be a collection of measurements, with outcome sets $\Omega_1, \dots, \Omega_n$. To each $y_i \in \Omega_i$ we associate an operator

$$K_{A_i}(y_i) := \sum_{x \in \Omega_i} A_i(x) c(x, y_i)$$

and to each tuple of outcomes $(y_1, \dots, y_n) \in \Omega_1 \times \dots \times \Omega_N$ we associate the joint numerical range

$$F(y_1, \dots, y_n) := \Big\{ \big(\mathrm{tr}(\rho K_{A_1}(y_1)), \dots, \mathrm{tr}(\rho K_{A_n}(y_n))\big) \Big| \rho \text{ is a quantum state} \Big\}.$$

In terms of the $K_{A_i}(y_i)$ , the uncertainty region can be written as

$$\begin{aligned} \mathcal{U}(A_1 \dots A_n) &= \big\{ (\nu(A_1|\rho), \dots, \nu(A_n|\rho)) \,\big|\rho \text{ is a quantum state} \big\} \\ &= \Big\{ \min_{y_1 \cdots y_n} \big(\mathrm{tr}(\rho K_{A_1}(y_1)), \dots, \mathrm{tr}(\rho K_{A_n}(y_n))\big) \Big| \rho \text{ is a quantum state} \Big\} \quad (3.4) \end{aligned}$$

Dropping the minimization in (3.4) will enlarge the set. Hence, we have

$$\mathcal{U}(A_1 \dots A_n) \subseteq \bigcup_{y_1, \dots, y_n} F(y_1, \dots, y_n).$$

However, by dropping the minimization in (3.4), only non-optimal points will be added to $\mathcal{U}(A_1 \dots A_n)$. Therefore, we have

$$\begin{aligned} \mathsf{PDR}\big[\mathcal{U}_P(A_1 \dots A_n)\big] &= \mathsf{PDR}\Big[ \bigcup_{y_1, \dots, y_n} F(y_1, \dots, y_n)\Big] \\ &= \bigcup_{y_1, \dots, y_n} \mathsf{PDR}\Big[F(y_1, \dots, y_n)\Big], \end{aligned}$$

which is a union of finitely many sets, if all $\Omega_i$ are finite. Here, we can compute the Pareto boundary by, firstly, computing the Pareto boundary for any set $F(y_1, \dots, y_n)$ individually and selecting the optimal points from these boundaries afterwards. Because $F(y_1, \dots, y_n)$ is a joint numerical range, its boundary can be computed efficiently. Furthermore, if we directly want to compute the minimal uncertainty $u_j$ of a $A_j$ measurement, under the assumption that the uncertainties of all other measurements are not bigger than some $u_i$, for $i \neq j$, we have to compute

$$\begin{aligned} u_j &= \min \left\{ \nu(A_j|\rho) \; |\rho : \nu(A_i|\rho) \geq u_i \forall i \neq j \right\} \\ &= \min \left\{ \mathrm{tr}\left(\rho K_{A_j}(x_j)\right) \; |x_j \in \Omega_j \,\&\, \forall x_i \in \Omega_i : \mathrm{tr}\left(\rho K_{A_i}(x_i)\right) \geq u_i \right\} \\ &= \min_{x_j \in \Omega_j} \min \left\{ \mathrm{tr}\left(\rho K_{A_j}(x_j)\right) |\forall x_i \in \Omega_i : \mathrm{tr}\left(\rho K_{A_i}(x_i)\right) \geq u_i \right\}, \end{aligned}$$

which can easily be done by solving the above SDP for any $x_j \in \Omega_j$ individually, and then taking the minimum over all $x_j$ afterwards.

**Examples:** As an example, the sets $F(x,y)$ are computed for two cases: The first example are spin-1 measurements with $c(x,y) = |x-y|^2$ (see Fig. 3.1). Here, the actual uncertainty region (red boundary) is depicted as well. In this figure, we can immediately see that the actual uncertainty region is way smaller than the union of all the $F(x,y)$, however the bigger set only contains additional points with high uncertainty. We will come back to this example in Fig. 3.3 and Fig. 4.1.

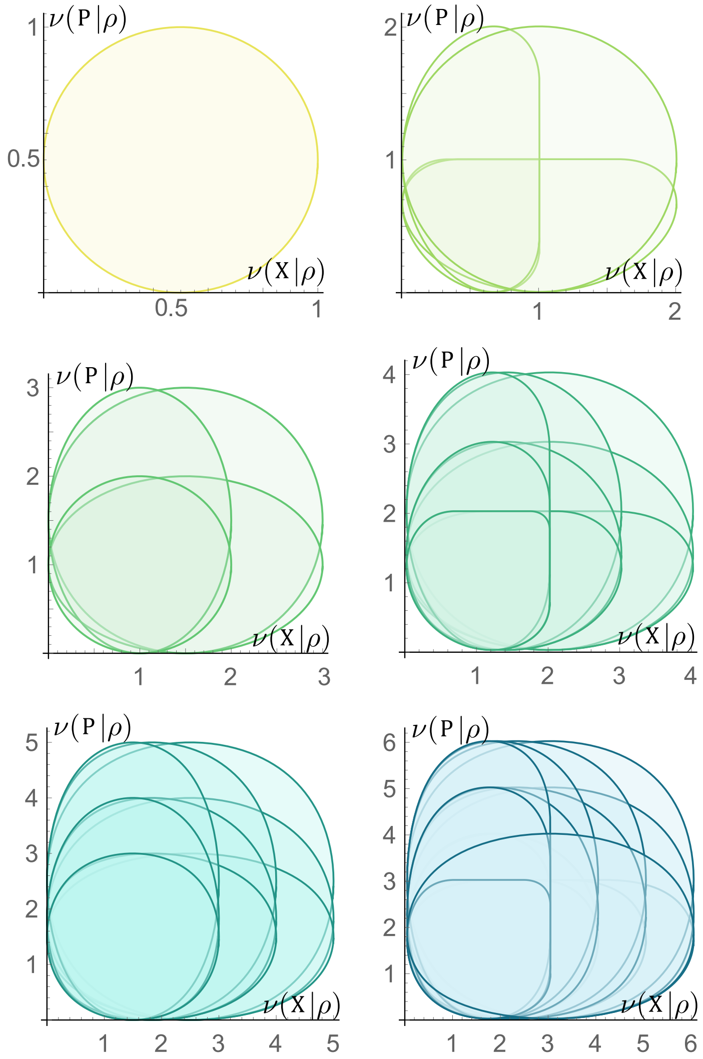


Figure 3.2: Preparation uncertainty regions for phasespace observables $(\mathtt{X},\mathtt{P})$ with dimensions $d = 2,\ldots,7$ (ordered from top left to bottom right). We have $c(x,y) = |x-y|$ and $\mathtt{X},\mathtt{P}$ projective and linked by the $d$-dimensional Fourier transformation. The outcome sets are $\Omega_{\mathtt{X}} = \Omega_{\mathtt{P}} = \{1,\ldots,d\}$. In this case all uncertainty regions are convex.

As a second example, (see Fig. 3.2), we depict the sets $F(x, y)$ for discrete phase space observables $\mathtt{X}$ $\mathtt{P}$ (see also [Wer16] for a comprehensive treatment of uncertainty relations in this case). Phase space observables are constructed by considering two sharp measurements with outcomes on $\{1, \ldots, d\}$, and eigenbases related by the $\mathbb{Z}_n$ Fourier-transformation. Hereby, each set of outcomes is treated as a set of labels corresponding to a representation of $\mathbb{Z}_n$.

## 3.2 Connections between preparation uncertainty and measurement uncertainty

In this section we will draw a connection between linear measurement and preparation uncertainty relations in the context of general cost functions. For the moment we will consider only sharp measurement and comment on unsharp measurements later.

For a collection of projective measurements, we can make the following observation: Preparation uncertainty relations within a collection of measurements vanish, whenever the observables share at least *one* common eigenvector, i.e the measurement operators of all observables commute on some subspace. In contrast to this, it can be shown that (see [BLPY16]), a joint measurement between sharp measurements only exists, i.e. we have a vanishing measurement uncertainty, if the observables commute on the *whole* Hilbert space. Therefore, the existence of a measurement uncertainty relation is more restrictive than the existence of a preparation uncertainty relation. Hence, we should ask for ways to lower bound measurement uncertainty given a preparation uncertainty relation.

We will answer this question by the following theorem (Thm. 3.4), which states that, for a fixed cost function, any linear uncertainty relation for the entanglement reference frame error can be lower bounded by a corresponding linear preparation uncertainty:

**Theorem 3.4.** *Let $A = (A_1 \ldots A_n)$ be a collection of sharp measurements, let $\rho$ be a quantum state, and let $\mathcal{R}(A_1, \ldots A_n)$ be the set of all joint measurements $R$ that approximate $A_1 \ldots A_n$ by marginals $A'_1 \ldots A'_n$. For the entanglement reference frame error, given by the vector*

$$\vec{\varepsilon}_E(A|A') := (\varepsilon_E(A_1|A'_1), \ldots, \varepsilon_E(A_n|A'_n)),$$

*and the preparation uncertainty, given by the vector*

$$\vec{\nu}(A|\rho) := (\nu(A_1|\rho), \ldots, \nu(A_n|\rho)),$$

*and for all positive weights $\vec{a} = (a_1 \dots a_n)$ we have:*

$$\inf_R \vec{a} \cdot \vec{\varepsilon}_E(A|A') \geq \inf_\rho \vec{a} \cdot \vec{\nu}(\rho|A).$$

*Proof.* Let $\Omega_i$ denote the outcome set of the measurement $A_i$. For shorthand notation we will always assume that the variables $x_i$ and $y_i$ come from $\Omega_i$ and that any summation over $x_i$ or $y_i$ runs over the $\Omega_i$. Since the measurements $A_i$ are sharp, we will assume w.l.o.g. that all $\Omega_i$ have the same size $d$, i.e. the dimension of the underlying Hilbert space. As usual, in the following, the joint measurement $R$ is given by a POVM on the joint outcome set $\Omega_1 \times \cdots \times \Omega_n$ with effect operators $R(y_1, \dots, y_n)$, which obey

$$\sum_{y_1,\dots,y_n} R(y_1, \dots, y_n) = \mathbb{I} \text{ and } R(y_1, \dots, y_n) \geq 0.$$

At first step, we note that the preparation uncertainty bound

$$c_{prep} := \inf_\rho \vec{a} \cdot \vec{\nu}(\rho|A) = \inf_\rho \operatorname{tr}\left(\rho \sum_i a_i \min_{y_i} \sum_{x_i} c(x_i, y_i) A_i(x_i)\right)$$

gives the best constant $c_{prep}$ such that,

$$\sum_i a_i \sum_{x_i} c(x_i, y_i) A_i(x_i) \geq c_{prep}\, \mathbb{I} \tag{3.5}$$

holds for all $y_1, \dots, y_n$. We will need this form in a moment for lower bounding the measurement error.

Secondly, we recall that for any collection of self-adjoint operators $T(z)$ and a minimization over POVMs $Q$ the duality relation

$$\begin{aligned} \inf &\left\{ \sum_z T(z)Q(z) \middle| \sum_z Q(z) = \mathbb{I}, Q(z) \geq 0 \right\} \\ &\geq \sup \left\{ \operatorname{tr}(Y) | \forall z : T(z) \succeq Y \right\} \end{aligned} \tag{3.6}$$

holds. This can be checked by bounding each $T(z)$ from below by a feasible $Y$, which gives:

$$\operatorname{tr}\left(\sum_z T(z)Q(z)\right) \geq \operatorname{tr}\left(\sum_z YQ(z)\right) = \operatorname{tr}(Y).$$

Note that in generic cases we even have equality in (3.6), i.e. a strong duality holds. However, for the purpose of this proof, we only need weak duality in (3.6).

Using the definition of the entanglement reference frame error, we expand:

$$\begin{aligned}\inf_R \vec{a}\cdot\vec{\varepsilon}_E(A|A') &= \inf_R \sum_{i=1}^N a_i \varepsilon_E(A_i|A_i') \\ &= \inf_R \sum_i a_i \frac{1}{d} \sum_{y_1,\ldots,y_N,x_i} \mathrm{tr}\left(A_i(x_i)R(y_1,\ldots,y_n)\right) c(x_i,y_i) \\ &= \inf_R \mathrm{tr}\left(\sum_{y_1,\ldots,y_N}\Big[\sum_i a_i \frac{1}{d}\sum_{x_i} A_i(x_i)c(x_i,y_i)\Big]R(y_1,\ldots,y_n)\right),\end{aligned}$$

where we can use the duality from (3.6) and get

$$\inf_R \vec{a}\cdot\vec{\varepsilon}_E(A|A') \geq \sup\left\{ tr(Y)\middle| Y \preceq \frac{1}{d}\sum_i a_i \sum_{x_i} c(x_i,y_i)A_i(x_i)\right\}. \tag{3.7}$$

Here we can use (3.5) to check that $Y = \frac{c_{prep}}{d}\mathbb{I}$ is a feasible point in optimization on the r.h.s. of (3.7). Hence, we get the desired statement

$$\inf_R \vec{a}\cdot\vec{\varepsilon}_E(A|A') \geq c_{prep}\,\mathrm{tr}(\mathbb{I}/d) = c_{prep}. \tag{3.8}$$

□

On the first view, the above theorem considers only the entanglement reference error. However, it follows directly from the definitions of the other error quantities (see [SRW16] eq. (11) ) that the entanglement reference error is smaller than the calibration error, which again is smaller than the measurement error. This ordering does not change when we consider positive weighted linear combinations of these errors. Hence we have

$$\inf_R \vec{a}\cdot\vec{\varepsilon}_M(A|A') \geq \inf_R \vec{a}\cdot\vec{\varepsilon}_C(A|A') \geq \inf_R \vec{a}\cdot\vec{\varepsilon}_E(A|A') \geq \inf_\rho \vec{a}\cdot\vec{\nu}(\rho|A).$$

In terms of positive convex hulls this implies that the respective sets include each other, i.e. we have

$$\mathcal{U}_M^+ \subseteq \mathcal{U}_C^+ \subseteq \mathcal{U}_E^+ \subseteq \mathcal{U}_\nu^+.$$

For the example of spin-1 measurements in orthogonal directions, this is depicted in Fig. 3.3. In Sec. 2.1 it is shown that the uncertainty regions, for a measurement uncertainty, are always convex. That directly implies, that the Pareto boundary of $\mathcal{U}_L^+$ always equals the Pareto Boundary of $\mathcal{U}_L$, for $L = M, C, E$. However, this is not always true for a preparation uncertainty region, which already could be seen in Fig. 3.1 and Fig. 3.2. Hence, it could happen that

$$\mathrm{PDR}(\mathcal{U}_\varepsilon) \nsubseteq \mathrm{PDR}(\mathcal{U}_\nu).$$

An example for this is provided in Fig. 3.3, here the red line, corresponding to the the Pareto boundary of the preparation uncertainty region, crosses the boundaries of the measurement uncertainty regions of the calibration and the entanglement reference error.

**Equality:** The crucial estimate in the proof of Thm. 3.4 happens between (3.7) and (3.8). Here, we estimate an expression of the form

$$\sup\left\{\mathrm{tr}(Y) | T(z) \succeq Y \,\forall z\right\} \tag{3.9}$$

from below by the trace of the best feasible $Y \propto \mathbb{I}$. In general this estimate does not need to be tight, i.e. the linear preparations uncertainty relation can be smaller than the corresponding linear measurement uncertainty relation, see e.g. Fig. 3.3. However, there are relevant cases of symmetries where equality holds:

Let $\mathcal{G}$ be a group with an irreducible representation $\{U_g\}_{g\in\mathcal{G}}$ and let the parameters $z$ be elements of $\mathcal{G}$. We have equality in the above estimate if the operators $T(z)$ are covariant with respect to the action of $U_g$, i.e. if we have that

$$U_g T(z) U_g^\dagger = T(z+g)$$

holds for all $z, g \in \mathcal{G}$. In this case we have: If $Y$ is feasible, i.e. $Y \preceq T(z) \forall z \in \mathcal{G}$, any rotation $U_g Y U_g^\dagger$ is feasible, too. Hence, we can conclude

$$\frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} U_g Y U_g^\dagger \preceq T(z),$$

the feasibility of group mean of $Y$. Because of

$$\mathrm{tr}\left(\frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} U_g Y U_g^\dagger\right) = \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} \mathrm{tr}\left(U_g Y U_g^\dagger\right) = \mathrm{tr}(Y),$$

taking the group mean does not change the trace of $Y$, i.e. the objective function in our optimization (3.9). Therefore, we can conclude that for any optimizer $Y^*$, its group mean $\frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} U_g Y^* U_g^\dagger$ is an optimizer, too. However, we have

$$\frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} U_g Y^* U_g^\dagger \propto \mathbb{I},$$

if the representation $\{U_g\}_{g\in\mathcal{G}}$ is irreducible, and, therefore, equality in the central estimate of Thm. 3.4.

An example of this symmetry are finite phase-space observables $\mathtt{X}$ and $\mathtt{P}$, together with a cost function that respects the underlying translation invariance, see the example for Fig. 3.2 and [Wer16]. Here, our measurements are given by POVMs with elements $\{X(q)\}$ and $\{P(p)\}$, and outcome tuples $z = (p, q) \in \{1, \dots d\}^2 \cong \mathbb{Z}_d \times \mathbb{Z}_d$, from which we get the discrete position and momentum operators

$$\mathtt{X} = \sum X(q)q \text{ and } \mathtt{P} = \sum P(p)p.$$

The group $\mathcal{G} = \mathbb{Z}_d \times \mathbb{Z}_d$ is represented by the Weyl operators $W(p, q)$, i.e. by

$$W(p, q) = e^{-ip\mathtt{X}+iq\mathtt{P}}, \tag{3.10}$$

which acts on the POVM elements $X(q')$ and $P(p')$ by

$$W(p, q)X(q')W(p, q)^\dagger = X(q' + q) \tag{3.11}$$

and

$$W(p, q)P(p')W(p, q)^\dagger = P(p' + p).$$

For weights $\lambda, (1 - \lambda)$, the corresponding operators $T$ are given by

$$T(p, q) = \frac{1}{d} \sum_{(p',q')\in\{1,\dots d\}^2} \Big(\lambda\, c(q', q)X(q') + (1 - \lambda)\, c(p', p)P(p')\Big),$$

with a cost function that obeys $c(x + d, y) = c(x, y + d) = c(x, y)$. Here, we can straight forwardly check the necessary covariance condition

$$W(p, q)T(p', q')W(p, q) = T(p + p', q + q')$$

by using the transformation rules (3.10) and (3.11). Therefore, we have the equality

$$\inf_R \lambda\varepsilon_E(\mathtt{X}|\mathtt{X}') + (1 - \lambda)\varepsilon_E(\mathtt{P}|\mathtt{P}') = \inf_\rho \lambda\nu(\mathtt{X}|\rho) + (1 - \lambda)\nu(\mathtt{P}|\rho)$$

Because of the high symmetry, we can make an even stronger statement for this particular example: Firstly, it can be checked easily that all test states attain the same individual error. Hence, the calibration error and the entanglement reference frame error coincide. Secondly, see [BLW14b] for a proof, the worst error is attained on test-states. Hence, the measurement error coincides with the other two errors, as well. Therefore, Fig. 3.2 actually shows the boundary of all four uncertainty regions $\mathcal{U}_M$,$\mathcal{U}_C$,$\mathcal{U}_E$, and $\mathcal{U}_\nu$.

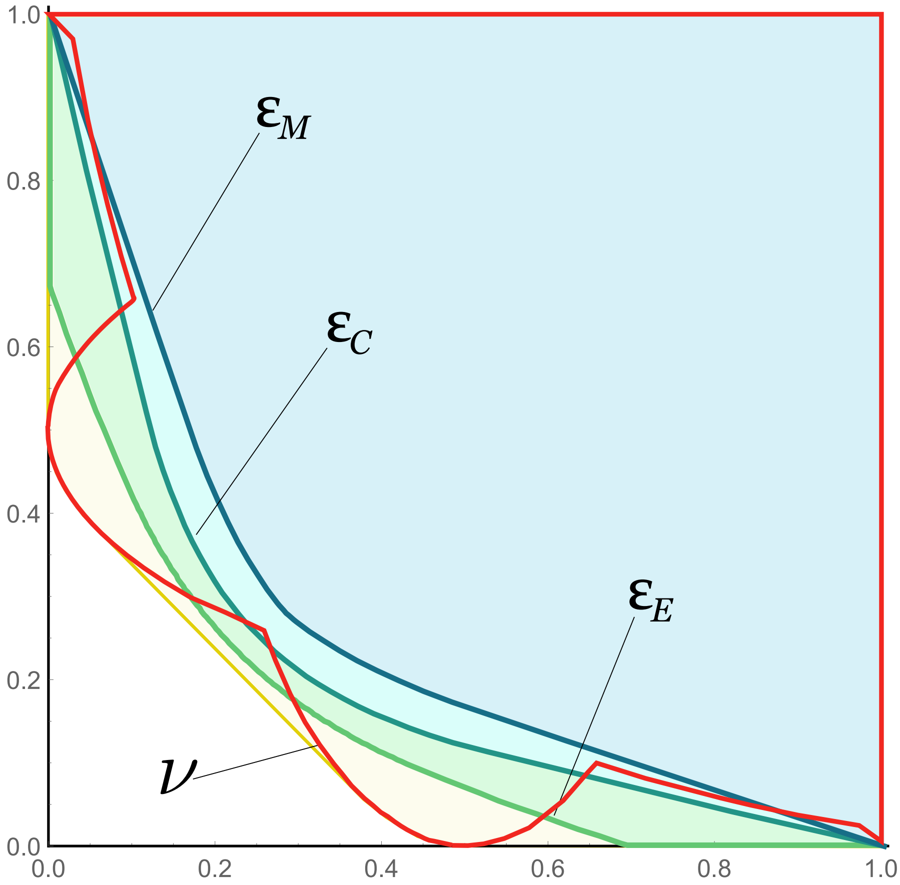


Figure 3.3: Preparation and measurement uncertainty regions of spin-1 angular momentum measurements $L_x$ and $L_y$ in orthogonal directions, with cost function $c(x, y) = |x - y|^2$. The red line indicates the shape of the preparation uncertainty region, and the yellow line its convex hull. This example shows that only the convex hull is a lower bound on the measurement uncertainty regions, indicated in green and blue.

**Nøisy measurements:** In Thm. 3.4 we only considered sharp measurements. The calibration and the entanglement reference frame error are only defined for this case. Albeit, for the general case of unsharp measurements, the measurement error $\varepsilon_M$ and the deviation $\nu$ are well defined, a statement analogous to Thm. 3.4 may fail: Consider the measurements from the above example modified to

$$\tilde{X}(q) := (1 - \delta)X(q) + \delta\mathbb{I}$$

and

$$\tilde{P}(p) := (1 - \delta)P(p) + \delta\mathbb{I},$$

with a noise parameter $\delta \in (0,1)$. For $\delta = 0$ we get equality between the measurement and the preparation uncertainty. However, if we increase the noise all preparation uncertainties will increase, too. In the limit $\delta = 1$ we will apply measurements which will give us a uniform distribution on all measurement outcomes, independent of the input state. In contrast, the measurement error will decrease when we increase the noise. Furthermore, for a sufficiently big $\delta$, the measurements $\tilde{\mathtt{X}}$ and $\tilde{\mathtt{P}}$ will become compatible. Hence, the measurement error will vanish completely.

# CHAPTER 4

## Uncertainty based on variances

In this chapter we will have a closer look at preparation uncertainty relations in terms of variances. In accordance to the chapters before, we will keep the restriction of considering only observables on finite Hilbert spaces and outcome sets. However, for computing the minimal deviation to a point measure (see(3.2)) we will allow to embed outcome sets into the real numbers and we will fix the euclidean distance as cost function. By this extension the moments of a probability distribution, especially the expectation value, get an explicit meaning. As mentioned before, this makes sense for quantities that have a discrete outcome set in quantum physics but a continuous outcome set in classical physics, like for example: energies and angular momenta. For a fixed cost function, we are now also in a position to compare $\Delta_\rho^2 A$, the variance of a measurement $A$, with the corresponding discrete deviation $\nu(A,\rho)$. Because the variance is computed as minimum over the bigger set (compare (3.2) and (3.3)), we have

$$\Delta_\rho^2 A \leq \nu(A,\rho),$$

which also carries over to an inclusion of the corresponding uncertainty regions, see Fig. 4.1. Furthermore, this shows that a state-independent variance based uncertainty relation gives a lower bound on the corresponding measurement uncertaintiy relation. This was already used, with equality, for the special case of position and momentum in [BLW14b, BLW13]. Unfortunately, the computational methods developed in chapter Ch. 3, for finite outcome sets, are inapplicable for

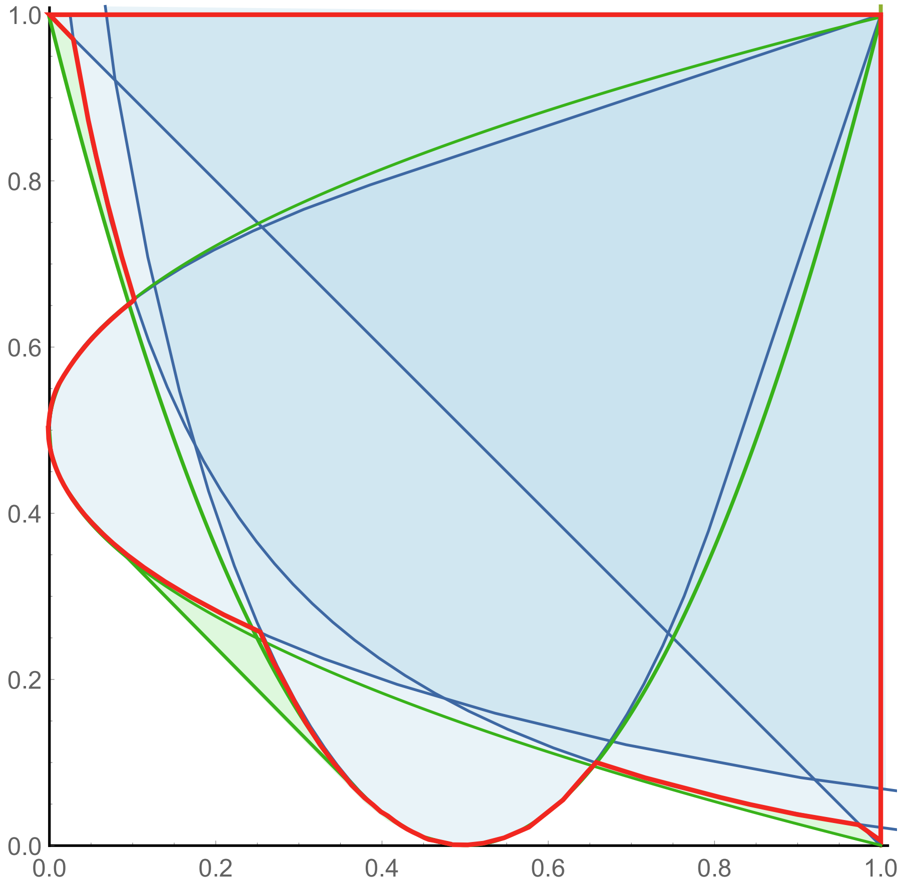


Figure 4.1: Comparison of the uncertainty regions for variances $(\Delta^2_\rho L_x, \Delta^2_\rho L_y)$ and the discrete uncertainty (with red boundary) $(\nu(L_x, \rho), \nu(L_y, \rho))$ for measurements of angular momentum components in spin-1 representation. The minimization, in the definition of the variance, is evaluated on the bigger set. Hence, the uncertainty region of the discrete uncertainty is included in the uncertainty region of variances.

variances, because in this case a minimization over the uncountable set of the real numbers has to be performed.

In general it is worth to concentrate on variances, because they are doubtlessly the most common measure of deviation used at many points in classical statistics and even though in quantum uncertainty, most prominently in the works of Kennard [Ken27], Robertson [Rob29], and Schrödinger [Sch35]. In reference to those works, the vast majority of publications on quantum uncertainty, published in the last decade, (see sources in the introduction), has variances as measure of choice, as well. Thereby, most of them do not, or only insufficiently, address the question of state-independent optimality.

The central part of this chapter is [SDW17], given in the next section. Here we addressed this question for arbitrary measurements. More precisely: we provide a simple and efficient algorithm that allows to compute linear uncertainty relations for any pre assigned precision, with estimates approaching the optimal bound from below.

## 4.1 [SDW17]

*State-independent Uncertainty Relations and Entanglement detection in noisy Systems*

- **Authors:** René Schwonnek, Lars Dammeier, and Reinhard F. Werner
- **Published in:** Physical Review Letters 119, 170404 (2017)
- **DOI:** 10.1103/PhysRevLett.119.170404
- **Presented version:** The presented version is identical to arXiv:1705.10679, the literature and the supplemental material are placed at the end of this thesis.
- **Contributions:** The central ideas and the implementation of the algorithm were contributed by René Schwonnek.
- **Main results:**
  - Linear variance based uncertainty relations can be represented as optimization problem over a three-dimensional joint numerical range.
  - An algorithm for solving this optimization is presented. This algorithm produces lower bounds, and an error estimate.
  - Linear uncertainty relations can also be computed for general POVMs. An improved entanglement detection scheme, based on the tomography of local noise sources, is provided.

# State-independent uncertainty relations and entanglement detection in noisy systems

René Schwonnek,* Lars Dammeier,† and Reinhard F. Werner‡
*Leibniz Universität Hannover - Institut für Theoretische Physik*

(Dated: April 7, 2018)

Quantifying quantum mechanical uncertainty is vital for the increasing number of experiments that reach the uncertainty limited regime. We present a method for computing tight variance uncertainty relations, i.e., the optimal state-independent lower bound for the sum of the variances for any set of two or more measurements. The bounds come with a guaranteed error estimate, so results of pre-assigned accuracy can be obtained straightforwardly. Our method also works for POVM measurements. Therefore, it can be used for detecting entanglement in noisy environments, even in cases where conventional spin squeezing criteria fail because of detector noise.

PACS numbers: 03.65.Ta, 02.60.Pn, 03.67.Mn

## INTRODUCTION

Uncertainty relations quantitatively express a phenomenon which is ubiquitous in quantum mechanics: Given two observables $A$ and $B$, it is usually impossible to prepare a state such that the respective outcome distributions of these observables are both sharp. Of course, for the best known example of this, the position and momentum observables, the relation is in every textbook. It was first established by Kennard [1], who turned Heisenberg's heuristic ideas [2] into a quantitative statement. In particular, it was his idea to consider the variances [3] of momentum and position in the state $\rho$ as the mathematical expression of sharpness. Kennard's relation $\Delta^2_\rho(P)\,\Delta^2_\rho(Q) \geq \hbar^2/4$ is tight, i.e., the constant on the right hand side is the best possible, because it is attained for Gaussian pure states.

The aim of our paper is to provide an efficient method to obtain the best possible bounds for any given pair of measurements $A$, $B$. This is of direct use in the increasing number of experiments that reach the uncertainty-limited regime. A particular application is the certification of entanglement via steering inequalities [4–6]. In such applications, even if one does not necessarily need an optimal bound, it is crucial to have a correct one, i.e., a bound valid for *all* states. Any algorithm based on computing the uncertainties "for sufficiently many states" will fail to guarantee this correctness. In particular, in high dimensional Hilbert spaces, typical states will not have uncertainties near the boundary, so it is actually hard to explore the set of uncertainty pairs $(\Delta^2_\rho(A), \Delta^2_\rho(B))$ "from within". Our method uses instead an "outer" approximation, which has the virtue that in every step it provides a correct bound. The bound is iteratively improved, converging to the optimal one. This feature sets our method apart from several recent works, in which ad hoc methods were used to provide uncertainty bounds. The problem of getting optimal uncertainty bounds becomes more difficult as the dimension $d$ of the Hilbert space increases. Indeed, naively it would seem to be a search problem on the $2d-2$ dimensional manifold of pure states, which in bad cases might scale exponentially with $d$. However, we can do much better. We reformulate the problem as a geometric problem in three dimensions, namely of getting a sequence of outer polyhedral approximation of a certain convex set, see Fig. 1. Any such approximation gives a valid uncertainty bound. In the iteration step, i.e., for computing a tighter approximation, one has to compute the lowest eigenvalue of a certain hermitian combination of the operators $A$ and $B$. Those eigenvalue problems now determine the scaling of our method as a function of dimension, which will be a low order polynomial in $d$. Moreover, if additional information is available about $A$ and $B$, for example, if they are both sparse in the same basis, eigenvalue computations can be speeded up considerably, and our method will speed up by the same factor.

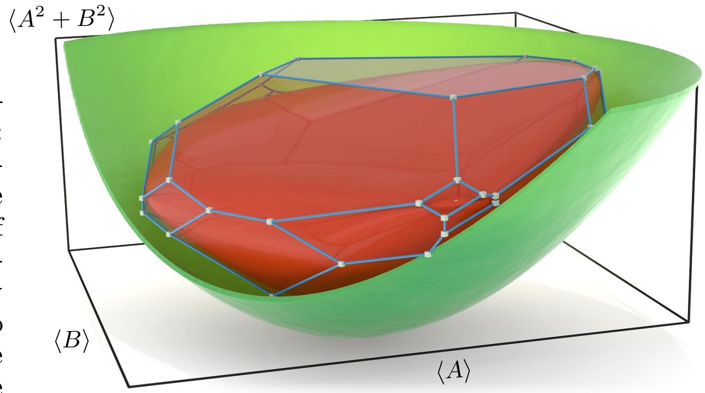


FIG. 1. Minimizing the sum of the variances of two observables $A$ and $B$ can be expressed entirely in terms of the set $\mathcal{C}$ of possible triples $(\langle A\rangle_\rho, \langle B\rangle_\rho, \langle A^2+B^2\rangle_\rho)$ (red solid convex body), namely as finding that vertical displacement of the surface $z = x^2+y^2$ (green paraboloid) which just touches $\mathcal{C}$ from below. We successively approximate $\mathcal{C}$ by polytopes (blue edges, boxed vertices) from the outside, and perform the minimization on this polytope. This gives a converging sequence of correct state-independent uncertainty relations.

Tight uncertainty bounds have only been obtained for a few specific pairs of observables. One example is angular momentum measurement, where bounds for two or three orthogonal spin components [4, 7, 8] are known. In those cases symmetry crucially helps to reduce the problem. Other examples are qubits [9], for which the low dimension allows an analytical solution.

There are also variants, in which the sharpness of a distribution is measured by other quantities than the usual variance [10–13], for instance entropies [14–25] and its generalization to majorazation uncertainty relations [26–28], or where more than two observables are considered simultaneously [29–31]. Quite different methods [32] are needed for optimal measurement uncertainty relations [10], or information-disturbance bounds [10], so we will not consider these aspects here.

## METHODS

### Linear state independent bounds

Since we are interested in state-independent bounds [33] we have no use for the often-cited general relation by Robertson [34] (and its improvements [35]), which have a state-dependent expression like $\langle i[A,B]\rangle_\rho^2$, or similar, on the right hand side. Indeed, any relation of product form $\Delta_\rho^2(A)\,\Delta_\rho^2(B) \geq c$ is useless for state-independent relations in finite dimension: $A$ and $B$ have discrete eigenvalues, so the trivial $c = 0$ is the best possible bound. We therefore consider bounds of the form

$$\Delta_\rho^2(A) + \Delta_\rho^2(B) \geq c. \qquad (1)$$

Here, $c$ is the largest constant for which the above holds on *any* quantum state $\rho$. Since our method handles arbitrary $A$ and $B$ we can also admit factors here, i.e., inequalities of the form $\alpha\Delta_\rho^2(A) + \beta\Delta_\rho^2(B) \geq c(\alpha,\beta)$. Each of these constrains the set of uncertainty pairs $(\Delta_\rho^2(A), \Delta_\rho^2(B))$ to a half-plane, and together they outline the uncertainty set (or, more precisely its "lower convex hull", see Fig. 4 and [7, 9, 32]).

To see the connection to eigenvalue problems we write the optimal constant in (1) as

$$c = \min_\rho \min_{a,b} \left\langle (A - a\mathbb{1})^2 + (B - b\mathbb{1})^2 \right\rangle_\rho . \qquad (2)$$

Here we just wrote the variance as the minimal quadratic deviation, using that the minimum with respect to $a$ is attained at the expectation $a = \langle A\rangle_\rho$. On the other hand, if we fix $a$ and $b$, the minimization with respect to $\rho$ is exactly the ground state problem for the operator in parentheses. This suggested our previous ansatz [7], which we call the *see-saw* algorithm: One alternatingly minimizes with respect to $\rho$ and $(a, b)$. In many practical cases this converges quickly, and with the safeguard of trying out several initial values it seems fairly reliable. However, in general the method of Alternating Minimization may easily fail to find the global minimum, and there is no proof of convergence. Intermediate results of the see-saw algorithm give an upper bound on $c$, but as an upper bound on a lower bound this is useless for applications. Moreover, there are indications that the see-saw algorithm actually may get trapped.

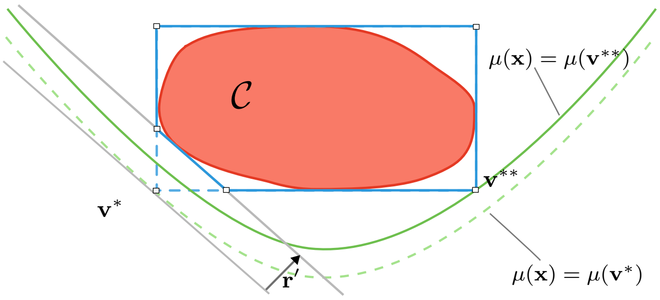

$\mathcal{C}$
$\mu(\mathbf{x}) = \mu(\mathbf{v}^{**})$
$\mathbf{v}^{**}$
$\mathbf{v}^{*}$
$\mathbf{r}'$
$\mu(\mathbf{x}) = \mu(\mathbf{v}^{*})$

FIG. 2. Two dimensional sketch of geometry and the basic algorithm: The set $\mathcal{C}$ (red) with its outer approximation $\mathcal{P}(\mathcal{R})$ (blue and blue dasehd) and the extremal points $\mathcal{E}(\mathcal{R})$ (white squares). By adding the direction $\mathbf{r}'$, the polyhedral approximation is refined and the lower bound $c_-(\mathcal{R})$ is improved from $\mu(\mathbf{v}^*)$ (dashed green parabola) to $\mu(\mathbf{v}^{**})$ (green parabola).

### Geometry of outer approximations

In contrast, the method described in this paper is an outer method, in which all intermediate steps give valid lower and upper bounds on $c$. Its geometric core is the joint numerical range

$$\mathcal{C} = \left\{ \left( \langle A\rangle_\rho , \langle B\rangle_\rho , \langle A^2 + B^2\rangle_\rho \right) \,\middle|\, \rho \in \mathcal{S}(\mathcal{H}) \right\}, \qquad (3)$$

where $\mathcal{S}(\mathcal{H})$ denotes the state space, i.e., the set of density operators. Notice first that this set contains all the information necessary to compute $c$ from (2). With the quadratic functional $\mu(\mathbf{x}) := z - x^2 - y^2$ of $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$ we find

$$c = \min_{\rho\in\mathcal{S}(\mathcal{H})} \Delta_\rho^2(A) + \Delta_\rho^2(B) = \min_{\mathbf{x}\in\mathcal{C}} \mu(\mathbf{x}). \qquad (4)$$

Now the set $\mathcal{C}$ is clearly convex and compact, because the state space $\mathcal{S}(\mathcal{H})$ has these properties, and they are preserved by the map taking $\rho$ to the tuple of expectations. The set $\mathcal{C}$ is therefore completely described by the linear inequalities it satisfies. To get such inequalities, let $\mathbf{r} = (r_1, r_2, r_3)$ be a real vector, and consider $H(\mathbf{r}) = r_1 A + r_2 B + r_3(A^2 + B^2)$. Let $h(\mathbf{r})$ denote the smallest eigenvalue of this operator. Then, for any state $\rho$, and hence the corresponding tuple $\mathbf{x} \in \mathcal{C}$ of expectations:

$$\mathbf{r} \cdot \mathbf{x} = \langle H(\mathbf{r})\rangle_\rho \geq h(\mathbf{r}). \qquad (5)$$

Now let $\mathcal{R} \subset \mathbb{R}^3$ be any finite set of vectors, and consider the polytope $\mathcal{P}(\mathcal{R})$ of those points $\mathbf{x}$, which just satisfy the inequalities (5) with $\mathbf{r} \in \mathcal{R}$. Since these vectors satisfy fewer constraints than $\mathcal{C}$, we have $\mathcal{C} \subset \mathcal{P}(\mathcal{R})$, i.e., this is an outer approximation of $\mathcal{C}$. Denote by $\mathcal{E}(\mathcal{R})$ the set of extreme points of $\mathcal{P}(\mathcal{R})$, which is also finite. Then

$$c \geq \min_{\mathbf{x} \in \mathcal{P}(\mathcal{R})} \mu(\mathbf{x}) = \min_{\mathbf{x} \in \mathcal{E}(\mathcal{R})} \mu(\mathbf{x}) =: c_-(\mathcal{R}). \tag{6}$$

Here we have used, firstly, that the minimum over a larger set is smaller, and, secondly, that the functional $\mu$ is concave, so that the minimum over a compact convex set is attained at an extreme point. Hence for every finite set $\mathcal{R}$ of directions, we get a lower bound on $c$, which is computed as a finite minimum over $\mathcal{E}(\mathcal{R})$. On the other hand, for each $r \in \mathcal{R}$ we get a point $\mathbf{x}^*(r)$, with equality in Eq. (5). Then

$$c \leq \min_{\mathbf{r} \in \mathcal{R}} \mu(\mathbf{x}^*(\mathbf{r})) =: c_+(\mathcal{R}). \tag{7}$$

So for every set $\mathcal{R}$, this procedure estimates the optimal constant $c$ up to a precision $\varepsilon = c_+(\mathcal{R}) - c_-(\mathcal{R})$.

### Basic algorithm

The idea of the algorithm is now to let the set $\mathcal{R}$ grow step by step, which shrinks $\mathcal{P}(\mathcal{R})$, so $c_-(\mathcal{R})$ increases and $c_+(\mathcal{R})$ decreases (see Fig. 2 and Fig. 3). The algorithm terminates when $\varepsilon$ is below the target accuracy.

Apart from the set $\mathcal{R}$ it is useful to keep track of the polytope $\mathcal{P}(\mathcal{R})$ in the form of a list of vertices $\mathcal{E}(\mathcal{R})$ and edges. To arrive at the next approximation $\mathcal{R}' = \mathcal{R} \cup \{\mathbf{r}'\}$:

1. Determine a vertex $\mathbf{v}^* \in \mathcal{E}(\mathcal{R})$ at which $\mu$ becomes minimal, and set
$$\mathbf{r}' = \nabla\mu|_{\mathbf{v}^*}. \tag{8}$$
2. Solve the minimum-eigenvalue problem for $H(\mathbf{r}')$. This provides the bound $h(\mathbf{r}')$ for the new inequality (5), and an expectation tuple $\mathbf{x}^*$ corresponding to the ground state.
3. Compute $\mu(\mathbf{x}^*)$ and update $c_+(\mathcal{R}')$, if this is smaller than the current value.
4. Take the new inequality (5), and compute the intersections with all current edges of $\mathcal{P}(\mathcal{R})$. This will give some new extreme points for $\mathcal{E}(\mathcal{R}')$, and corresponding edges.
5. Evaluate $\mu$ on the new extreme points in $\mathcal{E}(\mathcal{R}')$ and update $c_-(\mathcal{R}')$. Terminate if $c_+(\mathcal{R}') - c_-(\mathcal{R}')$ is as small as desired. Otherwise go to step 1.

All these steps except the choice in step 1 are dictated by the geometry of outer approximation. The rationale of the choice (8) (apart from its flavour of gradient search) is that, whenever possible, it will eliminate the vertex $\mathbf{v}^*$ from $\mathcal{P}(\mathcal{R}')$, and thus strictly increase $c_-(\mathcal{R})$, unless there are other vertices with the same value of $\mu$, which have first to be eliminated in a similar manner. A proof of this statement is provided in the appendix. As an application of our method, we derived the uncertainty relations for two non-orthogonal spin components, see the appendix.

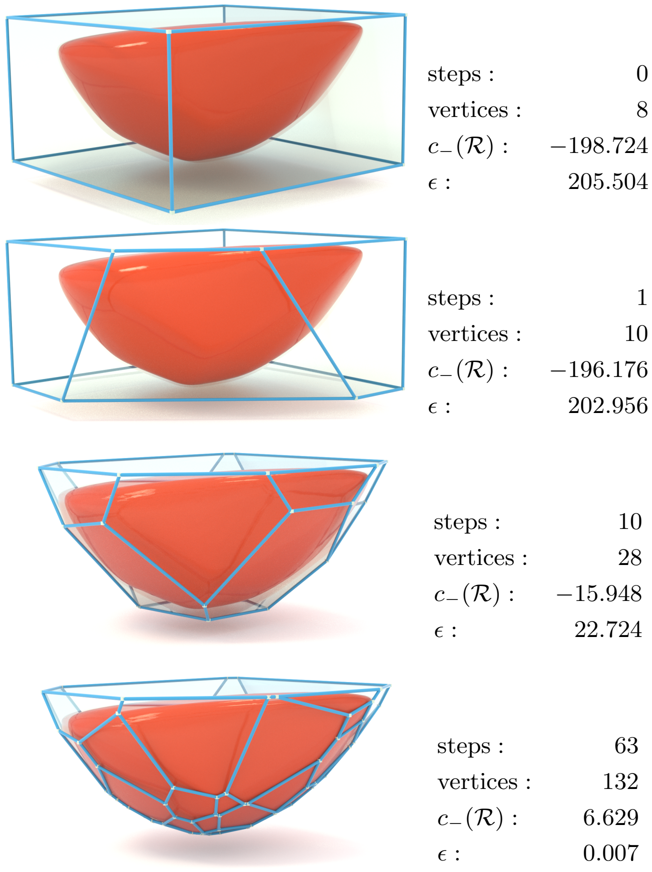


FIG. 3. Improving the outer approximation of $\mathcal{C}$ (red convex body) by adding more directions to the set $\mathcal{R}$. Every direction $\mathbf{r} \in \mathcal{R}$ gives a face of $\mathcal{P}(\mathcal{R})$ (blue polytope). New directions are chosen such that the vertex with the lowest value of $\mu$ will be cut off. Example generated from randomly chosen $A, B \in \mathbb{R}^{10\times10}$.

### Generalization to POVMs

Our method can be applied with minimal modifications to generalized measurements, i.e. observables given by positive operator valued measures (POVMs). In general, a POVM measurement $\mathcal{A}$ is described by its outcomes $\{a_i\}$ and corresponding effects $\{E_i\}$ [36, 37], where the probability of obtaining the outcome $a_i \in \mathbb{R}$ is given by $\mathrm{tr}(\rho E_i)$. The moments of an outcome distribution are then given by the expectations of the moment operators $A^{(n)} = \sum_i (a_i)^n E_i$. The only difference from the "standard" projection valued case is that the identity $A^{(n)} = \left(A^{(1)}\right)^n$ no longer holds. But this is not required for our method.

We therefore only need to express variances as $\Delta^2_\rho(A) = \langle A^{(2)}\rangle_\rho - \langle A^{(1)}\rangle^2_\rho$, and replace in (3) and the definition of $H(\mathbf{r})$: $A^2$ by $A^{(2)}$, $A$ by $A^{(1)}$, and analogously for $B$.

## APPLICATION TO ENTANGLEMENT DETECTION

In [4, 5], it was shown that every state-independent uncertainty relation like (4) yields a non-linear entanglement witness, when applied to local measurements in a bipartition. Here the following scenario is considered: Two parties, Alice and Bob, can perform local measurements $A_1, A_2$ such as $B_1, B_2$, on an unknown quantum state $\rho$. Their goal is to decide if $\rho$ is entangled or not. For this, they measure the 'sum observables' $M_1, M_2$, given by

$$M_i = A_i \otimes \mathbb{1} + \mathbb{1} \otimes B_i. \qquad (9)$$

In the POVM case this is generalized to measuring $A_i$ on Alice's side, $B_i$ on Bob's, and adding the outcomes, which results in

$$M_i^{(1)} = A_i^{(1)} \otimes \mathbb{1} + \mathbb{1} \otimes B_i^{(1)} \qquad (10)$$

$$M_i^{(2)} = A_i^{(2)} \otimes \mathbb{1} + 2A_i^{(1)} \otimes B_i^{(1)} + \mathbb{1} \otimes B_i^{(2)}. \qquad (11)$$

Now if $\rho = \rho_A \otimes \rho_B$ is uncorrelated, variances just add up, so

$$\Delta^2_\rho(M_1) + \Delta^2_\rho(M_2) \geq c_A + c_B, \qquad (12)$$

where $c_A$ and $c_B$ are the optimal uncertainty constants for the observable pairs $(A_1, A_2)$ and $(B_1, B_2)$, respectively. Since the variance is concave, this inequality holds also for all convex combinations of uncorrelated states, i.e., for all separable states [4].

Hence if (12) is violated, $\rho$ must be entangled. Of course, there is also an uncertainty bound $c_M$ for the observable pair $(M_1, M_2)$. So the interesting range allowing the conclusion "$\rho$ is entangled" is marked by

$$c_A + c_B > \Delta^2_\rho(M_1) + \Delta^2_\rho(M_2) \geq c_M. \qquad (13)$$

For angular momentum measurements, (12) can be seen [38] as a spin-squeezing criterion. As such, it requires the same experimental data as other spin squeezing criteria, see [39, 40], namely only a measurement of first and second moments of the total angular momentum. In contrast to entanglement criteria based on single outcomes, this requirement is very advantageous in typical experimental implementation, especially including many particle systems, see [41].

We further sharpen this criterion by applying it to the observable pairs $(\mu A_1, \lambda A_2)$ and $(\mu B_1, \lambda B_2)$. In this way we get two convex regions of pairs $(\Delta^2_\rho(M_1), \Delta^2_\rho(M_2))$: A larger one containing the pairs achievable with arbitrary states, given by the bounds of the type $c_M$, and a smaller one attainable by separable states, given by the bounds of the type $c_A + c_B$. As Fig. 4 shows, this increases the parameter range for which entanglement can be certified. The linear uncertainty bound with equal weights as a function of the local noise, evaluated for measurements $M_1$ and $M_2$ on separable and entangled states is shown by Fig. 8.

### Entanglement detection with noisy detectors

The generalization to POVMs increases the possibilities for entanglement detection. Suppose for the sake of discussion that before hitting the detector each subsystem goes through a known noisy channel. This typically increases variance [42], so traditional spin squeezing inequalities would often fail to detect entanglement. Indeed the state after the action of the noisy channels may well fail to be entangled. On the other hand, we might be interested in the presence of entanglement *before* the action of the noise. This is the appropriate view when the noise is inherent in the detection process. The noise is thus applied in the Heisenberg picture, turning even a standard projection valued measurement into

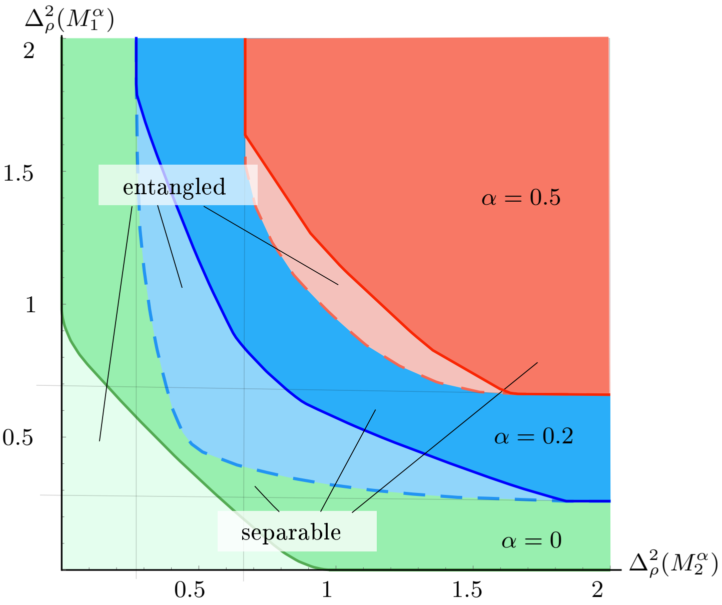


FIG. 4. Uncertainty regions for entangled and separable states. Superposition of the graphs for different noise levels $\alpha$: green= 0, blue= 0.2, red= 0.5. In this example we consider local measurements of orthogonal spin-1 components, i.e. $M_i = L_i^A + L_i^B$.

a proper POVM. This might easily find entanglement, which would go undetected by a direct application of the spin squeezing criterion.

These possibilities are shown in Fig. 4 by superimposing the entanglement detection regions for three different noise levels of a partially depolarizing channel $\rho \mapsto (1-\alpha)\rho + \alpha\rho_0$, where $\rho_0 \propto \mathbb{1}$ is the maximally mixed state, and $\alpha$ is a noise parameter. Increasing $\alpha$ shifts the diagram towards larger variances, but even for a modest noise level of $\alpha = 0.2$ the entanglement detection region lies entirely in the region where traditional spin squeezing (corresponding to $\alpha = 0$) would never find any entanglement.

## CONCLUSIONS AND OUTLOOK

We provided an algorithm for determining the optimal uncertainty bounds for two arbitrary observables. The precision of the bound is controlled as a duality gap, so terminating the iteration at any step gives a certified lower uncertainty bound together with an error estimate.

The method can, in principle, be extended to more observables, or to variances based not on quadratic but higher order deviations. However, this would increase the dimension of the geometric problem. Thus at every new approximation step one has to determine the intersection of the polytope with the new supporting hyperplane. This requires a better book-keeping of the topological structure of the polytopes, and a local version of the vertex enumeration problem [43].

The inequalities derived here have an immediate application to entanglement detection by generalized spin squeezing criteria. The possibility to use arbitrary observables (rather than orthogonal angular momentum components) greatly increases the versatility of this method.

It is an apparently open problem how strong the method becomes with arbitrary $A_i$, $B_j$, i.e. is every entangled state violating a local uncertainty relation. The problem has been studied carefully for orthogonal spin components [5, 6], but we do not know of a characterization of the (un-)detectable, possibly entangled states.

We gratefully acknowledge inspiring conversations and email exchange with Marcus Cramer, Otfried Gühne, Géza Tóth, Kais Abdelkhalek, David Reeb and Terry Farrelly.
We also acknowledge the financial support from the RTG 1991 and CRC 1227 DQ-mat funded by the DFG and the collaborative research project Q.com-Q funded by the BMBF.

## 4.2 A non-algorithmic bound on variance based uncertainty relations

In [SDW17] an algorithmic approach for computing lower bounds on linear uncertainty relations is provided. The central iteration step of this algorithm is to produce an increasing set of directions $\mathcal{R}$, which are used as face normals for approximating the joint numerical range $\mathcal{C}$ by an outer polyhedron.

Beside this algorithmic approach, we might alternatively directly aim to construct a fixed set of directions $\mathcal{R}_0$, in order to build an outer approximation and obtain bounds $c_-(\mathcal{R}_0)$ such as $c_+(\mathcal{R}_0)$. Those directions could also be used as an initial set-up for the above algorithm. However, not every set of directions guarantees non-trivial results, i.e.

(i) $c_-(\mathcal{R}_0)$ should always be non-negative

(ii) $c_-(\mathcal{R}_0)$ should be zero, if and only if the optimal bound $c$ is zero.

Triviality, in the above sense, usually occurs when $\mathcal{R}_0$ is chosen too small, i.e. if the resulting approximation is too inaccurate. In contrast, if $\mathcal{R}_0$ is too big, computations might become unhandy in practice.

In this section, we will provide an appropriate set $\mathcal{R}_0$, for projective measurements. This set leads to a non-trivial bound $c_-(\mathcal{R}_0)$, but is still efficiently small, i.e. $|\mathcal{R}_0| \approx d^2$. Furthermore, $\mathcal{R}_0$ can be directly deduced from the spectra of the observables $A$ and $B$, which makes it easy to compute in practice.

**Theorem 4.1.** *Let $\{a_i\}$ and $\{b_i\}$ be the outcomes of projective measurements $A$ and $B$ given in non-decreasing order. A non-trivial set of initial directions is given by the union $\mathcal{R}_0 = \mathcal{R}_{\mathcal{P}_0} \cup \mathcal{R}_{\mathcal{G}}$, with*

$$\begin{aligned}\mathcal{R}_{\mathcal{P}_0} =& \{(-a_i - a_{i+1}, -b_j - b_{j+1}, 1)\,|i,j \in \{1,\cdots,d\}\} \\ &\cup \{(1,0,0),(-1,0,0),(0,1,0),(0,-1,0)\} \\ &\cup \{(-a_1 - a_d, -b_1 - b_d, -1)\}\end{aligned}$$

*and*

$$\mathcal{R}_{\mathcal{G}} = \left\{\nabla\mu|_{P_{ij}} = (-a_i, -b_j, 1)\,|i,j \in \{1,\cdots,d\}\right\}.$$

*Proof.* Consider the points $P_{ij} = (a_i, b_j, a_i^2 + b_j^2)$, which only depend on the measurement outcomes $\{a_i\}$ and $\{b_i\}$, and construct a convex polytope $\mathcal{P}_0 = \operatorname{conv}(P_{ij})$ by taking the $P_{ij}$ as vertices.

This proof is based on the following three properties, which will be proven immediately:

(i) $\mathcal{P}_0$ is an outer approximation, i.e.: $\mathcal{C} \subseteq \mathcal{P}_0 = \mathrm{conv}(P_{ij})$

(ii) The variance functional $\mu$ vanishes on the vertices of $\mathcal{P}_0$, i.e. $\mu(P_{ij}) = 0$. Therefore, we have a non-negative $\mu$ in the interior of $\mathcal{P}_0$

(iii) $\mathcal{P}_0$ has only $\mathcal{O}(d^2)$ vertices, faces and edges.

Here our central idea, for constructing $\mathcal{R}_0$, is to approximate $\mathcal{C}$ by an refinement of $\mathcal{P}_0$. This is always possible (because of (i)), not to complex (because of (iii)), and will always lead to bounds with positive $\mu$ (because of (ii)).

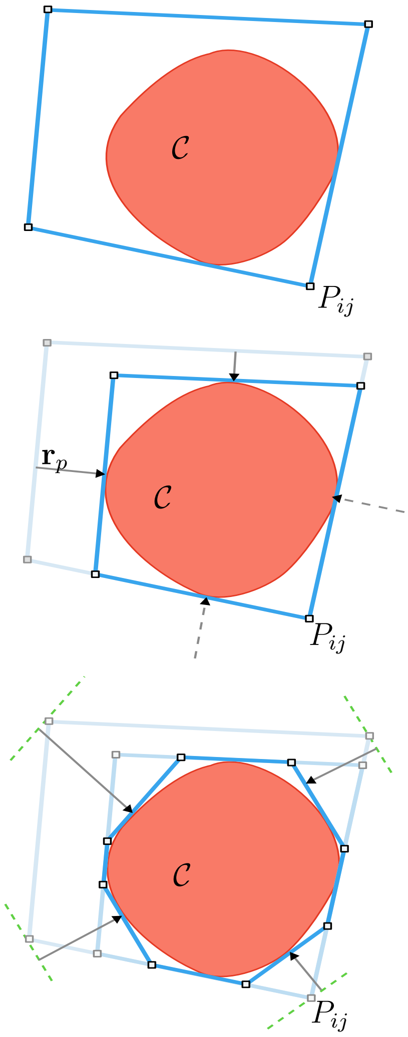


Figure 4.2: On one hand, the points $P_{ij}$ are chosen such that $\mu(P_{ij}) = 0$ and, on the other, such that $\mathcal{C}$ is included in the convex polytope $\mathcal{P}_0$, which is constructed by taking the $P_{ij}$ as vertices. Here, the set $\mathcal{R}_{\mathcal{P}_0}$ consists of the face normals of this polyhedron. This ensures non-negativity. The set $\mathcal{R}_{\mathcal{G}}$ consists of the gradients of $\mu$ evaluated at the $P_{ij}$, which grants that the points $P_{ij}$ are not included in $\mathcal{P}(\mathcal{R}_0)$. Hence, approximating $\mathcal{C}$ with directions $\mathcal{R}_0$ gives a strict improvement to $\mathcal{P}_0$. Therefore, a non-trivial bound $c_{-}(\mathcal{R}_0)$ is guaranteed.

At first, we will take the face normals of $\mathcal{P}_0$ as directions $\mathcal{R}_{\mathcal{P}_0}$ for an approximation of $\mathcal{C}$. Geometrically spoken (see Fig. 4.2): we shift the faces of $\mathcal{P}_0$ to its interior until they 'touch' the approximated set $\mathcal{C}$. This will give us a new polytope $\mathcal{P}(\mathcal{R}_{\mathcal{P}})$ that lies in-between $\mathcal{C}$ and $\mathcal{P}_0$. By this, we have the guarantee that the bound $c_-(\mathcal{R}_{\mathcal{P}})$ will be non-negative. The computation of the face normals $\mathcal{R}_{\mathcal{P}_0}$ can be found at the end of this proof.

However, non-triviality is not guaranteed by $\mathcal{R}_{\mathcal{P}}$, because it could happen that some vertices of the polygon $\mathcal{P}(\mathcal{R}_{\mathcal{P}})$ equal some of the $P_{ij}$ even if these points are not in $\mathcal{C}$, i.e. some faces of $\mathcal{P}_0$ could be 'touched' by $\mathcal{C}$ such that they will not 'move' when we try to shift them to the interior. If this is the case for all adjacent faces of a point $\mathcal{P}_{ij}$, this point will also not 'move'.

Secondly, in order to ensure that the above does not happen, we have to add further faces to $\mathcal{P}(\mathcal{R}_{\mathcal{P}})$, in order to 'remove' all points $P_{ij}$ that are not in $\mathcal{C}$. For this we take the gradients of the functional $\mu$ at the points $P_{ij}$ as further directions $\mathcal{R}_{\mathcal{G}}$ (see Fig. 4.2,below). Explicitly, these additional directions are given by

$$\mathcal{R}_{\mathcal{G}} = \left\{ \nabla\mu|_{P_{ij}} = (-a_i, -b_j, 1)\,|i, j \in \{1, \cdots, d\} \right\}.$$

All together, computing $c_-(\mathcal{R}_0)$ and $c_+(\mathcal{R}_0)$ for

$$\mathcal{R}_0 = \mathcal{R}_{\mathcal{P}} \cup \mathcal{R}_{\mathcal{G}}$$

guarantees to obtain a state independent and non trivial uncertainty relation by only requiring to solve $|\mathcal{R}_0| = \mathcal{O}(d^2)$ ground state problems for constructing $\mathcal{P}(\mathcal{R}_0)$.

*(i) $\mathcal{C}$ is included in* conv$(\{P_{ij}\}$*:* As in [SDW115], the joint numerical range $\mathcal{C}$ can be written as

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{R}^3 | \mathbf{r}.\mathbf{x} \geq h(\mathbf{r})\ \forall \mathbf{r} \in \mathbb{R}^3\},$$

with

$$h(\mathbf{r}) = \min_{\rho} \langle r_1 A + r_2 B + r_3 \left(A^2 + B^2\right)\rangle_{\rho}.$$

By setting

$$h_0(\mathbf{r}) := \min_{\mathbf{x}\in\mathcal{P}_0} \mathbf{r}.\mathbf{x}$$

we can write $P_0$, in a similar way:

$$\mathcal{P}_0 = \{\mathbf{x} \in \mathbb{R}^3 | \mathbf{r}.\mathbf{x} \geq h_0(\mathbf{r})\ \forall \mathbf{r} \in \mathbb{R}^3\}. \tag{4.1}$$

So, for proving the inclusion $\mathcal{C} \subset \mathcal{P}_0$, it suffices to show

$$h_0(\mathbf{r}) \leq h(\mathbf{r}) \quad \forall \mathbf{r} \in \mathbb{R}^3 \tag{4.2}$$

Because then, every point of $\mathcal{C}$ will also fulfil the defining constrains given in (4.1).

We can see the validity of (4.2) by,

$$\begin{aligned} h(\mathbf{r}) &= \min_\rho \langle r_1 A + r_2 B + r_3 \left(A^2 + B^2\right)\rangle_\rho \\ &\geq \min_\rho \langle r_1 A + r_3 A^2\rangle_\rho + \min_\rho \langle r_2 B + r_3 B^2\rangle_\rho \\ &\geq \min_{ij} \left(r_1 a_i + r_3 a_i^2 + r_2 b_j + r_3 b_j^2\right) \\ &= \min_{ij} \mathbf{r}.P_{ij} = h_0(\mathbf{r}) \end{aligned}$$

Here, we used the sub-additivity of the minimum in the first step. In the second step we used that the minimal expectation of any measurement will be attained on a point density and, in the last step, we used that the minimum of a linear functional on a convex set will be attained on its extreme points.

*(ii) $\mu(P_{ij}) = 0$:* The points $\mu(P_{ij})$ were constructed by taking the spectral tuples $(a_i, b_j)$ and projecting them to the paraboloid $\mu(x, y, z) = 0$. Hence, (ii) can easily be checked by plugging $P_{ij} = (a_i, b_j, a_i^2 + b_j^2)$ into $\mu(x, y, z) = z - x^2 - y^2$.

*(iii) number of vertices, edges and faces:* We have (including degeneracies) $d$ outcomes $a_i$ or $b_j$. Hence, we have $d^2$ pairs $(a_i, b_j)$.

*The facenormals of $\mathcal{P}_0$ (sketch):* We want to show that the face normals

$$\begin{aligned} \mathcal{R}_\mathcal{P} = &\left\{(-a_i - a_{i+1}, -b_j - b_{j+1}, 1) \,|\, i, j \in \{1, \cdots, d\}\right\} \\ &\cup \left\{(1,0,0), (-1,0,0), (0,1,0), (0,-1,0)\right\} \\ &\cup \left\{(-a_1 - a_d, -b_1 - b_d, -1)\right\}, \end{aligned}$$

with $\{a_i\}$ and $\{b_i\}$ given in non decreasing order, describe the faces of $\mathcal{P}_0$.

At first, we show that any of those directions gives a face of $\mathcal{P}_0$: Consider a direction $\mathbf{r}_{ij} = (-a_i - a_{i+1}, -b_j - b_{j+1}, 1)$. It is straight forward to check that this direction is the normal of a plane through the points $\{P_{i,j}, P_{i+1,j}, P_{i,j+1}, P_{i+1,j+1}\}$. This plane is a face of $\mathcal{P}_0$, if we have

$$\min_{P_{kl}} \mathbf{r}_{ij} P_{kl} = \mathbf{r}_{ij} P_{ij} = \mathbf{r}_{ij} P_{i+1,j} = \mathbf{r}_{ij} P_{i,j+1} = \mathbf{r}_{ij} P_{i+1,j+1}. \tag{4.3}$$

We expand the first minimum in (4.3) and get

$$\min_{(x,y,x^2+y^2) \in \{P_{kl}\}} \mathbf{r}_{ij}(x, y) = x^2 - a_i x - a_{i+1} x + y^2 - b_j y - b_{j+1} y. \tag{4.4}$$

For general $(x, y) \in \mathbb{R}^2$, the convex quadratic functional on the r.h.s. of (4.4) has its minimum on $x^* = 1/2(a_i + a_{i+1})$ and $y^* = 1/2(b_j + b_{j+1})$. For the discrete minimization, the minimum in (4.4) is attained at those points $P_{kl}$, which are the closest to $(x^*, y^*)$. These are the points $\{P_{i,j}, P_{i+1,j}, P_{i,j+1}, P_{i+1,j+1}\}$, because the $a_i$ and $b_j$ are given in non-decreasing order.

With the same argumentation as above, it can be checked that the points of the form $P_{1,j}$, $P_{d,j}$, $P_{i,1}$, and $P_{i,d}$, form faces corresponding to the normals $r_{side} = \{(1,0,0), (-1,0,0), (0,1,0), (0,-1,0)\}$, and that the points $\{P_{1,1}, P_{d,1}, P_{1,d}, P_{d,d}\}$ form a face with the normal $r_{top} = \{(-a_1 - a_d, -b_1 - b_d, -1)\}$.

We can check that all those faces correspond to a polyeder by Euler's formula [Eul58]

$$\#vertices - \#edges + \#faces = 2.$$

Careful counting shows (compare Fig. 4.3): The directions $r_{ij}$ belong to $(d-1)^2$ faces, with $2d(d-1)$ edges. The directions $r_{side}$ add 4 new faces with 4 new edges, and the face $r_{top}$ only adds a new face between points that are already connected by edges. Hence, we get

$$\underbrace{\left(d^2\right)}_{\#vertices} - \underbrace{\left(2d(d-1)+4\right)}_{\#edges} + \underbrace{\left((d-1)^2+4+1\right)}_{\#faces} = 2$$

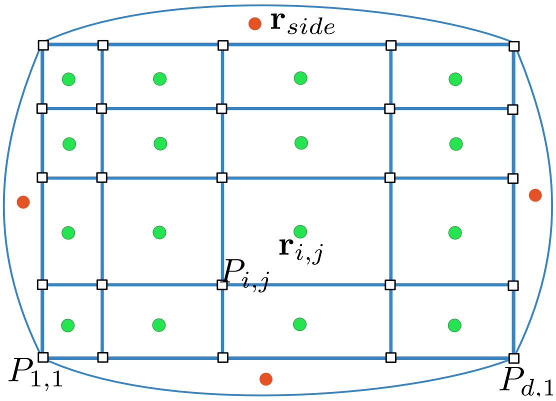


Figure 4.3: Planar graph representation of $\mathcal{P}_0$: The face normals ●$\mathbf{r}_{ij}$ and ●$\mathbf{r}_{side}$ are indicated. The face corresponding to $\mathbf{r}_{top}$ corresponds to the outer region of this graph.

□

# CHAPTER 5

# An information theoretic view on measurement uncertainty

In this chapter we will change our view on uncertainty relations to an information theoretic perspective. Here, we regard the outcomes of a measurement as elements of a finite set with non-specified structure. Without loss of generality, we assume that the elements of an outcome set are labelled by labels coming from the index set $\{1, ..., n\}$, also called *alphabet* in the following. From this perspective we concentrate our interest to the probability that a particular label occurs as outcome of a measurement.

In the following sections we will introduce different error measures and their resulting measurement uncertainty relations, and we will investigate their interplay with preparation uncertainty relations, as well. A diagram of all those interplays can be found at the end of this section. In the first section we will concentrate on the discrete metric as cost function. Here we will introduce the according errors $\varepsilon^{dm}(A, A')$ and deviations $\nu^{dm}(A|\rho)$. In the second section we will put our focus on entropic quantities. On one hand, we will show that the, so called, *self-information* $-\log(p_i)$ can be used as cost function that provides us an error measures $\varepsilon^{info}(A, \mathtt{A}')$ and will lead to preparation uncertainty relations, in terms of the Shannon entropy $H(A|\rho)$, in a natural way.

On the other, we will consider measurement uncertainty formulated in terms of conditional entropies $H(X|R^X)$. In contrast to all other relations provided in this thesis, those relations can be defined between arbitrary outcome spaces. Hence, we can drop the marginal construction from the definition of a joint measurement. However, this construction does not rely on an underlying cost function directly and we will see that only an analogue to the entanglement reference error is well defined in this case.

## 5.1 Uncertainty relations for the discrete metric

For this section let $I = \{1, \dots, n\}$ denote a finite alphabet with $n$ letters. In the following we will apply the constructions from [SRW16] and the corresponding discrete preparation uncertainty, from Ch.3, with discrete metric $c(i,j) = 1 - \delta_{ij}$ as cost function. For this case, the measurement error was also investigated in the bachelor thesis [Fra15]. It turned out that, due to the simple form of the discrete metric, the corresponding Wasserstein distance and the measurement error can be expressed in a simplified way, too. In this thesis, we will also consider the corresponding calibration and entangled reference frame errors. We will see that both errors admit an even simpler form.

**Measurement errors:** For the discrete metric, the minimization of the transport cost, between probability distributions $p$ and $p'$, can be solved analytically. Here, ( see e.g. [Vil09, p.22] or [Fra15] for a proof) the minimal transport cost is given by the half of the 1-norm distance:

$$\check{c}(p,p') = \frac{1}{2}\|p - p'\|_1.$$

Commonly, this quantity is also called the *total variational distance.* Alternatively (see [Fra15]), this norm distance can be written as minimization of a function, linear in $p, p'$, over all subsets of $I$, i.e. as

$$\frac{1}{2}\|p - p'\|_1 = \max_{X \subset I} \sum_{i \in X} p_i - p'_i.$$

This particular form is advantageous when we consider $p$ and $p'$ to be the outcome probabilities of measurements $A$ and $A'$ on a state $\rho$. We can use the linearity and write the measurement error as

$$\begin{aligned} \varepsilon_M^{dm}(A|A') &= \sup_\rho \max_{X \subset I} \sum_{i \in X} \operatorname{tr}(\rho A(i)) - \operatorname{tr}(\rho A'(i)) \\ &= \max_{X \subset I} \sup_\rho \operatorname{tr}\left(\rho \sum_{i \in X} A(i) - A'(i)\right). \end{aligned}$$

Here, we use the notation $A(X) = \sum_{i \in X} A(i)$ and interpret the supremum over all states as operator norm. Hence, we have

$$\varepsilon_M^{dm}(A|A') = \max_{X \subset I} \|A(X) - A'(X)\|_\infty$$

If $A$ is a sharp measurement, with eigenprojectors $\{\phi_i := A(i)\}_{i \in I}$, we are in a position to assign a calibration error $\varepsilon_C^{dm}(A|A')$ and an entangled reference frame error $\varepsilon_E^{dm}(A|A')$. Using the definitions in Eq.(6) and Eq.(10) from [SRW16] we get

$$\begin{aligned}\varepsilon_C^{dm}(A|A') &= \max_i \sum_{j\in I} \operatorname{tr}\left(A'(j)\phi_i\right) c(i,j) \\ &= \max_{i\in I} \sum_{j\neq i} \operatorname{tr}\left(A'(j)\phi_i\right) = 1 - \min_{i\in I} \operatorname{tr}\left(A'(i)\phi_i\right)\end{aligned}$$

for the calibration error and

$$\begin{aligned}\varepsilon_E^{dm}(A|A') &= \frac{1}{d}\sum_i \sum_{j\in I} \operatorname{tr}\left(A'(j)\phi_i\right) c(i,j) \\ &= \frac{1}{d}\sum_{i\in I}\sum_{j\neq i} \operatorname{tr}\left(A'(j)\phi_i\right) = 1 - \frac{1}{d}\sum_{i\in I} \operatorname{tr}\left(A'(i)\phi_i\right)\end{aligned} \tag{5.1}$$

for the entanglement reference frame error. We can see that both error quantities only depend on the probabilities $p_{j=i} := \operatorname{tr}\left(A'(i)\phi_i\right)$, in a way that maximizing these probabilities will decrease the error. Those probabilities have a clear operational meaning: If we test a device $A'$ with an eigenstate $\phi_i$, the probability for getting the correct outcome $i$ is given by $p_{j=i}$. Thereby, the probabilities for the other outcomes occur only indirectly by the relation $p_{i=j} = 1 - p_{j\neq i}$. Hence, for a test on $\phi_i$ all labels $j \neq i$ are treated equally by our error quantities. This nicely illustrates the nature of the underlying metric.

Despite the fact that all three error quantities now appear in a simple and intuitively interpretable form, the computation of a corresponding measurement uncertainty relation still demands to consider the semi definite boundary conditions of a joint measurement. Hence, this computation still has to be done by solving an SDP using the formulation given in [SRW16] Sec.3 .

**Preparation uncertainty:** According to the definition Def. 3.1, the generalized deviation in terms of the discrete metric is given by

$$\begin{aligned}\nu(A|\rho) &= \min_j \sum_i \operatorname{tr}(\phi_i\rho)c(i,j) \\ &= \min_j \sum_i \operatorname{tr}(\phi_i\rho)(1-\delta_{ij}) = 1 - \max_j \operatorname{tr}\left(\phi_j\rho\right).\end{aligned}$$

After performing the maximization over $j$, this quantity only depends on a single outcome probability $p_{j_{max}} = \operatorname{tr}\left(\phi_{j_{max}}\rho\right)$, too. Again this probability, i.e the probability of the most probable outcome, has a clear interpretation and also many applications in information theory. Indeed, by taking the negative logarithm of $p_{j_{max}}$, we retrieve the so called *min entropy*

$$H_\infty(p_\rho^A) = -\log\left(p_{j_{max}}\right),$$

which is usually considered instead of $p_{j_{max}}$ itself.

For a collection of measurements $A_1, \dots, A_m$ and weights $a_1, \dots, a_m$, the minimization problem for the corresponding linear uncertainty relation is given by

$$\begin{aligned}\inf_\rho \sum_{j=1\dots m} \alpha_j \nu(A_j|\rho) &= 1 - \sup_\rho \sum_{j=1\dots m} \alpha_j \max_i \operatorname{tr}\left(A_j(i)\ \rho\right) \\ &= 1 - \max_{i_1 \dots i_m} \left\| \sum_{j=1\dots m} \alpha_j A_j(i_j) \right\|_\infty, \end{aligned} \tag{5.2}$$

which can be solved in the general case by computing the norm above for all combinations $(i_1, \dots, i_m) \in \{1, \dots, d\}^m$ separately. For the special case of only two ideal sharp measurements, say $A$ and $B$, this can be further simplified:

Here, all necessary information on the problem is given by specifying the overlaps $\langle \phi_i^A | \phi_j^B \rangle$ between the respective eigenstates of $A$ and $B$.

**Lemma 5.1.** *Let $A$ and $B$ denote sharp measurements with eigenstates $\{\phi_i^A\}$ and $\{\phi_i^B\}$. The linear preparation uncertainty between $A$ and $B$ is given by*

$$\inf_\rho a\ \nu(A|\rho) + b\ \nu(B|\rho) = \frac{1}{2}\left(2 - a - b - \sqrt{a^2 + b^2 + 2ab(2c^* - 1)}\right)$$

*and therefore only depends on the maximal overlap*

$$c^* = \max_{ij} |\langle \phi_i^A | \phi_j^B \rangle|$$

*Proof.* The central part of this proof is to evaluate the norm from (5.2) for weights $(a, b)$ and eigenstates $A(i) = \phi_i^A$, $B(j) = \phi_j^B$. We have to compute

$$\eta_{ij} := \left\| a\phi_i^A + b\phi_j^B \right\|_\infty \tag{5.3}$$

This can be done completely within the two dimensional subspace spanned by the vectors $|\phi_i^A\rangle$ and $|\phi_j^B\rangle$. The norm (5.3) is unitarily invariant. Hence, we can represent those vectors, w.l.o.g., as

$$|\phi_i^A\rangle = (1, 0) \quad \text{and} \quad |\phi_j^B\rangle = (\cos(\theta), \sin(\theta))$$

with a relative angle $\theta$, i.e. $\cos(\theta) = |\langle \phi_i^A | \phi_j^B \rangle|$. Within this representation (5.3) is reduced to the norm of a $2 \times 2$ matrix, which can be computed analytically. We get:

$$\left\| \begin{pmatrix} a + b\cos^2(\theta) & b\cos(\theta)\sin(\theta) \\ b\cos(\theta)\sin(\theta) & b\sin^2(\theta) \end{pmatrix} \right\|_\infty = \frac{1}{2}\left(a + b + \sqrt{a^2 + b^2 + 2ab(2\cos(\theta) - 1)}\right)$$

For $\theta \in (0, \pi/2)$, i.e. for $\cos(\theta) \in (0, 1)$, this norm is a monotonously increasing function of $\cos(\theta)$. Therefore, the uncertainty relation

$$1 - \max_{i,j} \eta_{ij}$$

is saturated on the indices $(i, j)$ belonging to the largest angle/overlap

$$c^* = \max_{ij} |\langle \phi_i^A | \phi_j^B \rangle|$$

$\square$

For the special case of equal weights $a = b = 1$, the lemma Lem. 5.1 gives the simple uncertainty relation

$$\inf_\rho \nu(A|\rho) + \nu(B|\rho) = 1 - c^* \tag{5.4}$$

and by this a direct interpretation for the constant $c^*$. Furthermore, this constant, more precisely its negative logarithm, plays a prominent role for preparation uncertainty relations in terms of Shannon entropies, as well. Here, we have the Maassen and Uffink bound [MU88] :

$$\inf_\rho \frac{1}{2} H(A|\rho) + \frac{1}{2} H(B|\rho) \geq -log(c^*),$$

which is, in contrast to (5.4), only a lower bound, that is tight only for the very special cases documented in [ASM$^+$15] (Sec.7.1).

## 5.2 Entropic measurement uncertainty relations

In this section we will provide two constructions for the formulation of measurement errors in terms of entropic quantities. For sharp measurements, we will show that both types of error lead to measurement uncertainty relations that can be lower bounded by linear preparation uncertainty relations in terms of the Shannon entropy.

### The self-information is a cost function

The following lemma shows that we can reinterpret the Shannon entropy as deviation with respect the to the self-information as cost function.

**Lemma 5.2.** *Let $\mathcal{P}_n$ denote the set of all probability distributions over an alphabet $I$ of length $n$. Furthermore, let $c : I \times \mathcal{P}_n \mapsto \mathbb{R}_+$ be the function that assigns the self-information, with respect to a distribution $q \in \mathcal{P}_n$, to a label $i \in I$, i.e.*

$$c(i, q) = -\log(q_i).$$

*Let $X$ be a random variable with outcomes on $I$, distributed by some $p \in \mathcal{P}_n$. The Shannon entropy of $p$, i.e. $X$, is given by the deviation of $X$ with respect to $c$ as cost function, i.e. we have*

$$H(X) = -\sum_{i \in I} p_i \log(p_i) = \inf_{q \in \mathcal{P}_n} \langle c(X, q) \rangle_p = \nu(X, p)$$

*Proof.* Let $p, q \in \mathcal{P}_n$ be probability distributions. The relative entropy $D(p||q)$, between those distributions is always non negative, i.e. we have

$$0 \leq D(p||q) = \sum_{i \in I} p_i \log\left(\frac{p_i}{q_i}\right). \tag{5.5}$$

This estimate is also known as Gibbs' inequality. By expanding the logarithm, this is equivalent to

$$-\sum_{i \in I} p_i \log(p_i) \leq -\sum_{i \in I} p_i \log(q_i) \quad \forall q \in \mathcal{P}_n.$$

Equality in (5.5) is achieved by $p = q$. Therefore we have

$$-\sum_{i \in I} p_i \log(p_i) = \min_{q \in \mathcal{P}_n} \sum_{i \in I} p_i \left(-\log(q_i)\right) = \min_{q \in \mathcal{P}_n} \sum_{i \in I} p_i c(i, q) = \inf_{q \in \mathcal{P}_n} \langle c(X, q) \rangle_p$$

□

Given the self-information as cost function, we can employ the construction from [SRW16] to obtain the according measurement error quantities. Those quantities can be used to compare a measurement $A$, with outcomes $i$ from the alphabet $I$, with a measurement $\mathtt{A}'$, that has outcomes $q$ on the set of probability distributions $\mathcal{P}_n$, with $n = |I|$. Hence, we will have to model this measurement by an operator valued measure with continuous support. In the following we will write $\mathtt{A}'[\omega]$ to denote the operator valued measure corresponding to a measurable set $\omega \subset \mathcal{P}_n$. Furthermore, we will use the notation

$$\int_\omega \mathrm{tr}\left(\mathtt{A}'[dq]\rho\right)$$

to denote the probability that a measurement on a state $\rho$ gives a guess (probability distribution) $q \in \omega$ as outcome.

When tested on eigenstates $\{\phi_i\}$, such a device $\mathtt{A}'$ has a clear operational interpretation: for every test state $\phi_i$ the correct answer, a perfect measurement device would give, is $i$. An approximate device with outcomes on $I$, as in the last section, will give this correct answer with some probability, or a differing label $j \neq i$ with the opposite probability. A device $\mathtt{A}'$, as in this section, will instead respond to a test state $\phi_i$ with an educated guess of the form:

'with probability $q_1$ the input was $\phi_1$, with probability $q_2$ the input was $\phi_2$, and so on...'

Here, $\mathtt{A}'$ would be a perfect approximation, i.e. equivalent, to a sharp measurement $A$, if $\mathtt{A}'$ always gives the certain answer $q : q_j = \delta_{ij}$, resulting to the input $\phi_i$. In the following, we will restrict to sharp measurements and regard only error measures evaluated on eigenstates of an ideal measurement. Although, all quantities of this section are also well defined for the more general cases, an investigation of those cases is still open and, therefore, not presented here.

**Definition 5.3.** According to [SRW16], the *entropic calibration* and the *entropic entangled reference frame error* are given by

$$\varepsilon_C^{info}(A|\mathtt{A}') = \sup_{i \in I} \int_{\mathcal{P}_n} \operatorname{tr}\bigl(\mathtt{A}'[dq]\phi_i\bigr)\bigl(-\log(q_i)\bigr)$$

and

$$\varepsilon_E^{info}(A|\mathtt{A}') = -\frac{1}{d} \int_{\mathcal{P}_n} \operatorname{tr}\Bigl(\mathtt{A}'[dq] \sum_{i \in I} \phi_i \log(q_i)\Bigr). \tag{5.6}$$

Since $\mathcal{P}_n$ is continuous, we had to replace the sum over all outcomes by an integral with respect to the operator valued measure $\mathtt{A}'[dq]$. A careful check of Thm. 3.4 shows that this theorem is valid in this case as well. Therefore, we can directly conclude the following corollary:

**Corollary 5.4.** *Let $A_1, \dots, A_m$ be a collection of sharp measurements with outcomes on the alphabet $I$, and let $\vec{a} = (a_1, \dots, a_m)$ be a vector of positive weights. For any joint measurement $\mathtt{R}$, with outcomes on $\mathcal{P}_n \times \cdots \times \mathcal{P}_n$, let $\mathtt{A}'_j$ denote the respective marginals with outcomes on $\mathcal{P}_n$. We have*

$$\inf_{\mathtt{R}} \sum_{j=1,\dots,m} a_j \varepsilon_E^{info}(A_j|\mathtt{A}'_j) \geq \inf_{\rho} \sum_{j=1,\dots,m} a_j H(A_j|\rho).$$

It is clear from the definition that in this case the entangled reference error is smaller than the calibration error, as well. Hence, by using the shorthand notation of Ch. 3, we can state:

$$\inf_{\mathtt{R}} \vec{a} \cdot \vec{\varepsilon}_C^{\,info}(A|\mathtt{A}') \geq \inf_{R} \vec{a} \cdot \vec{\varepsilon}_E^{\,info}(A|\mathtt{A}') \geq \inf_{\rho} \vec{a} \cdot \vec{H}(A|\rho).$$

Furthermore, there is a direct correspondence between the devices from this section and the devices from the last: In a single shot, a device $\mathtt{A}'$ gives us a probability distribution $q$ as outcome. Given such a particular outcome, we can take a sample from this distribution and obtain a single outcome from the alphabet $I$. By this procedure, see Fig. 5.1, we emulate a device $A'$ with outcomes on $I$.

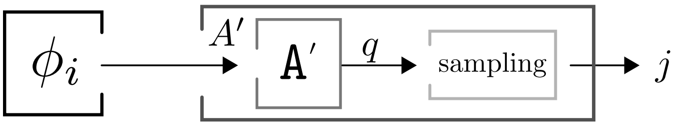


Figure 5.1: We can emulate a device $A'$, with outcomes on $I$, by sampling from the single-shot outcome $q$ of $\mathtt{A}'$.

In this context it also makes sense to compare the error quantities $\varepsilon_M^{info}(A_j|\mathtt{A}_j')$ and $\varepsilon_M^{dm}(A_j|A_j')$. As the following lemma shows, we have an explicit hierarchy:

**Lemma 5.5.** *Let $A_1, \ldots, A_m$ be a collection of sharp measurements with outcomes on $I$, and let $\vec{a} = (a_1, \ldots, a_m)$ be a vector of positive weights.*

*(i)For any joint measurement $\mathtt{R}$, with outcomes on $\mathcal{P}_n \times \cdots \times \mathcal{P}_n$, let $\mathtt{A}_j'$ denote the respective marginal measurement with outcomes on $\mathcal{P}_n$.*

*(ii)For any joint measurement $R$, with outcomes on $I \times \cdots \times I$, let $A_j'$ denote the respective marginal measurement with outcomes on $I$.*

*We have*

$$\inf_{\mathtt{R}} \sum_{j=1,\ldots,m} a_j \varepsilon_E^{info}(A_j|\mathtt{A}_j') \geq \inf_{R} \sum_{j=1,\ldots,m} a_j \varepsilon_E^{dm}(A_j|A_j') \tag{5.7}$$

*Proof.* At first, consider a fixed device $\mathtt{A}'$. To this device we assign a second device $A'$, which is given by the POVM elements

$$A'(j) := \int_{\mathcal{P}_n} \mathtt{A}'[dq] q_j.$$

If we sample from the outcomes $q$ of a measurement device $\mathtt{A}'$, the probability of getting the outcome $j$, as response to the input $\phi_i$, is given by

$$p_{ij} = \mathrm{tr}(A'(j)\phi_i) = \int_{\mathcal{P}_n} \mathrm{tr}(\mathtt{A}'[dq]\phi_i)q_j.$$

Hence, our emulated device $A'$ attains the error (see (5.1))

$$\begin{aligned}
\varepsilon_M^{dm}(A_j|A_j') &= 1 - \frac{1}{d}\sum_{i\in I} p_{ii} \\
&= 1 - \frac{1}{d}\sum_{i\in I}\int_{\mathcal{P}_n} \mathrm{tr}(\mathtt{A}'[dq]\phi_i)q_i \\
&= \frac{1}{d}\sum_{i\in I}\int_{\mathcal{P}_n} \mathrm{tr}(\mathtt{A}'[dq]\phi_i)(1-q_i)
\end{aligned}$$

We have $0 \leq q_i \leq 1$. Hence, we can employ the estimate $1 - q_i \leq -\log(q_i)$ in the above and get

$$\varepsilon_M^{dm}(A_j|A_j') \leq \frac{1}{d}\sum_{i\in I}\int_{\mathcal{P}_n} \mathrm{tr}(\mathtt{A}'[dq]\phi_i)\,(-\log(q_i)) = \varepsilon_M^{info}(A_j|\mathtt{A}_j').$$

This already gives us the necessary tool for proving (5.7). If we apply the above estimate for all marginal observables of any $\mathtt{R}$, we will directly obtain a joint measurement $R$ with smaller absolute error. □

**Interpretation of $\varepsilon_M^{info}$:** The entropic reference frame error $\varepsilon_E^{info}(A|\mathtt{A}')$, see (5.6), can be rewritten as

$$\varepsilon_E^{info}(A|\mathtt{A}') = \int_{\mathcal{P}_n} \frac{\mathrm{tr}(\mathtt{A}'[dq])}{d} \sum_{i\in I} \frac{\mathrm{tr}(\mathtt{A}'[dq]\phi_i)}{\mathrm{tr}(\mathtt{A}'[dq])} \log\left(\frac{1}{q_i}\right).$$

Here, we substitute the states

$$\rho_q = \frac{\mathtt{A}'[dq]}{\mathrm{tr}(\mathtt{A}'[dq])}$$

and the probabilities

$$(p_{\rho_q}^A)_i = \mathrm{tr}(\rho_q\phi_i) = \frac{\mathrm{tr}(\mathtt{A}'[dq]\phi_i)}{\mathrm{tr}(\mathtt{A}'[dq])}$$

and the probability measure

$$\mu[dq] = \operatorname{tr}\left(\frac{\mathtt{A}'[dq]}{d}\right)$$

to get

$$\varepsilon_E^{info}(A|\mathtt{A}') = \int_{\mathcal{P}_n} \mu[dq] \sum_{i \in I} \left(p_{\rho_q}^A\right)_i \log\left(\frac{1}{q_i}\right). \tag{5.8}$$

The sum in (5.8) is also known as *cross entropy*, denoted by $H(p;q)$. We get

$$\varepsilon_E^{info}(A|\mathtt{A}') = \int_{\mathcal{P}_n} \mu[dq] H\left(p_{\rho_q}^A; q\right), \tag{5.9}$$

which provides us with an explicit interpretation for $\varepsilon_E^{info}(A|A')$:

In statistics the cross entropy $H(p;q)$ serves as a quality measure for the approximation of an unknown distribution $p$ by a known distribution $q$. Explicitly in the field of machine learning, the cross entropy is commonly used as objective function for the optimization of models [DBKMR05]. In our case $\left(p_{\rho_q}^A\right)_i$ describes the conditional probability for having the input state $\phi_i$, given that the device $\mathtt{A}'$ produced the guess $q$. Here, $H\left(p_{\rho_q}^A; q\right)$ measures if the guess $q$ is a good model for the actual distribution $p_{\rho_q}^A$. The measure $\mu(\omega)$ is independent of $\phi_i$ and corresponds to the probability that a particular $q \in \omega$ is produced as guess. Therefore, $\varepsilon_E^{info}(A|\mathtt{A}')$ can be seen as the expected quality of this approximation.

Alternatively, the error $\varepsilon_E^{info}(A|\mathtt{A}')$ can be expressed in terms of the relative entropy. By the use of

$$D(p||q) = H(p;q) - H(p),$$

in (5.9) we get

$$\varepsilon_E^{info}(A|\mathtt{A}') = \int_{\mathcal{P}_n} \mu[dq] D\left(p_{\rho_q}^A \middle\| q\right) + \int_{\mathcal{P}_n} \mu[dq] H\left(A|\rho_q\right). \tag{5.10}$$

Here we can omit the first term and get the estimate

$$\varepsilon_E^{info}(A|\mathtt{A}') \geq \int_{\mathcal{P}_n} \mu[dq] H\left(A|\rho_q\right). \tag{5.11}$$

At the end of this chapter we will see that the optimal devices for a measurement uncertainty relation always attain tightness in this estimate.

**A second proof of Cor. 5.4:** As an alternative to the use of Thm. 3.4, the corollary Cor. 5.4 can be proven as a conclusion from the following observations:

For a collection of measurements $A = A_1, \dots, A_m$, a joint measurement $\mathtt{R}$ corresponds to a measure $\mathtt{R}\,[dq^1 \wedge \dots \wedge dq^m]$, that will give us an ensemble of states

$$\rho_{q^1,\dots,q^m} := \frac{\mathtt{R}\,[dq^1 \wedge \dots \wedge dq^m]}{\operatorname{tr}\left(\mathtt{R}\,[dq^1 \wedge \dots \wedge dq^m]\right)}$$

distributed by the measure

$$\mu\left[dq^1 \wedge \dots \wedge dq^m\right] = \operatorname{tr}(\mathtt{R}\left[dq^1 \wedge \dots \wedge dq^m\right])/d$$

The states corresponding to the marginal measurements $\mathtt{A}'_j$ are therefore given by the mixture

$$\begin{aligned} \rho_{q^j} &= \frac{R\,[\mathcal{P}_n \wedge \cdots dq^j \wedge \dots \wedge \mathcal{P}_n]}{\operatorname{tr}\left(R\,[\mathcal{P}_n \wedge \cdots dq^j \wedge \dots \wedge \mathcal{P}_n]\right)} \\ &= \int_{(\mathcal{P}_n)^{m-1}} \frac{\operatorname{tr}\left(R\,[dq^1 \wedge \dots \wedge dq^m]\right)}{\operatorname{tr}\left(R\,[\mathcal{P}_n \wedge \cdots dq^j \wedge \dots \wedge \mathcal{P}_n]\right)} \frac{R\,[dq^1 \wedge \cdots dq^j \wedge \dots \wedge dq^m]}{\operatorname{tr}\left(R\,[dq^1 \wedge \dots \wedge dq^m]\right)} \\ &= \int_{(\mathcal{P}_n)^{m-1}} \frac{\mu\,[dq^1 \wedge \dots \wedge dq^m]}{\mu\,[\mathcal{P}_n \wedge \cdots dq^j \wedge \dots \wedge \mathcal{P}_n]} \rho_{q^1,\dots,q^j} \end{aligned} \tag{5.12}$$

Hence, the linear uncertainty of some $R$, with respect to $A$ and weights $\vec{a}$, is given by

$$\sum_{j=1\dots m} a_j \varepsilon_E^{info}(A|A') = \sum_{j=1\dots m} a_j \int_{\mathcal{P}_n} \mu\left[\mathcal{P}_n \wedge \cdots dq^j \wedge \dots \wedge \mathcal{P}_n\right] H\left(p^A_{q^j}; q^j\right).$$

In the above we can use the estimate (5.11), the decomposition (5.12), and the concavity of the entropy, to get

$$\begin{aligned} \sum_{j=1\dots m} a_j \varepsilon_E^{info}(A|A') &\geq \sum_{j=1\dots m} a_j \int_{\mathcal{P}_n} \mu\left[\mathcal{P}_n \wedge \cdots dq^j \wedge \dots \wedge \mathcal{P}_n\right] H\left(A_j|\rho_{q^j}\right) \\ &\geq \sum_{j=1\dots m} a_j \int_{(\mathcal{P}_n)^m} \mu\left[dq^1 \wedge \dots \wedge dq^m\right] H\left(A_j|\rho_{q^1 \dots q^m}\right) \quad (5.13) \\ &\geq \inf_\rho \sum_{j=1\dots m} a_j H(A_j|\rho) \end{aligned}$$

which not only proves Cor. 5.4, but also gives us some new insight in the structure of this problem. Most interestingly, the integrand in (5.13) does not longer rely on the marginal observables $A'_j$, but rather directly on the outcomes of the joint measurement $R$. More precisely, the integral in (5.13) will not change if we regroup or reparametrize the joint outcome set $\mathcal{P}_n \times \dots \times \mathcal{P}_n$. In the next subsection we will build on this observation and provide a marginal independent approach to entropic measurement uncertainty.

### A marginal independent approach

From a more general perspective, the outcomes of a joint measurement $R$ can be seen as *raw data*, which is then post-processed (by reducing it to the marginals), in order to obtain an approximation for an ideal measurement. However, when testing a device $R$ by some test states $\phi_i$, it is intuitively clear that such a post-processing can *not* add new information on the desired outcome $i$ to the resulting data. More drastically spoken: in the best case we can hope that a post-processing does not delete parts of this information. This perspective suggests to drop this post-processing step and to work on the raw data directly. Furthermore, this raw data need not necessarily be an outcome on the joint set of outcomes of the ideal observables. It could be anything.

Hence, the motivation for the following subsection is to quantify the obtainable information about a set of incompatible ideal measurements, that is gained by performing a further measurement $R$. However, before starting this, we have to clarify what kind of information should be recovered by $R$. The way we take in this subsection was suggested in [AB16, BHOW14]: we take a classical random variable $X$ with outputs on $I$, in order to draw a random distribution of input test states $\phi_i$. Then the outcomes of $R$ are used to extract this information on $X$ from measuring this test ensemble. Later on, we repeat this with a second random variable $Y$ and test states $\psi_i$ corresponding to a second ideal observable $B$. Here we can compare the extracted information and start looking for an optimal device $R$ that realizes a trade off.

For the above setting, the common information theoretic quantity to judge the information on $X$, that is contained in the outcomes of $R$, is the *conditional information* of the joint distribution of inputs and outputs

$$H(X|R^X) = H(XR^X) - H(R^X),$$

where $R^X$ denotes the random variable corresponding to the outcomes of $R$. We note that, the above quantity strongly depends on the entropy of the test statistics $X$. A natural choice, as done in [AB16, BHOW14], is to assume a uniform distributed $X$, i.e. to assume a maximal entropy $H(X) = \log(d)$. A typical motivation for this, borrowed from information theory, is to consider a communication scenario where $X$ contains classical data that is encoded by the $\phi_i$'s and should be recovered by measuring $R$. If this classical data is optimally compressed, i.e. encoded at the Shannon limit, the according random variable $X$ will appear uniformly distributed.

Furthermore, we can motivate this choice by the entangled reference frame error: Here we prepare the input states by measuring one half of a maximally entangled state $\phi^+$. This will give us the same input ensemble as preparing a test ensemble of $\phi_i$'s by a uniform $X$.

We note that we have $H(X|R^X) \leq H(X)$ in general. Hence, the idea of a calibration error will fail in this setting, because any deterministic input has the entropy $H(X) = 0$. This is clear intuitively: if there is only one single input, there is no information to recover.

For the following let $k \in K$ be an outcome of the outcome set $K$ of $R$. Furthermore, let $X$ be distributed by some probability distribution $p$ on labels $i \in I$. The joint distribution for testing $(X, \{\phi_i\}_{i\in I})$ on $R$ is given by

$$p_{ik} = \frac{1}{d} \operatorname{tr}(R(k)\phi_i),$$

and the probability of obtaining the outcome $k$ is

$$p_k = \sum_{i\in I} p_{ik} = \operatorname{tr}\left(R(k) \sum_{i\in I} \frac{1}{d}\phi_i\right) = \frac{1}{d}\operatorname{tr}(R(k))$$

Within this notation the conditional entropy is given by

$$\begin{aligned} H(X|R^X) &= -\sum_{i\in I, k\in K} p_{ik}\log(p_{ik}) + \sum_{k\in K} p_k \log(p_k) \\ &= -\sum_{i\in I, k\in K} \frac{1}{d}\operatorname{tr}(R(k)\phi_i)\log\left(\frac{\operatorname{tr}(R(k)\phi_i)}{d}\right) + \frac{1}{d}\sum_{k\in K}\operatorname{tr}(R(k))\log\left(\frac{\operatorname{tr}(R(k))}{d}\right) \\ &= -\sum_{k\in K}\frac{1}{d}\operatorname{tr}(R(k))\sum_{i\in I}\frac{\operatorname{tr}(R(k)\phi_i)}{\operatorname{tr}(R(k))}\log\left(\frac{\operatorname{tr}(R(k)\phi_i)}{\operatorname{tr}(R(k))}\right) \end{aligned}$$

Here, we can again substitute the renormalized POVM elements as quantum states

$$\rho_k := \frac{\operatorname{tr}(R(k)\phi_i)}{\operatorname{tr}(R(k))}$$

and get

$$H(X|R^X) = \sum_{k\in K} p_k H\left(A|\rho_k\right) \tag{5.14}$$

For $K = \mathcal{P}_n$, the above expression is equal to the integral in (5.13), if we replace the sum by an integral. Hence, we found the generalization of this construction to arbitrary outcome sets $K$. However, we can also do the converse, i.e. convert a device $R$, with outcomes on $K$, into a device $\mathtt{A}'$, with outcomes in $\mathcal{P}_n$:

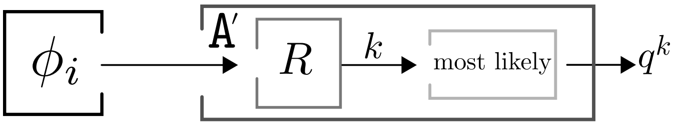


Figure 5.2: We can convert a device $R$, with outcomes on $K$, to a device $\mathtt{A}'$, with outcomes on $\mathcal{P}_n$, too.

Consider a fixed device $R$, which is tested with respect to an ideal sharp measurement $A$. We can assign a device $\mathtt{A}'$ by taking the conditional distribution, deduced from the $\rho_k$'s, as guesses, i.e. whenever we obtain a measurement outcome $k$, as response to an unknown input, we provide the guess

$$p^A_{\rho_k} = \frac{1}{\operatorname{tr}(R(k))} \left(\operatorname{tr}(R(k))\phi_1, \dots, \operatorname{tr}(R(k)\phi_n\right).$$

This procedure (see Fig. 5.2) gives us a device $\mathtt{A}'$, that is described by the POVM

$$\mathtt{A}'[dq] = \sum_{k\in K} R(k)\delta(q - p^A_{\rho_k})dq.$$

If we compute the entropic entanglement reference error of this device, the relative entropy in (5.10) vanishes by construction. Hence, we directly get (5.14), i.e.

$$\varepsilon^{info}_M(A|\mathtt{A}') = \sum_{k\in K} p_k H\left(A|\rho_k\right). \tag{5.15}$$

This construction also works if we consider more than one ideal sharp observable. Therefore, we get the following lemma:

**Lemma 5.6.** *Let $A_1, \dots, A_m$ be a collection of sharp measurements with outcomes on the alphabet $I$.*

*(i) Let $X_1 \dots X_m$ be a collection of uniform and independent distributed random variables on $I$. For any measurement $R$, with outcomes on a measurable set $K$, let $R^{X_1}, \dots, R^{X_m}$ denote the random variables corresponding to the outcomes of $R$, when tested on the respective eigenstates of $A_i$ with distribution $X_i$.*

*(ii) Let $\mathtt{R}$ be a joint measurement as in Cor. 5.4.*

*Let $\vec{a} = (a_1, \dots, a_m)$ be a vector of positive weights. Then*

$$\inf_R \sum_{j=1,\dots,m} a_j H(X_j|R^{X_j}) = \inf_{\mathtt{R}} \sum_{j=1\dots m} a_j \varepsilon^{info}_E(A|\mathtt{A}') \tag{5.16}$$

*Proof.* Since we consider arbitrary outcome sets $K$, the device $R$ comes from the bigger variation class, such that we get

$$\inf_R \sum_{j=1,\dots,n} a_j H(X_j|R^{X_j}) = \inf_{\mathtt{R}} \sum_{j=1\dots m} a_j \varepsilon_E^{info}(A|\mathtt{A}')$$

by (5.13). However, if we fixed a particular $R$, and consider a device $\mathtt{A}'$ with POVM elements

$$\mathtt{R}[dq^1 \wedge \cdots \wedge dq^n] = \sum_{k\in K} R(k)\delta(q^1 - p_{\rho_k}^{A_1}) \cdot \ldots \delta(q^n - p_{\rho_k}^{A_n}).$$

we get the desired equality in (5.16) by the same argumentation as in (5.15). □

Note that this lemma directly implies that linear measurement uncertainty relations in terms of the conditional entropy are bounded from below by entropic preparation uncertainty relations, as well. For completeness all linear uncertainty relations from this chapter and their interplay are organized in the following diagram:

$$\begin{tikzcd}
 & & \inf\limits_R \ \vec a\,\vec\varepsilon_E^{dm}(A|A') \arrow[dll, dash, "\geq"'] \arrow[drr, dash, "\geq"] \arrow[dddl, dash, "\leq"'] \arrow[dddr, dash, "\leq"] & & & \\
\inf\limits_R \ \vec a\vec\varepsilon_C^{dm}(A|A') \arrow[d, dash, "\leq"'] & & & & \inf\limits_\rho \ \vec a\vec\nu^{dm}(A|\rho) \arrow[r, dash, "\geq", "(m=2)"'] & 1 - c^* \arrow[d, dash, "\leq"] \\
\inf\limits_{\mathtt{R}} \ \vec a\vec\varepsilon_C^{info}(A|\mathtt{A}') \arrow[dr, dash, "\geq"'] & & & & \inf\limits_\rho \ \vec a\vec H(A|\rho) \arrow[r, dash, "\geq", "(m=2)"'] \arrow[dl, dash, "\geq"] & -\log(c^*) \\
 & \inf\limits_{\mathtt{R}} \ \vec a\vec\varepsilon_E^{info}(A|\mathtt{A}') \arrow[rr, dash, "="] & & \inf\limits_R \ \vec a\vec H(X|R^X) & &
\end{tikzcd}$$

Figure 5.3: The upper half of this diagram corresponds to the uncertainty measures based on the discrete metric and the lower one to entropic uncertainty measures. Equality from the left to the right can be realised by phase space observables, see the example in chapter Ch. 3. Equality from the top to the bottom can only be achieved if all uncertainties vanish.

# CHAPTER 6

## Entropic uncertainty relations

In this chapter we focus on the general structure of linear entropic preparation uncertainty relations in terms of the Shannon entropy. In the last chapter we saw that those uncertainty relations serve as lower bounds on both types of entropic measurement uncertainty relations. Beside this property, entropic preparation uncertainty relations are from interest by their own, because they are used as estimate all over quantum information theory (see the references in [S18] and [ASM$^+$15] for examples). However, there is only little known on their structure in the general case.

For the case of two sharp measurements, and equal weights, a couple of computable lower bounds on the optimal linear uncertainty relation are known (see the introducon in [ASM$^+$15]). The most prominent of them is the Maassen and Uffink bound [MU88]. In this chapter we will contribute to a better understanding of the structure of entropic uncertainty relations by [S18], placed in the next section. Here a proof of the additivity of entropic uncertainty relations is provided. The basic tool of this proof is to connect entropic uncertainty relations to so called $(p, q)$-*norms.*

In the second part of this chapter we will provide new numerical methods for computing linear entropic preparation uncertainty relations. These methods are based on an alternating minimization ansatz, which works fine for moderately small dimensions, but possibility fails in high dimension. Therefore, we will also comment on known results on the hardness of computing $(p, q)$-norms, which is closely connected to the hardness of this problem.

## 6.1 [S18]

*Additivity of entropic uncertainty relations*

- **Author:** René Schwonnek
- **Published in:** Quantum 2, 59 (2018)
- **DOI:** 10.22331/q-2018-03-30-59
- **Presented version:** The presented version is identical to arXiv:1801.04602v4, the literature is placed at the end of this thesis.
- **Main results:**
  - $(p, q)$ - norms are multiplicative for $p \leq q$
  - Linear entropic uncertainty relations for pairs of sharp measurements are additive.
  - Additivity fails for three measurements.
  - The Maassen and Uffink bound is generalized to linear relations with arbitrary weights.

# Additivity of entropic uncertainty relations

René Schwonnek

Institut für Theoretische Physik, Leibniz Universität Hannover, Germany

April 7, 2018

**We consider the uncertainty between two pairs of local projective measurements performed on a multipartite system. We show that the optimal bound in any linear uncertainty relation, formulated in terms of the Shannon entropy, is additive. This directly implies, against naive intuition, that the minimal entropic uncertainty can always be realized by fully separable states. Hence, in contradiction to proposals by other authors, no entanglement witness can be constructed solely by comparing the attainable uncertainties of entangled and separable states. However, our result gives rise to a huge simplification for computing global uncertainty bounds as they now can be deduced from local ones.**

**Furthermore, we provide the natural generalization of the Maassen and Uffink inequality for linear uncertainty relations with arbitrary positive coefficients.**

## Introduction

Uncertainty and entanglement are doubtless two of the most prominent and drastic properties that set apart quantum physics from a classical view on the world. Their interplay contains a rich structure, which is neither sufficiently understood nor fully discovered. In this work, we reveal a new aspect of this structure: the additivity of entropic uncertainty relations.

For product measurements in a multipartition, we show that the optimal bound $c_{ABC\cdots}$ in a linear uncertainty relation satisfies

$$c_{ABC\cdots} = c_A + c_B + c_C + \dots \quad , \tag{1}$$

where $c_A, c_B, c_C, \dots$ are bounds that only depend on local measurements. This result implies that minimal uncertainty for product measurements can always be realized by uncorrelated states. Hence, we have an example for a task which is not improved by the use of entanglement.

We will quantify the uncertainty of a measurement by the Shannon entropy of its outcome distribution. For this case, the corresponding linear uncertainty bound $c_{ABC\cdots}$ gives the central estimate in many applications like: entropic steering witnesses [1–4], uncertainty relations with side-information [5], some security proofs [6] and many more.

When speaking about uncertainty, we consider so called *preparation uncertainty relations* [7–14]. From an operational point of view, a preparation uncertainty describes fundamental limitations, i.e. a trade-off, on the certainty of predicting outcomes of several measurements that are performed on instances of the same state. This should not be confused [15] with its operational counterpart named measurement uncertainty[16–20]. A measurement uncertainty relation describes the ability of producing a measurement device which approximates several incompatible measurement devices in one shot.

The calculations in this work focus on uncertainty relations in a bipartite setting. However, all results can easily be generalized to a multipartite setting by an iteration of statements on bipartitions. The basic measurement setting, which we consider for bipartitions, is depicted in Fig. 1. We consider a pair of measurements, $X_{AB} = X_A X_B$ and $Y_{AB} = Y_A Y_B$, to which we will refer as the global measurements of (tensor) product form. Each of those global measurements of product form is implemented by applying local measurements at the respective sides of a bipartition between parties denoted by $A$ and $B$. Hereby,

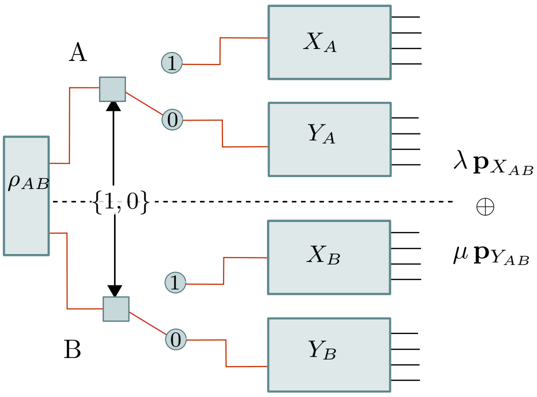


Figure 1: Basic setting of product measurements on a bipartition: pairs of measurements $X_A, X_B$ or $Y_A, Y_B$ are applied to a joint state $\rho_{AB}$ at the respective sides of a bipartition. One bit of information is transmitted for communicating whether the $X$ or the $Y$ measurements are performed. The weights $(\lambda, \mu)$ denote the probabilities corresponding to this choice.

the variables $X_A, X_B$ and $Y_A, Y_B$ will refer to those local measurements applied to the respective sides.

We only consider projective measurements, but beside this we impose no further restrictions on the individual measurements. So the only property that measurements like $X_A$ and $X_B$ have to share is the common label '$X$', besides this, they could be non-commuting or even defined on Hilbert spaces with different dimensions.

The main result of this work is stated in Prop.1 in Sec. 3. In that section, we also collect some remarks on possible and impossible generalizations and the construction of entanglement witnesses. The proof of Prop.1 is placed at the end of this paper, as it relies on two basic theorems stated in Sec.4 and Sec.5.

Thm.1, in Sec.4, clarifies and expands the known connection between the logarithm of $(p, q)$-norms and entropic uncertainty relations. As a special case of this theorem we obtain Lem.1 which states the natural generalization of the well known Maassen and Uffink bound [21] to weighted uncertainty relations. Thm.2, in Sec.5, states that $(p, q)$-norms, in a certain parameter range, are multiplicative, which at the end leads to the desired statement on the additivity of uncertainty relations.

Before stating the main result, we collect, in Sec.1, some general observations on the behavior of uncertainty relations for product measurements with respect to different classes of correlated states. Furthermore, in Sec. 2, we will motivate and explain the explicit form of linear uncertainty relations used in this work.

## 1 Uncertainty in bipartitions

All uncertainty relations considered is this paper are state-independent. In practice, finding a state-independent relation leads to the problem of jointly minimizing a tuple of given uncertainty measures, here the Shannon entropy of $X_{AB}$ and $Y_{AB}$, over all states. This minimum, or a lower bound on it, then gives the aforementioned trade-off, which then allows to formulate statements like: *"whenever the uncertainty of $X_{AB}$ is small, the uncertainty of $Y_{AB}$ has to be bigger than some state-independent constant"* .

Considering the measured state, $\rho_{AB}$, it is natural to distinguish between the three classes: uncorrelated, classically correlated and non-classical correlated. In regard of the uncertainty in a corresponding global measurement, states in these classes share some common features:

If the measured state is **uncorrelated**, i.e a product state $\rho_{AB} = \rho_A \otimes \rho_B$, the outcomes of the local measurements are uncorrelated as well. Hence, the uncertainty of a global measurement is completely determined by the uncertainty of the local measurements on the respective local states $\rho_A$ and $\rho_B$. Moreover, in our case, the additivity of the Shannon entropy, tells us that the uncertainty of a global measurement is simply the sum of the uncertainty of the local ones. In the same way any trade-off on the global uncertainties can be deduced from local ones.

If the measured state is **classically correlated**, i.e a convex combination of product states [22], additivity of local uncertainties does not longer hold. More generally, whenever we consider a concave uncertainty measure [23], like the Shannon entropy, the global uncertainty of a single global measurement is smaller than the sum of the local uncertainties. Intuitively this makes sense because a correlation allows to deduce information on the potential measurement outcomes of one side given a particular measurement outcome on the other. However, a linear uncertainty relation for a pair of global measurements is not affected by this, i.e a trade-off will again be saturated by product states. This is because the uncertainty relation between two measurements, restricted to some convex set of states, will always be attained on an extreme point of this set.

However, if measurements are applied to an **entangled state**, more precisely to a state which shows EPR-steering [24–26] with respect to the measurements $X_{AB}$ and $Y_{AB}$, it is in general not clear how a trade-off between global uncertainties relates to the corresponding trade-off between local ones. Just have in mind that steering implies the absence of any local state model, which is usually proven by showing that any such model would violate a local uncertainty relation.

In principle one would expect to obtain smaller uncertainty bounds by also considering entangled states, and there are many entanglement witnesses known based on this idea (see also Rem. 3 in the following section).

## 2 Linear uncertainty relations

We note that there are many uncertainty measures, most prominently variances [8, 10]. Variance, and similar constructed measures [17, 27], describe the deviation from a mean value, which clearly demands to assign a metric structure to the set of measurement outcomes. From a physicist's perspective this makes sense in many situations [11] but can also cause strange behaviours in situations where this metric structure has to be imposed artificially [28]. However, from the perspective of information theory, this seems to be an unnecessary dependency. Especially when uncertainties with respect to multipartitions are considered, it is not clear at all how such a metric should be constructed. Hence, it can be dropped and a quantity that only depends on probability distributions of measurement outcomes has to be used. We will use the Shannon entropy. It fulfills the above requirement, does not change when the labeling of the measurement outcomes are permuted, and has a clear operational interpretation [29, 30]. Remarkably, Claude Shannon himself used the term 'uncertainty' as an intuitive paraphrase for the quantity today known as 'entropy' [29]. Historically, the decision to call the Shannon entropy an 'entropy' goes back to a suggestion John von Neumann gave to Shannon, when he was visiting Weyl in 1940 (there are, at least, three versions of this anecdote known [31], the most popular is [32]).

Because we are not interested in assigning values to measurement outcomes, a measurement, say $X$, is sufficiently described by its POVM elements, $\{X_i\}$. So, given a state $\rho$, the probability of obtaining the $i$-th outcome is computed by $\mathrm{tr}(\rho X_i)$. The respective probability distribution of all outcomes is denoted by the vector $\mathbf{p}^x_\rho$. Within this notation the Shannon entropy of a $X$ measurement is given by $H(X|\rho) := -\sum_i \left(\mathbf{p}^x_\rho\right)_i \log \left(\mathbf{p}^x_\rho\right)_i$. As we restrict ourselves to non-degenerate projective measurements, all necessary information on a pair of measurements, $X$ and $Y$, is captured by a unitary $U$ that links the measurement basis. We will use the convention to write $U$ as transformation from the $\{X_i\}$ to the $\{Y_i\}$-basis, i.e. we will take $U$ such that $Y_i = UX_iU^\dagger$ holds.

Our basic objects of interest are optimal, state-independent and linear relations. This is, for fixed weights $\lambda, \mu \in \mathbb{R}^+$ we are interested in the best constant $c(\lambda, \mu)$ for which the linear inequality

$$\lambda H(X|\rho) + \mu H(Y|\rho) \geq c(\lambda, \mu) \qquad (2)$$

holds on *all* states $\rho$.

Such a relation has two common interpretations: On one hand one can consider a guessing game, see also [33]. On the other, a relation like (2) can be interpreted geometrically as in Fig. 2.

**Linear uncertainty: a guessing game**

For the moment, consider a player, called Eve, who plays against an opponent, called Alice. Dependent on a coin throw, in each round, Alice performs measurement $X_A$ or $Y_A$ on a local quantum state. Thereby the weights $\lambda$ and $\mu$ are the weights of the coin and the l.h.s. of (2) describes the total uncertainty Eve has on Alice's outcomes in each round. To be more precise, up to a $(\lambda, \mu)$-dependent constant, the l.h.s of (2) equals the Shannon entropy of the outcome distribution $\lambda \mathbf{p}^{X_A}_\rho \oplus \mu \mathbf{p}^{Y_A}_\rho$.

Eve's role in this game is to first choose a state $\rho$, observe the coin throw, wait for the measurements to be performed by Alice, and then ask binary questions to her opponent in order to get certainty on the outcomes. Thereby, the Shannon entropy sum on the l.h.s of (2) (with logarithm to the base 2) equals the expected amount questions Eve has to ask using an optimal strategy based on a fixed $\rho$. Hence, the value $c(\lambda, \mu)$ denotes the minimal amount of expected questions, attainable by choosing an optimal $\rho$.

For a bipartite setting, Fig. 1, a second player, say Bob, joins the game. Here, Eve will play the above game against Alice and Bob, simultaneously. Thereby, Alice and Bob share a common coin, and, therefore, apply measurements with the same labels ($X_{AB}$ or $Y_{AB}$). The obvious question that arises in this context is if Eve gets an advantage in this simultaneous game by using an entangled state or not. Prop. 1 in the next section answers the above question negatively, which is somehow unexpected as in principle the possible usage of non-classical correlations enlarges Eve's strategies. For example: Eve could have used a maximally entangled state, adjusted such that *all* measurements Alice and Bob perform are maximally correlated. In this case the remaining uncertainty Eve has, would only be the uncertainty on the outcomes of one of the parties. However, the marginals of a maximally entangled state are maximally mixed. Hence, Eve still has a serious amount of uncertainty ($\log d$), which turns out to be not small enough for beating a strategy based on minimizing the uncertainty of the local measurements individually. For the case of product-MUBs in prime square dimension [34], it turns out that the minimal uncertainty realizable by a maximally entangled state actually equals the optimal bound.

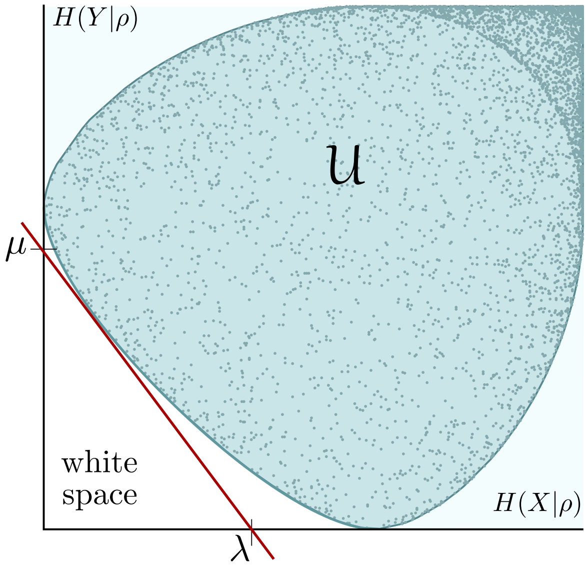


Figure 2: Uncertainty set for measurements performed on a qubit. Any linear uncertainty relation, (2), with weights $(\lambda, \mu)$, gives the description of a tangent to the uncertainty set. All attainable pairs of entropies lie above this tangent.

**Linear uncertainty: the positive convex hull**

The second interpretation comes from considering the set of all attainable uncertainty pairs, the so called uncertainty set

$$\mathcal{U} = \{(H(X|\rho), H(Y|\rho))\,|\rho \textit{ is a quantum state}\}. \quad (3)$$

In principle this set contains all information on the uncertainty trade-off between two measurements. More precisely, the white space in the lower-left corner of a diagram like Fig. 2 indicates that both uncertainties cannot be small simultaneously. In this context, a state-independent uncertainty gives a quantitative description of this white space. Unfortunately, it turns out that computing $\mathcal{U}$ can be very hard, because the whole state-space has to be considered. Here a linear inequality, like (2), gives an outer approximation of this set. More precisely, if $c(\lambda, \mu)$ is the optimal constant in (2), this inequality describes a halfspace bounded from the lower-left by a tangent on $\mathcal{U}$. This tangent has the slope $\mu/\lambda$. The points on which this tangent touches the boundary of $\mathcal{U}$ corresponds to states which realize equality in (2). Those states are called minimal-uncertainty states. Given all those tangents, i.e. $c(\lambda, \mu)$ for all positive $(\lambda, \mu)$, we can intersect all corresponding halfspaces and get a convex set which we call the *positive convex hull* of $\mathcal{U}$, denoted by $\overline{\mathcal{U}}$ in the following. Geometrically, the positive convex hull can be constructed by taking the convex hull of $\mathcal{U}$ and adding to it all points that have bigger uncertainties then, at least, some point in $\mathcal{U}$.

If $\mathcal{U}$ is convex, like in the example above, $\overline{\mathcal{U}}$ contains the full information on the relevant parts of $\mathcal{U}$. If $\mathcal{U}$ is not convex, $\overline{\mathcal{U}}$ still gives a variety of state independent uncertainty relations, but there is still place for finding improvements, see [34].

## 3 Additivity, implications and applications

We are now able to state our main result

**Proposition 1** (Additivity of linear uncertainty relations)**.** *Let $c_A(\lambda, \mu)$ and $c_B(\lambda, \mu)$ be state-independent lower bounds on the linear entropic uncertainty for local measurements $X_A, X_B$ and $Y_A, Y_B$, with weights $(\lambda, \mu)$. This means we have that*

$$\begin{aligned} \lambda H(X_A|\rho_A) + \mu H(Y_A|\rho_A) &\geq c_A(\lambda, \mu) \\ \lambda H(X_B|\rho_B) + \mu H(Y_B|\rho_B) &\geq c_B(\lambda, \mu) \end{aligned} \tag{4}$$

*holds on any state $\rho_A$ from $\mathcal{B}(\mathcal{H}_A)$ and $\rho_B$ from $\mathcal{B}(\mathcal{H}_B)$. Let $X_{AB}$ and $Y_{AB}$ be the joint global measurements that arise from locally performing $X_A, X_B$ and $Y_A, Y_B$ respectively. Then*

$$\lambda H(X_{AB}|\rho_{AB}) + \mu H(Y_{AB}|\rho_{AB}) \geq c_A(\lambda, \mu) + c_B(\lambda, \mu) \tag{5}$$

*holds for all states $\rho_{AB}$ from $\mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$. Furthermore, if $c_A$ and $c_B$ are optimal bounds, then*

$$c_{AB}(\lambda, \mu) := c_A(\lambda, \mu) + c_B(\lambda, \mu) \tag{6}$$

*is the optimal bound in (5), i.e. linear entropic uncertainty relations are additive.*

The proof of this proposition is placed at the end of Sec. 5. We will proceed this section by collecting some remarks related to the above proposition:

**Remark 1** (Product states)**.** Assume that $c_A(\lambda, \mu)$ and $c_B(\lambda, \mu)$ are optimal constants, and $\phi_A$ and $\phi_B$ are the states that saturate the corresponding uncertainty relations (4). Then the product state $\phi_{AB} := \phi_A \otimes \phi_B$ saturates (5), due to the additivity of the Shannon-entropy. However, this does not imply that all states that saturate (4) have to be product states. Examples for this, involving MUBs of product form, are provided in [34].

**Remark 2** (Minkowski sums of uncertainty regions)**.** Prop. 1 shows how the uncertainty set $\mathcal{U}_{AB}$, of the product measurement, relates to the uncertainty sets $\mathcal{U}_A$ and $\mathcal{U}_B$ of corresponding local measurements: For the case of an optimal $c_{AB}(\lambda, \mu)$, and fixed $(\lambda, \mu)$, equality in (5) can always be realized by product states (see Rem. 1). In an uncertainty diagram, like Fig. 3, those states correspond to points on the lower-left boundary of an uncertainty set, and, in general, they produce the finite extreme points of the positive convex hull of an uncertainty set.

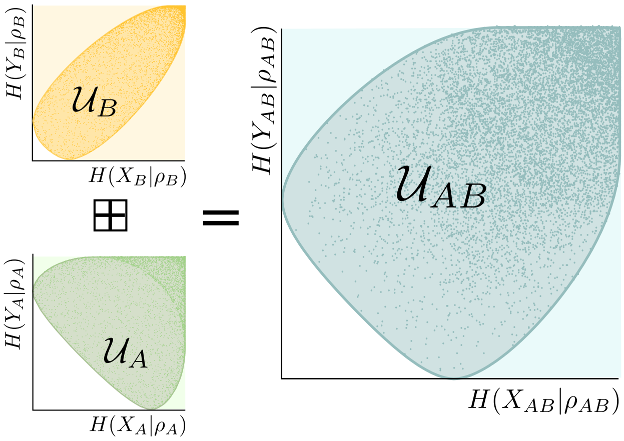


Figure 3: Uncertainty sets of local measurements can be combined by the Minkowski sum: Uncertainty sets (green and yellow) for two pairs of local measurements on Qubits and the uncertainty set of the corresponding global measurements (blue).

For product states we have the additivity of the Shannon entropy, which gives

$$\begin{pmatrix} H(X_{AB}|\phi_A \otimes \phi_B) \\ H(Y_{AB}|\phi_A \otimes \phi_B) \end{pmatrix} = \begin{pmatrix} H(X_A|\phi_A) \\ H(Y_A|\phi_A) \end{pmatrix} + \begin{pmatrix} H(X_B|\phi_B) \\ H(Y_B|\phi_B) \end{pmatrix} \tag{7}$$

This implies that we can get every extreme point of $\overline{\mathcal{U}}_{AB}$ by taking the sum of two extreme points of $\overline{\mathcal{U}}_A$ and $\overline{\mathcal{U}}_B$. Due to convexity the same holds for all points in $\overline{\mathcal{U}}_{AB}$ and we can get this set as Minkowski sum [35].

$$\overline{\mathcal{U}}_{AB} = \overline{\mathcal{U}}_{AB} \boxplus \overline{\mathcal{U}}_B \tag{8}$$

For convex uncertainty regions, arising from local measurements, this is depicted in Fig. 3. For this example, it is also true that $\mathcal{U}_{AB}$ itself is given as

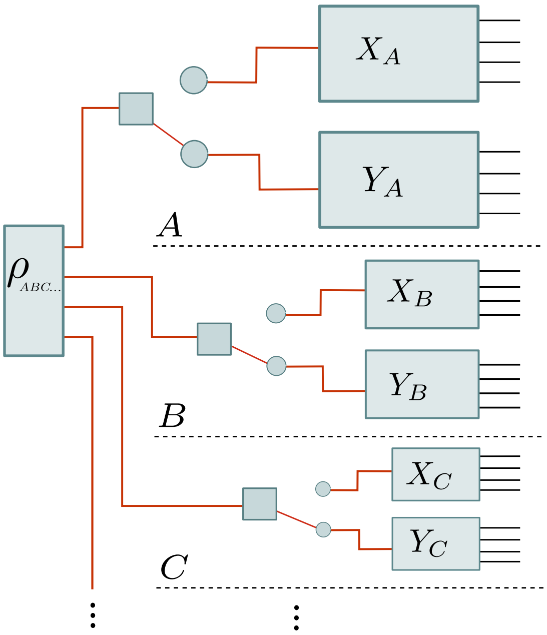


Figure 4: Multiparite setting: Additivity of entropic uncertainty relations also holds if a pair of global product measurements for many local parties is considered.

Minkowski sum of local uncertainty sets. However, we have to note, this behavior cannot be concluded from Prop. 1 alone.

**Remark 3** (Relation to existing entanglement witnesses)**.** A well know method for constructing nonlinear entanglement witnesses is based on computing the minimal value of a functional, like the sum of uncertainties [36–38], attainable on separable states. Given an unknown quantum state, the value of this functional is measured. If the measured value undergoes the limit set by separable states, the presence of entanglement is witnessed. For uncertainty relations based on the sum of general Schur concave functionals this method was proposed in [4], including Shannon entropy, i.e. the l.h.s. of (5), as central example.

Our result Prop. 1 shows that this method will not work for Shannon entropies, because there is no entangled state that undergoes the limit set by separable states. We note that there is no mathematical contradiction between Prop. 1 and [4]. We only show that the set of examples for the method proposed in [4] is empty.

For uncertainty relations in terms of Shannon, Tsallis and Renyi entropies a similar procedure for constructing witnesses was proposed by [37, 39]. Here explicit examples for states, that can be witnessed to be entangled, were provided. Again, our proposition Prop. 1 is not in contradiction to this work because in [37, 39] observables with a non-local degeneracy where considered.

Prop. 1 can easily be generalized to a multipartite setting, see Fig. 4 :

**Corollary 1** (Generalization to multipartite measurements)**.** Assume parties $A_1 \ldots A_n$ that locally perform measurements, $X_{A_1}, \ldots, X_{A_n}$ or $Y_{A_1}, \ldots, Y_{A_n}$, with weights $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$. In analogy to (4), let $c_{A_1}(\vec{\lambda}), \ldots, c_{A_n}(\vec{\lambda})$ denote optimal local bounds and let $c_{A_1 \cdots A_n}(\vec{\lambda})$ be the optimal bound corresponding to product measurements $X_{A_1 \ldots A_n}$ and $Y_{A_1 \ldots A_n}$. We have

$$c_{A_1 \ldots A_n}(\vec{\lambda}) = \sum_{i=1}^{n} c_{A_i}(\vec{\lambda}) \tag{9}$$

This follows by iterating (6).

**Remark 4** (Generalization to three measurements)**.** The generalization of Prop. 1 to three measurements, say $X_{AB}$, $Y_{AB}$ and $Z_{AB}$, fails in general. The following counterexample was provided by O. Gühne [40]: For both parties we consider local measurements deduced from the three Pauli operators on a qubit and take all weights equal to one. In short hand notation we write $X_{AB} = \sigma_X \otimes \sigma_X$, $Y_{AB} = \sigma_Y \otimes \sigma_Y$, and $Z_{AB} = \sigma_Z \otimes \sigma_Z$. In this case, the minimal local uncertainty sum is attained on eigenstates of the Pauli operators. If such a state is measured, the entropy for one of the measurements is zero and maximal for the others. Hence, the local uncertainty sum is always bigger than 2 [bit]. Therefore we have

$$\begin{aligned} &H\left(\sigma_X \otimes \sigma_X | \phi_A \otimes \phi_B\right) + \\ &H\left(\sigma_Y \otimes \sigma_Y | \phi_A \otimes \phi_B\right) + \\ &H\left(\sigma_Z \otimes \sigma_Z | \phi_A \otimes \phi_B\right) \geq 4 \end{aligned} \tag{10}$$

for all product states. In contrast to this a Bell state, say $\Psi^-$, will give the entropy of 1[bit], for all above measurements. Hence we have,

$$\begin{aligned} &H\left(\sigma_X \otimes \sigma_X | \Psi^-\right) + \\ &H\left(\sigma_Y \otimes \sigma_Y | \Psi^-\right) + \\ &H\left(\sigma_Z \otimes \sigma_Z | \Psi^-\right) = 3 \ngeq 4. \end{aligned} \tag{11}$$

## 4 Lower bounds from $(p, q)$-norms

The quite standard technique for analyzing a linear uncertainty relation is to connect it to the $(p, q)$-*norm* (see (12) below) of the basis transformation $U$. Thereby, the majority of previous works in this field is concentrating only on handling the case of equal weights $\lambda = \mu = 1$, which is connected to the $(p, q)$-norm for the case $1/p + 1/q = 1$. However, for the purpose of this work, i.e. for proving Prop. 1, we have to extend this connection to arbitrary $(\lambda, \mu)$. We will do this by Thm. 1 on the next page.

A historically important example for the use of the connection between $(p, q)$-norms and entropic uncertainties, is provided by Bialynicki-Birula and Mycielski [41]. They used Beckner's result [42], who computed the $(p, q)$-*norm* of the Fourier-Transfromation, for proving the corresponding uncertainty relation, between position and momentum, conjectured by Hirschmann [43]. Also Maassen and Uffink [21] took this way for proving their famous relation. Our result gives a direct generalization of this, meaning we will recover the Maassen and Uffink relation at the end of this section as special case of (50). Albeit, before stating our result, we will start this section by shortly reviewing the previously known way for connecting $(p,q)$-norms with linear uncertainty relation, see also [44, 45] for further details:

The $(p,q)$-norm, i.e the $l^p \to l^q$ operator norm, of a basis transformation $U$ is given by

$$\|U\|_{q,p} := \sup_{\phi\in\mathcal{H}} \frac{\|U\phi\|_q}{\|\phi\|_p}. \tag{12}$$

Here, the limit of $\|U\|_{q,p}$ for $(p,q) \to (2,2)$ goes to 1. However, when $p$ and $q$ are fixed on the curve $1/p + 1/q = 1$, the leading order of $\|U\|_{q,p}$ around $(p,q) = (2,2)$ recovers the uncertainty relation (2) in the case of equal weights $\lambda = \mu = 1/2$, see [41, 43].

More precisely, taking the negative logarithm of (12) gives

$$-\log\|U\|_{q,p} = \inf_{\phi\in\mathcal{H}} \log\|\phi\|_p - \log\|U\phi\|_q\,. \tag{13}$$

Here, we can identify the squared modulus of the components of $\phi$ as probabilities of the $X$ and $Y$ measurement outcomes

$$\begin{aligned}|(\phi)_i|^2 &= \langle\phi|\,X_i\,|\phi\rangle = (\mathbf{p}^X_\phi)_i\\ |(U\phi)_i|^2 &= \langle\phi|\,Y_i\,|\phi\rangle\ = (\mathbf{p}^Y_\phi)_i\end{aligned} \tag{14}$$

and substitute

$$\|\phi\|_p = \left(\|\mathbf{p}^X_\phi\|_{p/2}\right)^2 \quad\text{and}\quad \|U\phi\|_q = \left(\|\mathbf{p}^Y_\phi\|_{q/2}\right)^2\,. \tag{15}$$

By this, (13) gives a linear relation in terms of the $\alpha$-Renyi entropy [46], $H_\alpha(\mathbf{p}) = \frac{\alpha}{1-\alpha}\log(\|\mathbf{p}\|_\alpha)$. Here we get

$$\inf_{\phi\in\mathcal{H}} \frac{2-p}{p} H_{p/2}(X|\phi) - \frac{2-q}{q} H_{q/2}(Y|\phi) = -\log\|U\|^2_{q,p}\,. \tag{16}$$

If we evaluate this on the curve $1/p + 1/q = 1$, for $p \leq 2 \leq q$, we can use

$$\frac{2-p}{p} = \frac{1}{p} - \frac{1}{q} = \frac{q-2}{q}, \tag{17}$$

which can be employed to (16), in order to get

$$\inf_{\phi\in\mathcal{H}} H_{p/2}(X|\phi) + H_{q/2}(Y|\phi) = \left(\frac{1}{q} - \frac{1}{p}\right)^{-1} \log\|U\|^2_{q,p}\,. \tag{18}$$

Here, the limit $(p,q) \to (2,2)$, in the l.h.s of (18), gives the limit from the Renyi to the Shannon entropy. This gives the l.h.s. of the uncertainty relation (2) for $\lambda = \mu = 1$. Hence, the functional dependence of $\|U\|_{q,p}$ on $(p,q)$ in the limit $(p,q) \to (2,2)$ gives the optimal bound $c(1,1)$, in (2). For the case of the $\mathcal{L}^2(\mathbb{R})$-Fourier transformation the norm $\|U_\mathcal{F}\|_{q,p} = \sqrt{p^{1/p}}/\sqrt{q^{1/q}}$ was computed by Beckner [42], leading to $c(1,1) = \log(\pi \mathrm{e})$. However, to the best of our knowledge, computing $\|U\|_{q,p}$, for general $U$ and $(p,q)$, is an outstanding problem, and presumably very hard [47, 48]. Albeit, for special choices of $(p,q)$ this problem gets treatable, see [49] for a list of those. The known cases include $p = q = 2$, $p = \infty$ or $q = \infty$ such as $p = 1$ or $q = 1$.

The central idea of Maassen's and Uffink's work [21] is to show that the easy case of $(p = 1, q = \infty)$, here we have $\|U\|_{1,\infty} = \max_{ij}|U_{ij}|$, gives a lower bound on $c(1,1)$. More precisely, they show that, for $1 \leq p \leq 2$ and on the line $1/p + 1/q = 1$, the r.h.s. of (18) approaches $c(1,1)$ from below. Note that this is far from being obvious. Explicitly, for $p \leq 2 \leq q$ we have $H_{q/2}(Y|\phi) \geq H(Y|\phi)$ and $H_{p/2}(X|\phi) \leq H(X|\phi)$, so one term approaches the limit from above and the other approaches the limit from below. Whereas Maassen and Uffink showed, using the Riesz-Thorin interpolation [50, 51], that the $\inf_\phi$ of the sum of both approaches the limit from below.

The following Theorem, Thm.1, extends the above to the case of arbitrary $(\lambda,\mu)$. Notably, we have to take $(p,q)$ from curves with $1/p + 1/q \neq 1$, those are depicted in Fig. 5. In contrast to Maassen and Uffink, the central inequality we use is the $\infty$-norm versions of the Golden Thompson inequality (see [52–54] and the blog of T.Tao [55] for a proof and related discussions).

**Theorem 1.** *Let $c(\lambda,\mu)$, with $\lambda,\mu \in \mathbb{R}_+$, be the optimal constant in the linear weighted entropic uncertainty relation*

$$c(\lambda,\mu) := \inf_\rho \lambda H\left(X|\rho\right) + \mu H\left(Y|\rho\right). \tag{19}$$

*Then:*

*(i) $c(\lambda,\mu)$ is bounded from below by $-N\log\left(\omega_N(\lambda,\mu)\right)$*

$$\textit{with}\quad \omega_N(\lambda,\mu) = \sup_{\substack{\mathbf{x}\in B_r(\mathbb{C}^d)\\ \mathbf{y}\in B_s(\mathbb{C}^d)}} \left|\mathbf{x}^\dagger U\mathbf{y}\right| \tag{20}$$

$$\textit{and}\quad r = \frac{2N}{N+2\lambda} \qquad s = \frac{2N}{N+2\mu} \tag{21}$$

*where*

$$B_r(\Omega) := \{\mathbf{x}\in\Omega|\ 1 \geq \|\mathbf{x}\|_r\}$$

*denotes the unit $r$-norm Ball on $\Omega$.*

*(ii) For $\lambda,\mu \leq N/2$ we can write*

$$\omega_N(\lambda,\mu) = \sup_{\phi\in\mathbb{C}^d} \frac{\|U\phi\|_{r'}}{\|\phi\|_s} = \sup_{\phi\in\mathbb{C}^d} \frac{\|U\phi\|_{s'}}{\|\phi\|_r} \tag{22}$$

$$\textit{with}\ r' = \frac{2N}{N-2\lambda} \qquad s' = \frac{2N}{N-2\mu} \tag{23}$$

$$\tag{24}$$

*(iii) For $\mu, \lambda \in \mathbb{R}^+\backslash\{0\}$, we have*

$$c(\lambda,\mu) = \lim_{N\to\infty} -N\log\left(\omega_N(\lambda,\mu)\right) \tag{25}$$

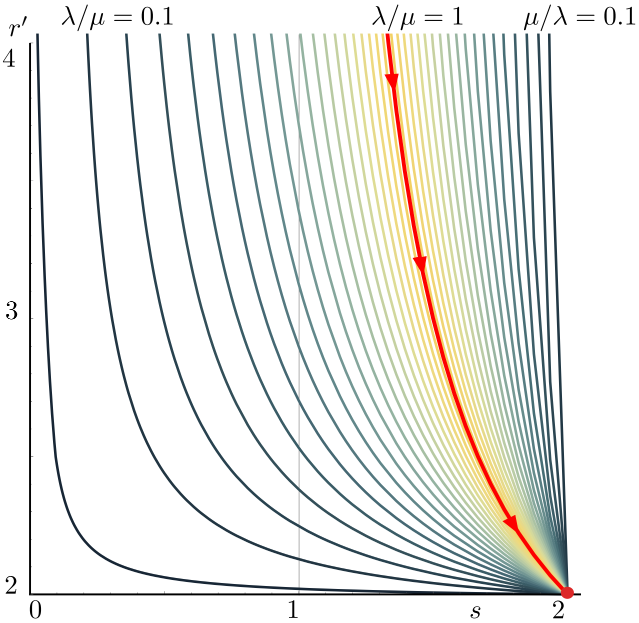


Figure 5: Evaluating $\|U\|_{r',s}$ on the depicted curves gives a lower bound for $c(\lambda,\mu)$, (see Thm. 1). Because $c(\lambda,\mu)$ is a linear bound it is 1-homogenious in $(\lambda,\mu)$. Hence all information on the optimal bound $c(\lambda,\mu)$ can be recovered by knowing it for any fixed ratio $\lambda/\mu$. The thick red curve corresponds to the case $1/r'+1/s=1$ which gives bounds $c(1,1)$ from below. For $s=1$ the norm $\|U\|_{r',s=1}$ can be computed analytically, this gives a generalization of the Masssen and Uffink bound (see Lem. 1).

*Proof.* The starting point of this proof is a modification of a technique, used by Frank and Lieb in [56], for reproving the Maassen and Uffink bound (see also the talk of Hans Maassen [44], for a finite dimensional version).

For probability distributions $\mathbf{p},\mathbf{q} \in B_1(\mathbb{R}^d_+)$ we define the operators

$$A(\mathbf{p}) := -\sum X_i \log(p_i) \text{ and } B(\mathbf{q}) := -\sum Y_i \log(q_i) \tag{26}$$

such that we can rewrite the Shannon entropy as

$$H\left(X|\rho\right) = \mathrm{tr}(\rho A(p^X_\rho)) \text{ and } H\left(Y|\rho\right) = \mathrm{tr}(\rho B(p^Y_\rho)) \tag{27}$$

Based on this, we can further rewrite the Shannon entropy as an optimization over a linear function in $\rho$ by using the positivity of the relative entropy, i.e. we have $D(\mathbf{p}||\mathbf{q}) = \sum p_i \log(p_i) - \sum p_i \log(q_i) \geq 0$, which implies $-\sum p_i \log(q_i) \geq H(\mathbf{p})$. We obtain

$$H\left(X|\rho\right) = \inf_{\mathbf{p}} \mathrm{tr}(\rho A(\mathbf{p})), \tag{28}$$

such as the respective statement for $H\left(Y|\rho\right)$ and $B(\mathbf{q})$. If we employ this rewriting to $c(\lambda,\mu)$, we obtain the minimal entropy sum as a minimization over a parametrized eigenvalue problem, namely

$$\begin{aligned} c(\lambda,\mu) &= \inf_\rho \lambda H\left(X|\rho\right) + \mu H\left(Y|\rho\right) \\ &= \inf_{\mathbf{p},\mathbf{q},\rho} \mathrm{tr}\left(\rho\left(\lambda A(\mathbf{p}) + \mu B(\mathbf{q})\right)\right) \end{aligned} \tag{29}$$

Now we will turn the minimization, over $\rho$, into a maximization by applying the convex function $e^{-x/N}$, with $N \geq 1$, to the weighted sum of $A$ and $B$. This will map the smallest eigenvalue of $\lambda A + \mu B$ to the largest of $e^{-\frac{\lambda A(p)+\mu B(q)}{N}}$ and so on. In order to get back the correct value of $c$ we will have to apply the inverse function, $-N\log(x)$, afterwards. We get

$$c(\lambda,\mu) = -N\log\left(\sup_{\mathbf{p},\mathbf{q},\rho} \mathrm{tr}\left(\rho e^{-\frac{\lambda A(p)+\mu B(q)}{N}}\right)\right). \tag{30}$$

Due to the positivity of the operator exponential, i.e. $A$ and $B$ are hermitian, the optimization over $\rho$ is equivalent to the Schatten-$\infty$ norm. We have

$$c(\lambda,\mu) = -N\log\left(\sup_{\mathbf{p},\mathbf{q}} \left\| e^{-\frac{\lambda A(p)+\mu B(q)}{N}} \right\|_\infty\right) \tag{31}$$

At this point we apply the Golden-Thompson inequality

$$\|e^{S+T}\|_p \leq \|e^S e^T\|_p \tag{32}$$

and expand the resulting exponentials, as well as the Schatten norm. We get

$$\begin{aligned} c(\lambda,\mu) &\geq \text{-}N\log\left(\sup_{\mathbf{p},\mathbf{q}} \left\| e^{-\frac{\lambda A(p)}{N}} e^{-\frac{\mu B(q)}{N}} \right\|_\infty\right) & (33) \\ &= \text{-}N\log\left(\sup_{\mathbf{p},\mathbf{q}} \left\| \sum_{ij} X_i p_i^{\lambda/N} Y_j q_j^{\mu/N} \right\|_\infty\right) & (34) \\ &= \text{-}N\log\left(\sup_{\substack{\mathbf{p},\mathbf{q} \\ |x\rangle,|y\rangle}} \langle x|\sum_{ij} X_i p_i^{\lambda/N} Y_j q_j^{\mu/N}|y\rangle\right) & \\ & & (35) \end{aligned}$$

Now we substitute $p_i^{\lambda/N} =: \chi_i$ and $q_j^{\mu/N} =: \xi_j$, and expand $|x\rangle = \sum x_i |e_i\rangle$ and $|y\rangle = \sum y_j |f_j\rangle$, with component vectors $\mathbf{x},\mathbf{y} \in B_2(\mathbb{C}^d)$. By this the r.h.s of (35) becomes

$$-N\log\left(\sup_{\chi,\mathbf{x}} \sup_{\xi,\mathbf{y}} \left| \sum_{ij} \chi_i x_i \langle e_i|f_j\rangle \xi_j y_j \right|\right). \tag{36}$$

Here we can identify $\langle e_i|f_j\rangle = U_{ij}$, i.e. the overlaps are the components of $U$ when represented in the basis $X$. At this point, it is straightforward to check that $\chi \in B_{N/\lambda}(\mathbb{R}^d_+)$ and $\xi \in B_{N/\mu}(\mathbb{R}^d_+)$. Using the generalized Hölder inequality we can fuse some of the maximizations above as follows: On one hand, we have

$$\Big(\sum |\chi_i x_i|^r\Big)^{\frac{1}{r}} \leq \|\mathbf{x}\|_2 \|\chi\|_{N/\lambda} \leq 1 \tag{37}$$

$$\text{and} \quad \Big(\sum |\xi_j y_j|^s\Big)^{\frac{1}{s}} \leq \|\mathbf{y}\|_2 \|\xi\|_{N/\mu} \leq 1$$

$$\text{for} \quad \frac{1}{r} = \frac{1}{2} + \frac{\lambda}{N} \quad \text{and} \quad \frac{1}{s} = \frac{1}{2} + \frac{\mu}{N} \quad , \tag{38}$$

which means that the vectors $\mathbf{v}$ and $\mathbf{w}$, with $v_i = \chi_i x_i$ and $w_j = \xi_j y_j$, are in $B_r(\mathbb{C})$ and $B_s(\mathbb{C})$ respectively.

On the other hand, the converse is also true, i.e. every $\mathbf{v}$ and $\mathbf{w}$ from $B_r(\mathbb{C}^d)$ and $B_s(\mathbb{C}^d)$ can be realized by suitable choices of $\mathbf{x}, \chi$ and $\mathbf{y}, \xi$. For example, we can always set

$$x_i = |v_i|^{r\lambda/N} \text{ and } \chi_i = |v_i|^{2/r} e^{i \arg(v_i)} \tag{39}$$

for getting $\mathbf{x}$ and $\chi$ from $\mathbf{v}$, componentwise. For this particular choice we can check that

$$\begin{aligned} x_i \chi_i &= |v_i|^{r(\lambda/N+1/2)} e^{i \arg(v_i)} \\ &= |v_i|^{r/r} e^{i \arg(v_i)} = v_i \end{aligned} \tag{40}$$

holds, such that we will get back $\mathbf{v}$. Furthermore $\mathbf{x} \in B_{N/\lambda}(\mathbb{R}^d_+)$ and $\chi \in B_2(\mathbb{C}^d)$ follows by writing out

$$\sum_i x_i^{N/\lambda} = \sum_i v_i^r \leq 1 \text{ and } \sum_i \chi_i^2 = \sum_i v_i^r \leq 1. \tag{41}$$

If we use the above in (36), we can replace $\sup_{x,\chi}$ by $\sup_{\mathbf{v}}$ and $\sup_{y,\xi}$ by $\sup_{\mathbf{w}}$, in order to get the statement (i) with

$$\omega_N := \sup_{\substack{\mathbf{v} \in B_r(\mathbb{C}^d) \\ \mathbf{w} \in B_s(\mathbb{C}^d)}} \left|\mathbf{v}^\dagger U \mathbf{w}\right| \quad . \tag{42}$$

For showing the statement (ii), we take $r'$, with $1 = 1/r + 1/r'$. If $\lambda \leq N/2$ holds we have $r' \geq 0$ and we can use the tightness of the Hölder inequality to rewrite

$$\sup_{\mathbf{v} \in B_r} \left|\mathbf{v}^\dagger U \mathbf{w}\right| = \|U\mathbf{w}\|_{r'} \,, \tag{43}$$

i.e. the maximization over $B_r$ gives the dual norm of $r$. Substituting $\mathbf{w}$ by $\phi = \mathbf{w}\|\phi\|_s$ then gives

$$\omega_N = \sup_{\phi \in \mathbb{C}^d} \frac{\|U\phi\|_{r'}}{\|\phi\|_s} \tag{44}$$

Here the analogous rewriting applies with $s'$ given by $1 = 1/s + 1/s'$, if $\mu \leq \lambda/2$ holds.

For showing (iii), i.e.

$$c = \lim_{N \to \infty} -N \log(\omega_N) \quad , \tag{45}$$

it suffices to expand all exponentials in (31) and (33) up to the first order in $N$. On this order the Golden-Thomson inequality is a equality. □

**Remark 5** (The Maassen and Uffink bound)**.** For the case of $N = 2$ and $\lambda = \mu = 1$, in Thm.1, we get $s = r = 1$ and $s' = r' = \infty$. Hence, we recover the Maassen-Uffink bound [21]. Explicitly, we have

$$\omega_2(1,1) = \sup_{\substack{\mathbf{x} \in B_1(\mathbb{C}^d) \\ \mathbf{y} \in B_1(\mathbb{C}^d)}} \left|\mathbf{x}^\dagger U \mathbf{y}\right| = \max_{ij} |U_{ij}| \ . \tag{46}$$

Here we used that $\left|\mathbf{x}^\dagger U \mathbf{y}\right|$ is convex in $\mathbf{x}$ and $\mathbf{y}$. Hence, $\sup_{\mathbf{x},\mathbf{y}}$ is attained at the extreme points of $B_1(\mathbb{C}^d)$. Up to a phase, those extreme points have the form $(0, \cdots, 0, 1, 0 \cdots, 0)$ , i.e. they have their support only on a single site. So, choosing $\mathbf{x}$ and $\mathbf{y}$, with support on the $i-th$ and $j-th$ site, will give $\left|\mathbf{x}^\dagger U \mathbf{y}\right| = |U_{ij}|$.

**Remark 6** (Renyi-Entropies)**.** Alternatively, the bound obtained in Thm. 1 can be expressed in terms of Renyi-entropies: Using statement (i), (ii) and (iii) together directly gives

$$\begin{aligned} c(\lambda,\mu) &\geq -N \log \|U\|_{r',s} \\ &= \inf_{\phi \in \mathcal{H}} N \log \|\phi\|_r - N \log \|U\phi\|_{s'}. \end{aligned} \tag{47}$$

Here a straightforward computation shows

$$\frac{2-r}{r} = \lambda/N \text{ and } \frac{2-s'}{s'} = -\mu/N. \tag{48}$$

So, when we proceed as in (13), substituting the Renyi entropy in (47) gives

$$c(\lambda,\mu) \geq \inf_{\phi \in \mathcal{H}} \lambda H_{r/2}(X|\phi) + \mu H_{s'/2}(Y|\phi). \tag{49}$$

**Lemma 1** (Generalization of the Maassen and Uffink bound)**.** *Let $\mathbf{u}_i$ denote the $i$-th column of the basis transformation $U$ that links the measurements $X$ and $Y$. Then, for $1 \geq \lambda \geq \mu \geq 0$ and all states $\rho$ we have*

$$\lambda H(X|\rho) + \mu H(Y|\rho) \geq -2\lambda \log\left(\sup_{i=1\cdots d} \left\|\mathbf{u}_i\right\|_t\right). \tag{50}$$

*with*

$$t = \frac{2}{(1-\mu/\lambda)} \tag{51}$$

*Note that for the case $1 \geq \mu \geq \lambda \geq 0$ the same holds, if $U$ is replaced by $U^\dagger$, i.e. by the transformation between $Y$ and $X$.*

*Proof.* The linear uncertainty bound $c(\lambda,\mu)$ is homogeneous in $(\lambda,\mu)$. Hence, we can consider

$$c(\lambda,\mu) = \lambda c(1,\mu/\lambda) \tag{52}$$

We will apply Thm. 1, with $N = 2$, in order to get a lower bound. Here, we have $s = \frac{2}{1+\mu/\lambda}$ and

$$\omega_2(1,\mu/\lambda) = \sup_{\substack{\mathbf{x} \in B_1(\mathbb{C}^d) \\ \mathbf{y} \in B_s(\mathbb{C}^d)}} \left|\mathbf{x}^\dagger U \mathbf{y}\right| = \sup_{\substack{i=1,\cdots,d \\ \mathbf{y} \in B_s(\mathbb{C}^d)}} |\mathbf{u}_i\ \mathbf{y}| \,. \tag{53}$$

Here the second equality stems from the same argumentation as in Rem. 5. The sup over $B_s(\mathbb{C}^d)$ on the most right of (53), gives the norm dual to $s$, given by $t = \frac{2}{1-\mu/\lambda}$. All in all we have,

$$\begin{aligned} c(1, \mu/\lambda) &\geq -2\log\left(\omega(1, \mu/\lambda)\right) \\ &= -2\log\left(\sup_{i=1\cdots d}\left\|\mathbf{u}_i\right\|_t\right) \end{aligned} \tag{54}$$

□

**Remark 7** (More than two observables)**.** As mentioned in Sec. 3, the proposition Prop. 1 does not generalize to three measurements. A reasoning, or at least a hint, for this can be found by carefully following the proof of Thm. 1. In principle, the ansatz in (29) can be generalized to more than two measurements as well, and all following steps work out in a similar way, up to (33). Here the Golden-Thompson inequality was used. It is well known, that the direct generalization of this inequality to three operators fails to hold. Hence, the technique of our proof cannot be generalized for this case. We note that there is an ongoing work of exploring more sophisticated generalizations of this inequality [57–60]. However, we leave relating this to entropic uncertainty for future work.

## 5 Additivity of bounds from multiplicativity of $(p, q)$-norms

In this section we will provide the proof of Prop. 1, i.e. the additivity of linear uncertainty relations. By using Thm.1 from the section before we can formulate the linear uncertainty in terms of the logarithm of a $(p, q)$-norm. At this point, it is straightforward to check that the additivity of the linear uncertainty is equivalent to the multiplicativity of the $(p, q)$-norm. In fact, the following theorem Thm.2 provides that, for $p$ and $q$ coming from the correct range: The $(p, q)$ norm of a transformation which admits a product form $U_{AB} = U_A \otimes U_B$ is multiplicative.

**Theorem 2** (Global bounds from local bounds)**.** *Let $X_{AB}$ and $Y_{AB}$ be tensor-product bases of a Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, i.e. we have $X_{AB} = \{X_A^i \otimes X_B^i\}_{i=1,\cdots,d}$ and $Y_{AB} = \{Y_A^i \otimes Y_B^i\}_{i=1,\cdots,d}$, such as $U_{AB} = U_A \otimes U_B$. Furthermore let $\eta_A$ and $\eta_B$ denote the optimal constants for*

$$\begin{aligned} \|U_A\phi\|_q \leq \eta_A \|\phi\|_p \quad \forall \phi \in \mathcal{H}_A \\ \|U_B\phi\|_q \leq \eta_B \|\phi\|_p \quad \forall \phi \in \mathcal{H}_B\ . \end{aligned} \tag{55}$$

*If $1 \leq p \leq q$ then*

$$\|U_{AB}\ \phi\|_q \leq \eta_A\eta_B \|\phi\|_p \quad \forall \phi \in \mathcal{H}_{AB} \tag{56}$$

*holds with $\eta_A\eta_B = \eta_{AB}$ as optimal constant.*

*Proof.* We note that a related result, for pointwise positive maps between Lebesque spaces, was discovered by Grey and Sinnamon [61].

The basic object of this proof will be the $p \otimes q$-norm which will be defined immediately. The basic work of this proof is devoted to show some properties of this norm from which the statement directly follows.

Let $|\phi\rangle \in \mathcal{H}$ with components $\phi = \{\phi_{ij}\}$ sorted within the product base $X_{AB}$ by $\phi_{ij} = \langle\phi|e_i^A \otimes e_j^B\rangle$ and consider the norm

$$\|\phi\|_{q\otimes p} := \left(\sum_i \left(\sum_j |\phi_{ij}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}\ . \tag{57}$$

This norm shares the following properties

$$\begin{aligned} &(i) && \|\phi\|_{q\otimes q} = \|\phi\|_q && (58)\\ &(ii) && \|(\mathbb{I}\otimes V\phi)\|_{r\otimes q} \leq \|\phi\|_{r\otimes p}\,\eta_V && (59)\\ &(iii) && \|\phi\|_{q\otimes p} \leq \|\mathbb{F}\phi\|_{p\otimes q} && (60)\\ &\text{with} && \mathbb{F}\phi_1\otimes\phi_2 = \phi_2\otimes\phi_1 \text{ and } p \leq q. \end{aligned}$$

We will show the validity of $(i - iii)$ in a moment. First notice that, if $(i - iii)$ are valid we can easily conclude

$$\begin{aligned} \|U_{AB}\phi\|_q &= \|U_A \otimes U_B\phi\|_{q\otimes q} \\ &= \|(\mathbb{I}\otimes U_B)(U_A\otimes\mathbb{I})\phi\|_{q\otimes q} \\ &\leq \eta_B\|U_A\otimes\mathbb{I}\phi\|_{q\otimes p} \\ &\leq \eta_B\|\mathbb{I}\otimes U_A\mathbb{F}\phi\|_{p\otimes q} \\ &\leq \eta_B\eta_A\|\mathbb{F}\phi\|_{p\otimes p} \\ &= \eta_B\eta_A\|\phi\|_p\ . \end{aligned} \tag{61}$$

Furthermore, if we consider states that realize equality in (55), i.e. states that belong to optimal $\eta_A$ and $\eta_B$. The tensor-product of two of those states will realize, due to multiplicativity of the $p$-norm, equality in (56) as well. Hence, (61) will prove the main statement of this Theorem.

Property $(i)$ follows directly by plugging $p = q$ in the definition of the $p \otimes q$ norm, here is nothing more to prove. The property $(ii)$ follows by expressing $\mathbb{I}\otimes V$

as $\delta_{ik}V_{jl}$ in $X$-Basis and

$$\|(\mathbb{I}\otimes V\phi)\|_{r\otimes q} = \left(\sum_i \left(\sum_j \left|\sum_{lk} \delta_{ik} V_{jl}\phi_{kl}\right|^q\right)^{\frac{r}{q}}\right)^{\frac{1}{r}} \tag{62}$$

$$= \left(\sum_i \left(\sum_j \left|\sum_{l} V_{jl}\phi_{il}\right|^q\right)^{\frac{r}{q}}\right)^{\frac{1}{r}} \tag{63}$$

$$= \left(\sum_i \|V\phi_i\|_q^r\right)^{\frac{1}{r}} \tag{64}$$

$$\leq \eta_V \left(\sum_i \left(\sum_j |\phi_{ij}|^p\right)^{\frac{r}{p}}\right)^{\frac{1}{r}}$$

$$= \eta_V \|\phi\|_{r\otimes p} \ . \tag{65}$$

As a last step, $(iii)$ is a direct consequence of Minkowski's inequality / $l^p$-triangle inequality (see [62]), i.e. if $p \geq 1$ :

$$\left(\sum_y \left|\sum_x a_{xy}\right|^p\right)^{\frac{1}{p}} \leq \sum_x \left(\sum_y |a_{xy}|^p\right)^{\frac{1}{p}} \tag{66}$$

So, if $1 \leq q/p$ we can use this inequality as follows

$$\|\phi\|_{q\otimes p} = \left(\sum_i \left(\sum_j |\phi_{ij}|^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}}$$

$$= \left(\sum_i \left|\sum_j |\phi_{ij}|^p\right|^{\frac{q}{p}}\right)^{\frac{1}{q/p}\frac{1}{p}}$$

$$\leq \left(\sum_j \left(\sum_i |\phi_{ij}|^{p\frac{q}{p}}\right)^{\frac{p}{q}}\right)^{\frac{1}{p}} = \|\mathbb{F}\phi\|_{p\otimes q} \tag{67}$$

and show the validity of $(iii)$. □

**Lemma 2** (Multiplicativity of the $(p,q)$-norm)**.** *For $1 \leq p \leq q$, the $(p,q)$-norm of a product unitary $U_{AB} = U_A \otimes U_B$ is multiplicative, i.e. we have*

$$||U_{AB}||_{q,p} = ||U_A||_{q,p}||U_B||_{q,p}. \tag{68}$$

*Proof.* This is a direct consequence of Thm. 2. Using the definition of the $(p,q)$-norm we can parse $\eta_A = ||U_A||_{q,p}$, $\eta_B = ||U_B||_{q,p}$ and $\eta_{AB} = ||U_{AB}||_{q,p}$, if we consider $\eta_A$, $\eta_B$ and $\eta_{AB}$ to be optimal bounds. □

### Proof of Prop. 1

*Proof.* For proving Prop. 1 it suffices to proof the additivity of the optimal case, i.e. we will consider $c_A$, $c_B$ and $c_{AB}$ to already be constants for the best linear uncertainty bound. If the additivity

$$c_{AB} = c_A + c_B \tag{69}$$

holds we can directly conclude that the sum of lower bounds on $c_A$ and $c_B$ gives a valid lower bound on $c_{AB}$ as well.

Given measurements $X_{AB}$ and $Y_{AB}$, specified by a product unitary $U_{AB} = U_A \otimes U_B$, we use Thm. 1 to rewrite $c_A$, $c_B$ and $c_{AB}$ as the limit of logarithms of $(p,q)$-norms. We assume $\lambda \leq \mu$ both to be finite and $N$ to be sufficiently large such that we can use Thm. 1 part (ii) (here we needed $\lambda, \mu \leq N/2$), and get

$$c_A = -\lim_{N\to\infty} \log\left(||U_A||_{r,s}\right) \tag{70}$$

$$c_B = -\lim_{N\to\infty} \log\left(||U_B||_{r,s}\right) \tag{71}$$

$$c_{AB} = -\lim_{N\to\infty} \log\left(||U_{AB}||_{r,s}\right) \tag{72}$$

Using $r, s$, as given in (20) it is straightforward to check that $\lambda, \mu \leq N/2$ implies $1 \leq r \leq s$. Therefore, we can use Lem. 2 and get

$$c_{AB} = -\lim_{N\to\infty} \log\left(||U_A||_{r,s}||U_B||_{r,s}\right)$$

$$= -\lim_{N\to\infty} \log\left(||U_A||_{r,s}\right) + \log\left(||U_B||_{r,s}\right)$$

$$= c_A + c_B \ . \tag{73}$$

□

## Outlook and conclusion

In this work we showed that linear uncertainty relations between product type measurements in multipartions are additive. Prop. 1 gives some clear structure to the problem of computing entropic uncertainty bounds. Especially in the context of quantum-coding in cryptography, this result might turn out to be useful, because now it is possible to compute uncertainty bounds in the limit of infinite system sizes for block-coding schemes [6, 63, 64].

The generalization of the Maassen and Uffink bound for arbitrary weights $(\lambda, \mu)$, provided in Lem. 1, can also be directly employed in a multipartite setting in order to obtain valid state-independent uncertainty relations for this case. However, this bound is easy computable, it is only a lower bound and presumably only tight in high symmetrical cases (see [34] for a characterization of tightness for the usual Maassen and Uffink bound). The more general problem of providing a 'good' method for computing the optimal bound $c_{AB}$ remains open. We note that there are only few and special cases, including angular momentum and mutual unbiased bases, where this optimal bound is actually known. Thereby, the cases where the optimal bound can be computed analytically are even fewer [34, 65, 66] and the known numerical methods only work for very small dimensional problems [67]. Here the proof of Thm. 1 might give a new ansatz for better numerics. Explicitly, the minimization in (29) and maximization in (42) are giving rise to apply the method of alternating minimization.

In Sec. 4 we presented an extension to the known connections between the logarithm of $(p, q)$-norms and linear uncertainty relations in terms of the shannon entropy. However, the technique used seems to apply only for the special case we considered. An adaption of this technique to sets of more than two local measurements is not possible without major modifications. As mentioned in Rem. 7, this would require to incorporate generalizations of the Golden-Thompson inequality which seems to be a fruitful topic for future work. The technique from the proof of Thm. 1 might also fail if general POVMs instead of projective measurements are considered. Moreover, it is not clear if Prop. 1 will hold in this case. A third generalization, that does not hold, arises by considering arbitrary Schur-concave functions. Here, the natural question is to ask if at least any entanglement witness can be constructed. A very recent result [68] shows that such witnesses, in fact, can be constructed from Tsallis entropies.

## Acknowledgements

R.S. acknowledges K. Abdelkhalek, O. Gühne, A. Costa, I. Siemon, and R.F. Werner for all the helpful discussions and comments. Furthermore, R.S. acknowledges Coco, Inken and Lars for a careful reading and correcting of this manuscript. R.S. also acknowledges financial support by the RTG 1991 and the SFB DQ-Mat, both founded by the DFG, and the project Q.com-q founded by the BMBF. Finally R.S. acknowledges the hospitality of the Centro de Ciencias de Benasque Pedro Pascual and Petronilla, granted at the beginning of this project. The publication of this article was funded by the Open Access Fund of the Leibniz Universität Hannover.

## 6.2 The computation of entropic uncertainty relations

In the last section, [S18], an central additivity result for entropic uncertainty relations was proven. For two sharp measurement, this result allows to reduce the computation of global linear entropic uncertainty relations to the computation of local bounds. However, the computation of those local bounds stays open. The aim of this section is to provide some insight into to this problem: In the following we will first collect known results on the computational hardness of a closely related problem, and then provide some numerical methods.

### Known hardness results on $(p, q)$-norms

For two sharp observables, $A$ and $B$, related by a unitary $U$, the $(p, q)$-norm of $U$

$$\|U\|_{p,q} = \sup_{\phi \in \mathbb{C}^d} \frac{\|U\phi\|_q}{\|\phi\|_p}$$

turned out to be the proper tool for analysing entropic uncertainty relations. Beside their connection to entropic uncertainty, $(p, q)$-norms have many more applications, especially in the field of operational research (see [Ste05] for an overview). Hence, there is a wide range of literature focusing on computing $(p, q)$-norms. In the following we will list known results on the computability and in-computability of those norms. However, for the parameter range of $(p, q)$'s that relates to uncertainty relations many questions are still open and some of the proof techniques, that were used for other parameters, fail. Therefore, we can, unfortunately not, have a conclusive statement on the computational hardness of uncertainty relations, yet.

The $(p, q)$-norm of a generic $d \times d$-matrix $M$ can be computed efficiently for the cases:

- $p = q = 2$ In this case the norm reduces to the 'usual' operator norm of $M$, which will be attained on the biggest singular value.
- $p = 1$ In this case the extreme points of the set $\|\phi\|_1 = 1$ are of the form $(0, \dots, 0, e^{it}, 0 \dots)$. The functional $\|U\phi\|_q$ is convex. Hence, the norm $\|U\|_{1,q}$ will be attained on one of those extreme points.
- $q = \infty$ This case is dual to $p = 1$, by the duality described in [S18] Thm.1 (22).

In Fig. 6.1, those parameter tuples are marked by blue and green lines and boxed.

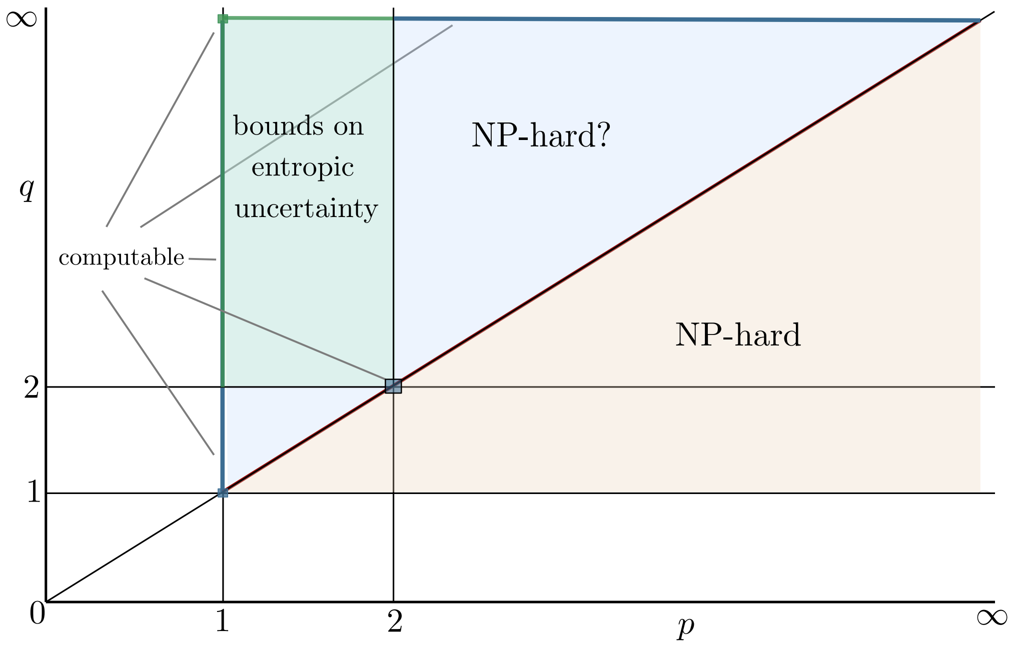


Figure 6.1: Known results on the hardness of computing $(p, q)$-norms.

Computing the $(p, q)$-norm of a generic $d \times d$-matrix $M$ is proven to be NP-hard for:

- $p = \infty, q = 1$ This was proven in [Roh00]. This hardness result can be understood intuitively for the corresponding real valued problem in: The set $\|\phi\|_\infty = 1$, with $\phi \in \mathbb{R}^d$, has $2^d$ extreme points of the form $(x_1, \ldots, x_n)$, with $x_i \in \{1, -1\}$. The only information available in the generic case is, that the convex functional $\|U\phi\|_1$ will attain its maximum on one of those extreme points. Hence, one has to check all of them separately. Explicitly, [Roh00] provided a mapping of the max-cut problem to this case. In [HO10], the more general case $p = \infty, 1 \leq q \leq \infty$ was proven to be NP-hard with the same line of argumentation.

- $p = q \neq 1, 2, \infty$ This was proven in [HO10], by providing instances for general $p = q$ that can be reduced to the case $p = \infty, 1 \leq q \leq \infty$.

- $1 \leq q < p \leq \infty$ This case was proven in [Ste05], by decoding the Knapsack problem into this case.

In Fig. 6.1, the corresponding parameter regions are marked with red colors.

We recall from [S18] Sec. 4, that the $(p, q)$-norm relates to linear entropic uncertainty relations, in terms of Rényi entropies, by

$$-2\log\left(\|U\|_{p,q}\right) = \inf_{\rho} \frac{2-p}{p} H_{p/2}(A|\rho) + \frac{q-2}{q} H_{q/2}(A|\rho) \tag{6.1}$$

Here, we have a proper uncertainty relation if the pre-factors in (6.1) correspond to positive weights. This is the case for $p \leq 2$ and $q \geq 2$. In Fig. 6.1, this region is marked by a green shaded box. Within this area, the point $(2, 2)$ is special when we consider the $(p, q)$-norm of an unitary, since this norm will always be equal to one. Here, Thm. 1 from [S18] tells us that we will need to compute the $(p, q)$-norm in a region around this point in order to get an uncertainty relation. In contrast, the other computable points within the green region are giving valid uncertainty relations:

- $p = 1, q = \infty$ Here, we get $\|U\|_{1,\infty} = \max_{ij} |U_{ij}|$, which leads to the Maassen and Uffink bound.
- $p = 1, q > 2$ Here, we get $\|U\|_{1,\infty} = \max_{ij} \|\mathbf{u}_i\|_q$, which leads to the generalized the Maassen and Uffink bound for arbitrary positive weights provided in [S18] Lem. 1.

However, a hardness result for the missing region, $1 < p < 2$ and $2 < q < \infty$, which is, in [Ste05], phrased to be the 'bad case', remains outstanding.

### Algorithmic methods for computing entropic uncertainty relations and $(p, q)$-norms

In the following two algorithms are presented. Both algorithms are extremely simple and efficient to implement and they both attain a (local) optimum very quickly. In contrast to the algorithm presented in Ch. 4, the algorithms in this section do not provide a precision estimate $\epsilon$. Furthermore, both algorithms rely on a randomly chosen starting point and a convergence to a global optimum is not guaranteed. Hence, they can not be used to proof a hardness statement.

$(p, q)$**-norms:** By use of the (Hölder) identity

$$\|z\|_q = \sup_x \frac{|\langle x, z\rangle|}{\|x\|_{q'}} \text{ with } 1/q' + 1/q = 1,$$

the $(p, q)$-norm can be rewritten as

$$\|U\|_{p,q} = \sup_{\substack{y \in \mathcal{B}_p(\mathbb{C}^d) \\ x \in \mathcal{B}_{q'}(\mathbb{C}^d)}} |\langle x|U|y\rangle| := \sup_{\substack{y \in \mathcal{B}_p(\mathbb{C}^d) \\ x \in \mathcal{B}_{q'}(\mathbb{C}^d)}} L_U(x, y).$$

From this perspective, we have to solve a double maximization of the functional $L_U(x,y)$, which is, up to a phase, linear in $x$ and $y$, where both variables are coming from a convex set. This suggests the following algorithm:

0. Pick a feasible tuple $(x_0, y_0)$ at random.

1. Fix $y_0$ and compute
$$x^* = \operatorname{argmax}\{|\langle x|Uy_0\rangle|\ |x \in \mathcal{B}_{q'}(\mathbb{C}^d)\},$$

2. Fix $x^*$ and compute
$$y^* = \operatorname{argmax}\{|\langle U^\dagger x^*|y\rangle|\ |y \in \mathcal{B}_p(\mathbb{C}^d)\},$$

3. Set $x_0 = x^*$ and $y_0 = y^*$ and proceed with step 1 .

The method that underlies the above algorithm is commonly called *alternating maximization* (see for example [GB05]). The value of the objective functional $L_U(x,y)$ is increased by improving $x$ and $y$ alternatingly. Hence, a full loop of the above algorithm guarantees to output a tuple $(x^*, y^*)$ with
$$L_U(x^*, y^*) \geq L_U(x_0, y_0).$$
The crucial point in an alternating maximization is the ability to compute an improvement of $L_U(x,y)$, if one of the parameters $x$ or $y$ is fixed. In the above case we have to maximize a linear functional over the $p$-norm ball $\mathcal{B}_p(\mathbb{C}^d)$, i.e. we have to solve problems of the form
$$z^* = \operatorname{argmax}\{|\langle z|\lambda\rangle|\ |z \in \mathcal{B}_p(\mathbb{C}^d)\},$$
which is an easy problem: We have
$$|\langle z|\lambda\rangle| = \left|\sum_i z_i\lambda_i\right| \leq \sum_i |z_i\lambda_i|\,.$$
Here, equality is attained when the vectors $z$ and $\lambda$ have componentwise coinciding phases. We can assume $\lambda \in \mathbb{R}^d_+$ with out loss of generality. In this case we only have to consider $z \in \mathbb{R}^d_+$, as well. We know by convexity, that the optimizer $z^*$ has to be on the boundary of $\mathcal{B}_p(\mathbb{C}^d)$, which is described by the functional $\|z\|_p^p = 0$. The linear functional $\langle z|\lambda\rangle$ is maximized if the gradient of the boundary is parallel to $\lambda$. Computing this gradient gives componentwise conditions
$$\alpha\lambda_i = \partial_i\|z\|_p^p = \partial_i \sum_i z_i^p = p z_i^{p-1},$$

with a constant $\alpha$ that regulates the norm of $z$. Hence, we get

$$z^* = \frac{\lambda^{\frac{1}{p-1}}}{\|\lambda^{\frac{1}{p-1}}\|_p},$$

where the power of a vector has to be understood componentwise.

This algorithm converges very quickly to a stationary point. Thereby, it is not guaranteed that this stationary point corresponds to the global maximum.

In practice, this algorithm will be performed many times with many random starting points and the hope to meet the global optimum in one of those runs. Interestingly, all tested examples show a finite and small set of those stationary points, which raises the hope to actually attain the global optimum by a finite amount of runs.

**Linear entropic uncertainty relations:** The following algorithm is based on an alternating optimization too. It can compute linear entropic uncertainty relations for collections of arbitrary, not necessarily sharp, measurements:

For a measurement $A$, with POVM elements $A(i)$ and outcomes on an alphabet $\Omega$, and a probability distribution $p \in \mathcal{P}_{|\Omega|}$, define the operator

$$K_A(p) = -\sum_{i\in\omega} \log(p_i) A(i)$$

The expectation of this operator, with respect to a state $\rho$, gives the cross entropy between the distribution $p_\rho^A$ and $p$, i.e.

$$\operatorname{tr}(\rho K_A(p)) = -\sum_i \operatorname{tr}(\rho A(i)) \log(p_i)$$

Hence, we have

$$H(A|\rho) = \min_{p\in\mathcal{P}_{|\Omega|}} \operatorname{tr}(\rho K_A(p)) .$$

Let $A_1, \ldots, A_n$ be a collection of measurements. Within the formulation above, the optimal linear entropic uncertainty relation can be computed by:

$$\begin{aligned}\min_\rho \sum_j a_j H(A_j|\rho) &= \min_\rho \min_{p^1,\ldots,p^n} \operatorname{tr}\left(\rho \sum_j a_j K_{A_j}(p^j)\right) \\ &:= \min_\rho \min_{p^1,\ldots,p^n} L(p^1, \ldots, p^n, \rho),\end{aligned}$$

where $p^1, \ldots, p^n$ denotes a collection of probability distributions on the alphabets $\Omega_1, \ldots, \Omega_n$.

The objective functional $L(p^1, \dots, p^n, \rho)$ can be minimized by the following algorithm:

0. Pick feasible starting parameters $\rho_0$ and $p_0^1, \dots, p_0^j$ at random.

1. Fix $p_0^1, \dots, p_0^n$ and

$$\rho^* = \mathrm{argmin}\Big\{ \mathrm{tr}\Big(\rho \sum_j a_j K_{A_j}(p^j)\Big) \Big| \rho \in \text{ quantum states } \Big\}$$

2. Fix $\rho^*$ and compute

$$\big(p_*^1, \dots, p_*^n\big) = \mathrm{argmin}\Big\{ \mathrm{tr}\Big(\rho^* \sum_j a_j K_{A_j}(p^j)\Big) \Big| \forall j : p^j \in \mathcal{P}_{|\Omega_j|} \Big\}$$

3. Set $p_0^1 = p_*^1, \dots, p_0^n = p_*^n$ and $\rho_0 = \rho^*$, and proceed with step 1.

Here, the minimum in step 1. is attained on the smallest eigenvector of the positive operator $\sum_j a_j K_{A_j}(p^j)$. Hence, computing $\rho^*$ is easy. For a fixed $\rho^*$, the minimum in step 2. is attained if the distributions $p^j$ match the distributions corresponding to a $A_j$ measurement of $\rho$, i.e. by

$$p^j = p_\rho^{A_j} = \big(\mathrm{tr}(A_j(1)\rho), \mathrm{tr}(A_j(2)\rho), \dots\big).$$

This circumstance was already used in the proof of Thm. 1 in [S18] and can be directly deduced from the positivity of the relative entropy.

Also this algorithm converges quickly to a stationary point and, in all tested examples, the set of stationary points was moderately small, as well. However, in a direct comparison the first algorithm has less stationary point and performs numerically much more stable because no logarithms have to be computed.

The unfortunate drawback of both algorithms is that in finite runtime only upper bounds on the optimal uncertainty are produced. Hence, in contrast to the method in chapter Ch. 4, they can not be used to derive reliable security criteria or steering witnesses. However, it should be made clear that the above algorithms still provide a step forward with respect to the methods used in previous works. There uncertainty relations were either obtained by random sampling pure states or by parametrizing a (small) dimensional statespace in order to use 'black box algorithms' like mathematica's `NMinimize[]`, see for example [RMM17b].

# CHAPTER 7

# Entropic uncertainty relations for mutual unbiased bases

In this chapter we will study entropic preparation uncertainty relations for an explicit class of examples. We will consider pairs of sharp measurements with mutually unbiased eigenbases, thereby we will mostly concentrate on bases linked by a finite Fourier transformation. For those cases the Maassen and Uffink bound represents the optimal linear uncertainty relations. However, the corresponding uncertainty regions are not convex. Hence, there is still place for non-linear improvements.

The central part of this chapter is [ASM$^+$15] 'Optimality of entropic uncertainty relations', here we investigate those non-linear bounds numerically. Interestingly, we observe the same additivity behaviour (see Fig. 1.9) which was proven for linear bounds in [S18], for non-linear bounds, too. As a general result we characterize the class of measurement for which the Maassen and Uffink bound gives the best linear bound, and we provide a basic theorem, which characterizes all states that achieve equality for the Maassen and Uffink inequality.

## 7.1 [ASM$^+$15]

*Optimality of entropic uncertainty relations*

- **Authors:** Kais Abdelkhalek, René Schwonnek, Hans Maassen, Fabian Furrer, Jörg Duhme, Philippe Raynal, Berthold-Georg Englert, Reinhard F. Werner
- **Published in:** International Journal of Quantum Informormation 13, 1550045 (2015)
- **DOI:** 10.1142/S0219749915500458
- **Presented version:** The presented version is identical to arXiv:1509.00398, the literature is placed at the end of this thesis.
- **Contributions:** The main contributions to this work were provided by Kais Abdelkhalek and René Schwonnek. Theorem IV.1 elaborates on an unpublished work contributed by Hans Maassen. Central ideas for a proof of the theorem V.2 and V.3 were contributed by René Schwonnek. The additivity conjectures V.6 and V.7 are inspired by numerical investigations done by Kais Abdelkhalek and René Schwonnek.
- **Main results:**
  - Measurements with a tight Maassen and Uffink bound are characterized.
  - The states that achieve equality in the Maassen and Uffink bound are characterized.
  - The optimal non-linear bound can always be realized by pure states.
  - The structure of non-linear bounds for MUBs is investigated numerically.

,

# Optimality of entropic uncertainty relations

Kais Abdelkhalek,[1] René Schwonnek,[1] Hans Maassen,[2] Fabian Furrer,[3] Jörg Duhme,[1] Philippe Raynal,[4, 5] Berthold-Georg Englert,[4, 6, 7] and Reinhard F. Werner[1]

[1] *Institut für Theoretische Physik, Leibniz Universität Hannover, Germany*
[2] *Department of Mathematics, Radboud University, Nijmegen, The Netherlands*
[3] *Department of Physics, University of Tokyo, Japan*
[4] *Centre for Quantum Technologies, National University of Singapore, Singapore*
[5] *University Scholars Programme, National University of Singapore, Singapore*
[6] *Department of Physics, National University of Singapore, Singapore*
[7] *MajuLab, CNRS-UNS-NUS-NTU International Joint Unit, UMI 3654, Singapore*

The entropic uncertainty relation proven by Maassen and Uffink for arbitrary pairs of two observables is known to be non-optimal. Here, we call an uncertainty relation optimal, if the lower bound can be attained for any value of either of the corresponding uncertainties. In this work we establish optimal uncertainty relations by characterising the optimal lower bound in scenarios similar to the Maassen-Uffink type. We disprove a conjecture by Englert *et al.* and generalise various previous results. However, we are still far from a complete understanding and, based on numerical investigation and analytical results in small dimension, we present a number of conjectures.

## I. INTRODUCTION

As a characteristic trait, quantum systems possess properties that are incompatible — properties that are equally real but mutually exclusive. In a pair of incompatible properties, if we have precise knowledge about one property, what we know about the other is necessarily imprecise. More generally, we can trade knowledge about one property for knowledge about the other and so know both imperfectly, and quantify the lack of knowledge by a suitable measure of uncertainty. Then, the compromises allowed by nature have their mathematical expressions in the form of *uncertainty relations*, which are inequalities that follow from the formalism of quantum theory.

The study of uncertainty tradeoffs originated in Heisenberg's pioneering work[1] of 1927 and was soon brought into a clear mathematical form by Kennard[2]. Weyl gave another early proof[3]. He was apparently unaware of Heisenberg's paper and gives credit for the idea to Pauli, who seems to have learned it from Heisenberg in a letter prior to the publication of [1]. The modern textbook proof combining the Schwarz inequality with the commutation relations is due to Robertson[4]. In this tradition the "uncertainty of observable $X$ in the state $\rho$" is identified with the root of the variance of the probability distribution of the outcomes of an $X$-measurement on particles prepared according to $\rho$, i.e.,

$$\delta X = \sqrt{\mathrm{tr}\,(\rho X^2) - \mathrm{tr}(\rho X)^2}\,, \tag{1}$$

The key requirement for Heisenberg's uncertainty relation $\delta Q\,\delta P \geq \hbar/2$ to hold is that these variances are evaluated in the same state. The relation is thus a quantitative expression of the observation that there are no dispersion-free states, and is hence of the type "preparation uncertainty relation". This is in contrast to "measurement uncertainty relations" which express the feature of quantum mechanics that some observables may not be measured jointly, which also implies that any measurement of one observable $X$ implies a disturbance of the other in the sense that it cannot be inferred from a measurement on the state after an $X$-measurement. This aspect, although more prominent in Heisenberg's paper than the preparation side, was made precise only recently[5] (also [6, 7]).

In this paper we are interested in preparation uncertainty relations for quantum systems of finite dimension $d$. A standard scenario in which this is of interest is the tradeoff between Welcher-Weg information and interference patterns at a multiport interferometer. In this minimalistic instance of wave-particle duality[8] one observable would detect particles on each of the internal paths of the interferometer, thus detecting a particle-like property, whereas the detectors at the end pick up wave-like interference. Uncertainty in this situation expresses the physical fact that if we prepare incoming particles so that they all go along the same path, we loose the interference contrast and, conversely, that large interference contrast is only possible when all paths are "traversed" with roughly equal probability. Another standard context for finite-dimensional uncertainty relations is quantum information theory, particularly quantum key distribution. Large parts of this theory have been developed in finite dimension, and there are many situations in which the incompatibility as expressed by uncertainty relations plays an important role (e.g. in security proofs[9] of cryptographic protocols).

What is common to these motivating instances of finite-dimensional uncertainty is that the outcomes of the respective observables are labelled in a completely arbitrary way. However, a variance depends not only on the abstract outcomes and their probabilities, but also on the real numbers we assign to them. For example, by multiplying all these numbers by the same factor we also multiply $\delta X$. Moreover, variance will change if we permute the outcomes, which is as easy to do with beams in optical fibers as with freely re-codable bits of information. Basically motivated by such considerations, Deutsch[10] suggested to use entropies to quantify the (lack of) sharpness of a probability distribution. This led to the famous entropic uncertainty relation established by Maassen and Uffink[11], to which we will refer to as the *MU bound*. It describes the sharpness tradeoff for the outcome distributions $p_X^\rho$ and $p_Y^\rho$ of two observables $X, Y$ in the same state $\rho$ in terms of their Rényi entropies $H_\alpha$, $H_\beta$ (see (6)), provided that these parameters satisfy the *duality relation*

$$\frac{1}{\alpha} + \frac{1}{\beta} = 2\ . \tag{2}$$

When the observables $X$ and $Y$ are given in terms of their eigenbases $\{x_i\}$ and $\{y_j\}$, so that $p_X^\rho(i) = \langle x_i|\rho|x_i\rangle$ and $p_Y^\rho(j) = \langle y_j|\rho|y_j\rangle$, the MU bound is

$$H_\alpha(p_X^\rho) + H_\beta(p_Y^\rho) \geq -\log \max_{j,k} |\langle y_k|x_j\rangle|^2\,. \tag{3}$$

The bound becomes zero when the two bases share a vector, and maximal (namely $\log d$) if the bases are mutually unbiased, so that all scalar products $\langle y_k|x_j\rangle$ have the same modulus.

An alternative to entropies would again be variances, once one realizes that for defining a variance it is not really necessary to have $\mathbb{R}$-valued random variables. It suffices to have outcomes in a metric space $\Omega$ with metric $\Delta$, so that the variance of a probability measure $\mu$ on $\Omega$ becomes

$$\mathrm{var}(\mu) = \inf_{\eta\in\Omega} \int \mu(d\omega)\ \Delta(\omega,\eta)^2\,. \tag{4}$$

When $\Omega = \{1,\ldots,d\}$ the only permutation invariant metrics are $\Delta(i,j) = c(1-\delta_{ij})$, and we will just fix the constant $c = 1$. Then

$$\mathrm{var}(p) = \min_j \sum_i p(i)\,(1-\delta_{ij})^2 = 1 - \max_j p(j)\,. \tag{5}$$

Up to a rescaling this is the so-called min entropy $H_\infty(p) = -\log \max_j p(j)$.

How then should we write an uncertainty relation in this general context? We will see that it is not wise to fix in advance the functional form of the tradeoff relation between $H_\alpha(p_X)$ and $H_\beta(p_Y)$. Instead, the best and most intuitive representation of the tradeoff is the diagram of all pairs $(H_\alpha(p_X), H_\beta(p_Y))$, ranging over all choices of input states $\rho$. An advantage of this representation is also that it changes in a simple way by a rescaling like the replacement of the variance (5) by $H_\infty$. For qubits ($d = 2$), all measures of sharpness are functions of each other, so all such diagrams are equivalent. Figure 1 is drawn for the Shannon entropy $H = H_1$. Some details of the diagram of *all* pairs of entropies, shown on the left, are clearly not relevant for the uncertainty tradeoff, in which we ask *how small* we can simultaneously make the entropies. For this question it is the lower left corner of the diagram which matters, i.e., the set in the right diagram. It can be described as adding to any pair of entropies the full closed positive (north-east) quadrant above it. It is completely described by its lower left boundary, consisting of those entropy pairs with the property that for no other state one can have one entropy strictly smaller and the other at least as small. We consider the resulting curve as the complete description of the uncertainty tradeoffs between the entropies involved. Characterising this curve is the aim of this paper.

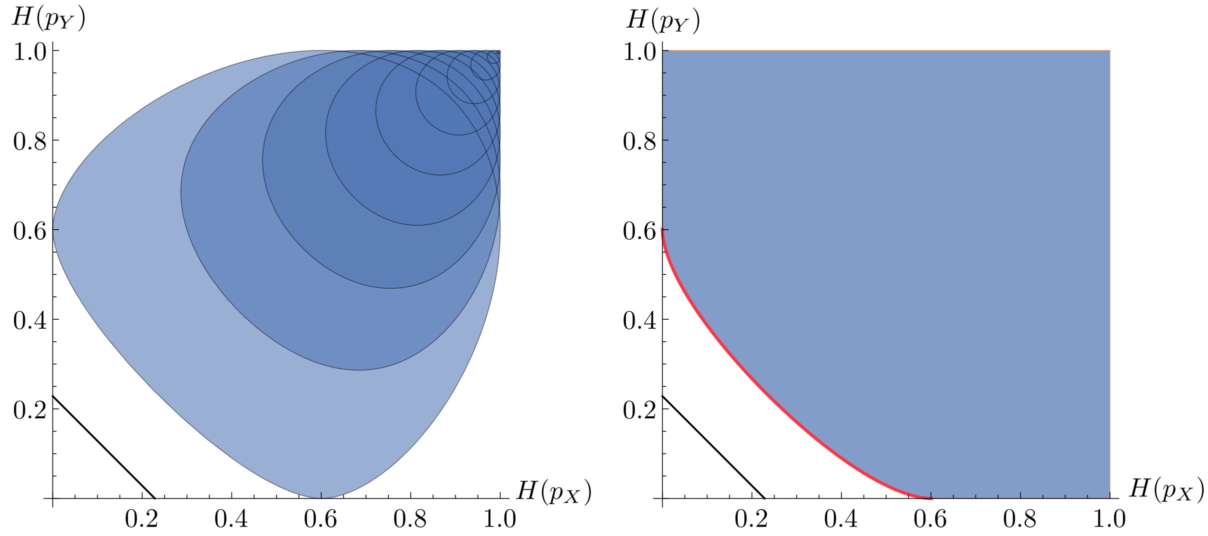


FIG. 1: Entropy pairs for $d = 2$ and the observables $X = \sigma_z$ and $Y = (\sigma_x + \sigma_z)/\sqrt{2}$. Left panel: The shaded set gives all pairs $(H(p_X^\rho), H(p_Y^\rho))$. The contours describe the subsets which can be reached by pure states with a fixed admixture of $\rho = \mathbb{1}/2$. Right panel: The shaded set is the "monotone closure" of the one on the left (see text). The solid curve represents the optimal bound: For entropy pairs on this bound it is impossible to reduce one entropy without enlarging the other. The thin line closer to the origin is the MU bound.

We will always consider a quantum system in a $d$-dimensional Hilbert space, and consider two projection valued observables with $d$ outcomes. This amounts to the choice of two bases $\{x_i\}$ and $\{y_j\}$, and for the question at hand the choice is completely described by the unitary overlap matrix $U_{ij} = \langle x_i | y_j \rangle$ modulo multiplication by diagonal unitary matrices or permutation matrices from either side. In the motivating standard case, closest to the case of position and momentum of continuous variables, the $U$ represents the discrete Fourier transform of either the cyclic group of $n$ elements or, if $n$ is composite, another finite abelian group of order $n$. More generally, we also consider complex Hadamard matrices, i.e., unitary operators such that $|U_{ij}| = 1/\sqrt{d}$ for all

$i, j$. The bases are then called mutually unbiased, and we can think of a multiport interferometer generalizing a 50:50-beam splitter. Such bases also represent complementary pairs of measurements from the informational point of view. However, we will not restrict our study to these special classes of unitary matrices — several results will hold for arbitrary unitary matrices. For generalized observables (POVMs) or $k$-tuples of observables similar questions can be asked, but we will not consider them in this paper. For the quantification of uncertainty or unsharpness we use the Rényi entropies $H_\alpha$ ($1/2 \leq \alpha \leq \infty$), and denote by $H = H_1$ the standard case of the Shannon entropy. Mostly we assume that the Rényi parameters $\alpha$ and $\beta$ used for $X$ and $Y$, respectively, satisfy the duality relation (2). Again, the questions make sense also for other measures, e.g., related to majorization, or for variances, but these will not be considered here. We will also restrict ourselves to state-independent bounds, i.e., to the entropy pairs achievable by arbitrary states. When more is known about the state, for example about further expectation values, the entropy diagram for the subset may be quite different. Thus we do not consider inequalities like the Robertson inequality for variances, where the lower bound depends on the expectation of a commutator.

*Outline.* In Sect. II we briefly define all the relevant quantities and state our problem in precise mathematical terms. We present a brief review of previous results in Sect. III. In Sect. IV we provide a characterization of the case of equality in the MU bound and thereby show that the MU bound is not optimal in almost all cases. Our main results are presented in Sect. V. We are not able to completely solve the problem in all its generality. However, we provide strong conjectures (Sect. V E) which, if true, heavily reduce the complexity of the problem.

## II. PRELIMINARIES AND NOTATION

For $\alpha \in [\frac{1}{2}, \infty]$ the *$\alpha$-Rényi entropy* of a probability distribution $p \in (0,1)^d$ is defined by

$$H_\alpha(p) = \begin{cases} \frac{1}{1-\alpha} \log \sum_{j=1}^d p(j)^\alpha & \text{if } \alpha \neq 1, \infty \\ -\sum_{j=1}^d p(j) \log p(j) & \text{if } \alpha = 1 \\ -\log \max_j p(j) & \text{if } \alpha = \infty \,. \end{cases} \tag{6}$$

The logarithms can be taken in any base (as long as it is always the same base). We follow the information theory convention of using base-2 logarithms, although base $d$ would also be natural in our context, as it would normalize the range to $0 \leq H_\alpha(p) \leq \log d = 1$. Monotone functions of the entropies tell the same story. In this sense we also cover "Tsallis entropies" $T_\alpha(p) = (1-\alpha)^{-1}(1 - \sum_j p(j)^\alpha)$.

Each entropy diagram will be drawn for a fixed choice of observables (i.e., bases) $X, Y$ and values of the Rényi parameters $\alpha, \beta$, so that we consider a map $f$ from the state space to $\mathbb{R}^2_+$ given by

$$f(\rho) = \big(f_1(\rho), f_2(\rho)\big) = \big(H_\alpha(p_X^\rho), H_\beta(p_Y^\rho)\big) \,. \tag{7}$$

For any choice we can define the order relation $\sqsubseteq$ on the state space, so that $\rho \sqsubseteq \rho'$ stands for "$f_1(\rho) \leq f_1(\rho')$ and $f_2(\rho) \leq f_2(\rho')$". The *uncertainty diagram* is the monotone closure of the range $\{f(\rho)\}$, i.e., it is the set $S$ containing precisely the pairs $(h_1, h_2) \in S$ for which there is a state $\rho$ with $f_i(\rho) \leq h_i$ for $i = 1, 2$ (compare FIG. 1). We call a state $\rho$ *optimal* if $\rho' \sqsubseteq \rho$ implies $\rho \sqsubseteq \rho'$, and hence $f(\rho) = f(\rho')$. The corresponding *optimal points* in the entropy plane are characterized by the property that the uncertainty diagram contains no points to their south-west. We call the set of all optimal points the curve of minimal entropies or the *optimal bound.* Therefore the optimal bound corresponds to a function $\gamma : (0, \log d) \to (0, \log d)$ for which

$$H_\alpha(p_X^\rho) \geq \gamma\big(H_\beta(p_Y^\rho)\big) \tag{8}$$

with the property that equality can be obtained for all possible values of $H_\beta(p_Y^\rho)$.

If for some state the MU bound is saturated we call this state an *equality state.* The corresponding point in the entropy plane is an *equality point.* If an equality point exists we call the MU bound *tight.* The MU bound is said to be *optimal*, whenever it completely coincides with the optimal bound.

A Hadamard matrix is a unitary matrix $U$ with elements satisfying $|U_{jk}| = 1/\sqrt{d}$. The Fourier matrix is the matrix $U^F$ with components satisfying

$$U_{jk}^F = \frac{1}{\sqrt{d}}\, \mathrm{e}^{\frac{2\pi \mathrm{i}}{d} jk} \ , \quad j, k = 0, ..., d-1 \ . \tag{9}$$

The Fourier matrix is hence a special instance of a Hadamard matrix. This example generalizes to the wider setting of *finite abelian groups*, rather than just the cyclic group of $d$ elements as in (9). To this end we consider the index set $J$ for the first matrix index of $U$ to equipped with a commutative binary operation "+" turning it into a group. The second index is similarly labelled by the so-called dual group, denoted here by $K$. A symmetric way to express the relation between these groups is via the canonical bicharacter of the pair $(J, K)$, which is a function $\zeta : J \times K \to \mathbb{C}$. It has the property that the for every $k$ the function $j \mapsto \zeta(j, k)$ is a homomorphism from $J$ to the complex numbers with modulus 1, and that, conversely every such homomorphism is of this form for some unique $k \in K$. Moreover, the same is true vice versa for the functions $k \mapsto \zeta(j, k)$ with fixed $j \in J$. The Fourier matrix in this case is $U_{jk} = d^{-1/2}\zeta(j, k)$, where $d = |J| = |K|$. It is unitary and obviously a Hadamard matrix. When $d$ is not a prime there are several non-isomorphic abelian groups of order $d$.

## III. PREVIOUS RESULTS

There has been considerable work to generalize and improve the MU bound, e.g. by using more general entropy functions [12] or more than two observables [13–16] (see also [17] for a review on entropic uncertainty relations). Most efforts, however, considered only the sum of the entropies (e.g. [18–25]), thereby already fixing the functional form of the tradeoff relation and not capturing all the information contained in the entropy diagram.

In this work we are instead interested in characterising the curve of minimal entropies which we consider the optimal lower bound on the two entropies involved. There are, to the best of our knowledge, only very few results in the literature about the curve of minimal entropies in the finite-dimensional setting. In [26, 27] the authors note that the MU bound is not optimal in the simplest case of dimension $d = 2$ and compute the optimal bound for general unitary operators, but restricted to the Shannon case $\alpha = \beta = 1$. In [8] a conjecture about the entropy minimizing states is presented. We will see that this conjecture needs improvement.

## IV. EQUALITY IN THE MAASSEN-UFFINK UNCERTAINTY RELATION

The MU bound provides a lower bound on the sum of two Rényi entropies that satisfy the duality relation (2). When characterising the curve of minimal entropies, it is natural to first investigate the case of equality in the MU bound. If the unitary operator linking the observables is a Hadamard matrix, it is clear that the MU bound is tight. Indeed, any eigenvector of the observables, $\{x_i\}$ or $\{y_i\}$, is an equality state. But can one also find equality points for arbitrary unitary operators?

There already exist some results in the literature discussing this question, most importantly [28] and [29]. In the latter work the authors show the link between the two concepts of uncertainty principle and data processing inequality. Using this link the characterisation of all states that saturate the uncertainty relation reduces to the question of characterising all states for which the application of a certain channel does not imply loss of information. Employing this technique the authors can characterize all quantum states that saturate the MU bound in the restricted setting of observables related by Fourier transformation and Shannon entropies. A more general result was obtained in [28], namely a complete characterisation of all equality points in the special case $\alpha = \beta = 1$, i.e. for Shannon entropies. Here we present an alternative proof of the uncertainty relation which allows us to generalize these from Shannon entropies to the case of arbitrary pairs of Rényi entropies that satisfy the duality relation.

The main result of this section is the following Theorem. In its formulation the "support" of a probability distribution is the set of points with non-zero probability, and $|M|$ denotes the number of elements of a set $M$.

**Theorem IV.1.** *Let $\alpha, \beta > \frac{1}{2}$ be such that $1/\alpha + 1/\beta = 2$, and let $X, Y$ be bases with $c = \max_{j,k} |\langle y_k|x_j\rangle|$. Let $\rho$ be a state, and denote by $s_X$ and $s_Y$ the supports of the distributions $p_X^\rho$ and $p_Y^\rho$. Then equality in the MU uncertainty relation*

$$H_\alpha(p_X^\rho) + H_\beta(p_Y^\rho) \geq \log \frac{1}{c^2} \tag{10}$$

*is reached if and only if $\rho = |\psi\rangle\langle\psi|$ is a pure state and, possibly after multiplying the basis vectors $x_i, y_j$ with suitable phases, the following condition holds:*

$$\langle x_i|\psi\rangle = |s_X|^{-1/2}\ , \quad \langle y_j|\psi\rangle = |s_Y|^{-1/2}, \quad \textit{and} \quad \langle y_j|x_i\rangle = c \quad \textit{for } i \in s_X \textit{ and } j \in s_Y. \tag{11}$$

*Moreover,*

$$|s_X|\,|s_Y| = \frac{1}{c^2}. \tag{12}$$

*Proof.* We assume first that $\rho = |\psi\rangle\langle\psi|$ is pure, and will show that this choice is even necessary at the end of the proof. We fix $\psi$ from now on, and choose phases for the basis elements so that, for $i \in s_X$, $j \in s_Y$ we have

$$\varphi_i = \langle x_i|\psi\rangle > 0 \quad \text{and} \quad \widehat{\varphi}_j = \langle y_j|\psi\rangle > 0. \tag{13}$$

Note that this will change neither $c$ nor the probability distributions. Furthermore, we assume without loss of generality that $\alpha \leq \beta$. We usually eliminate $\beta$ by the duality relation, so the basic parameter to choose is $\alpha$ with $1/2 < \alpha \leq 1$.

Our proof is inspired by interpolation theory, and involves the application of the maximum principle to a certain analytic "magic function" $F$. We do not pretend that finding this function is straightforward, since we also came by it in several stages of generalization and simplification. We define

$$F(z) \;=\; c^{1-z}\lambda^z \sum_{i,j\in s} \varphi_i^{\alpha z}\ \langle x_i|y_j\rangle\ \widehat{\varphi}_j^{\beta z} \tag{14}$$

$$\text{with} \quad \lambda \;=\; \big(\|\varphi^\alpha\|_2\, \|\widehat{\varphi}^\beta\|_2\big)^{-1}, \tag{15}$$

where "$i,j \in s$" is short hand for $i \in s_X$ and $j \in s_Y$, and $\varphi^\alpha$ is the componentwise power of $\varphi$, so that

$$\|\varphi^\alpha\|_2^2 = \sum_i \varphi_i^{2\alpha}, \tag{16}$$

and similarly for $\widehat{\varphi}$. The domain $\mathcal{G}$ on which this function is analyzed is the strip

$$\mathcal{G} = \{z \in \mathbb{C} \,|\, 1 \leq \operatorname{Re} z \leq 2\}, \tag{17}$$

which is also depicted in FIG. 2. Now since the sum (14) is finite and $|r^{\alpha z}| = r^{\alpha \operatorname{Re} z}$ is bounded on $\mathcal{G}$ for every $r > 0$, $F$ is also bounded on $\mathcal{G}$, and the restriction of an entire analytic function. We claim that it is bounded in absolute value by 1. We estimate this separately for the two boundary lines. That is, for $r \in \mathbb{R}$ we have, with $U_{ij} = \langle x_i | y_j \rangle$

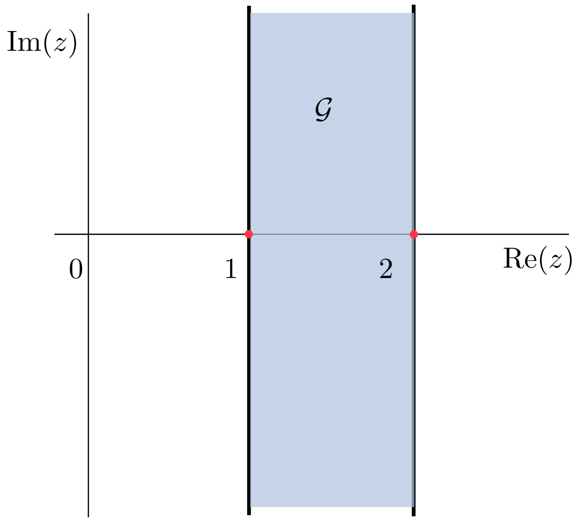


FIG. 2: Domain of $F$ in the complex plane

$$\begin{aligned}
|F(1+\mathrm{i}r)| &= \lambda \left| \sum_{i,j\in s} \varphi_i^{\alpha(1+\mathrm{i}r)} \; U_{ij} \widehat{\varphi}_j^{\beta(1+\mathrm{i}r)} \right| \\
&= \lambda \left| \langle \varphi^{\alpha(1-\mathrm{i}r)} | U | \widehat{\varphi}^{\beta(1+\mathrm{i}r)} \rangle \right| \\
&\leq \lambda \|\varphi^{\alpha(1-\mathrm{i}r)}\|_2 \|\widehat{\varphi}^{\beta(1+\mathrm{i}r)}\|_2 \\
&= \lambda \|\varphi^\alpha\|_2 \, \|\widehat{\varphi}^\beta\|_2 = 1. \qquad (18)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
|F(2+\mathrm{i}r)| &= c^{-1}\lambda^2 \left| \sum_{i,j\in s} \varphi_i^{\alpha(2+\mathrm{i}r)} \; U_{ij} \widehat{\varphi}_j^{\beta(2+\mathrm{i}r)} \right| \\
&\leq \lambda^2 \sum_{i,j\in s} \varphi_i^2 \; |U_{ij}/c| \; \widehat{\varphi}_j^{2\beta} \qquad (19) \\
&\leq \lambda^2 \sum_{i,j\in s} \varphi_i^{2\alpha} \; \widehat{\varphi}_j^{2\beta} \qquad (20) \\
&= \lambda^2 \|\varphi^\alpha\|_2^2 \, \|\widehat{\varphi}^\beta\|_2^2 = 1. \qquad (21)
\end{aligned}$$

Hence, by the maximum principle, $|F(z)| \leq 1$ for all $z \in \mathcal{G}$.

In order to relate this to entropies we consider the special value $z = 1/\alpha$, which always lies in the strip, but for $\alpha = 1$ is a boundary point. We get

$$\begin{aligned}
F\left(\frac{1}{\alpha}\right) &= c^{1-1/\alpha} \lambda^{1/\alpha} \sum_{ij\in s} \varphi_i \; U_{ij} \widehat{\varphi}_j^{\beta/\alpha} \\
&= c^{1-1/\alpha} \lambda^{1/\alpha} \sum_j \widehat{\varphi}_j^{1+\beta/\alpha} \qquad (22) \\
&= c^{1-1/\alpha} \big( \|\varphi^\alpha\|_2^{-1/\alpha} \, \|\widehat{\varphi}^\beta\|_2^{-1/\alpha} \big) \|\widehat{\varphi}^\beta\|_2^2 \qquad (23) \\
&= c^{1-1/\alpha} \|\varphi^\alpha\|_2^{-1/\alpha} \, \|\widehat{\varphi}^\beta\|_2^{1/\beta}, \qquad (24)
\end{aligned}$$

where at (22) we used that $\sum_i \varphi_i U_{ij} = \widehat{\varphi}_j$, and at (23) the definition of $\lambda$ and duality of $\alpha$ and $\beta$. For taking the logarithm of this expression we use that

$$\begin{aligned} \log\bigl(\|\varphi^\alpha\|_2^{-1/\alpha}\bigr) &= -\frac{1-\alpha}{2\alpha} H_\alpha(\varphi^2) \\ \text{and} \quad \log\bigl(\|\widehat{\varphi}^\beta\|_2^{1/\beta}\bigr) &= \frac{1-\beta}{2\beta} H_\beta(\widehat{\varphi}^2) = H_\beta(\widehat{\varphi}^2) \end{aligned} \tag{25}$$

and get, equivalently to $F(1/\alpha) \leq 1$, the inequality

$$\log F\left(\frac{1}{\alpha}\right) = -\frac{1-\alpha}{2\alpha}\Bigl(\log(c^2) + H_\alpha(\varphi^2) + H_\beta(\widehat{\varphi}^2)\Bigr) \leq 0. \tag{26}$$

For $\alpha \neq 1$ we cancel the common factor and get the MU inequality. For $\alpha = 1$ we always get $F(1) = 1$, and the MU inequality is obtained by taking the limit $\alpha \to 1$. However, it is better to express it instead by the derivative of $F$. For $\alpha = \beta = 1$ we get

$$F'(1) = -\log c - \frac{1}{2} H_1(\varphi^2) - \frac{1}{2} H_1(\widehat{\varphi}^2) \leq 0, \tag{27}$$

because for small $\varepsilon$ we must have $F(1+\varepsilon) \leq 1$.

The advantage of this derivation of the MU inequality is that we have powerful characterizations of the equality case. So suppose that equality holds in the MU inequality. Then for $\alpha < 1$ this means that $F$ attains its maximum modulus 1 at the interior point $1/\alpha$ of the strip $\mathcal{G}$, and the Phragmén-Lindelöf Theorem[30] tells us that $F = 1$ is the constant function. For $\alpha = 1$ we need a variant of the maximum principle due to Hopf[31] (see, e.g. Thm. 2.7 in [32]), saying precisely that if the maximum is attained at the boundary with vanishing derivative we once again must have a constant function. In either case we conclude that $F(z) = 1$ for all $z \in \mathcal{G}$.

With this information we can go back to the above estimates for (21), which must now be tight. The first step, the triangle inequality (19), is tight if all terms in the sum have the same argument, so up to a common phase the $U_{ij}$ for $i \in s_X$ and $j \in s_Y$ must be positive. With the phase convention (13) this means $U_{ij} > 0$ for all $i, j$ in the supports. The second estimate (20) is only tight when all $U_{ij}$ also have the maximum allowed modulus $c$. Hence $U_{ij} = c$. If we consider $U$ as an operator on vectors with support $s_Y$ it thus maps to constant functions, so $\varphi$ must be constant on $s_X$. By the same token $\widehat{\varphi}$ must be constant on $s_Y$. Taking into account the normalizations we get all assertions of the theorem in the pure case $\rho = |\psi\rangle\langle\psi|$.

It remains to show that all equality states must be pure. So let $\psi$ now be any unit vector in the support of $\rho$ and $\sigma = |\psi\rangle\langle\psi|$. Then we can write $\rho = \lambda\sigma + (1-\lambda)\rho'$ with $\lambda > 0$, $\rho'$ some other state, and similar convex relationships for the probability distributions. By concavity of the entropies, $\sigma$ must also be an equality state. Moreover, by strict concavity, $\sigma$ and $\rho$ must have the same distributions $p_X^\sigma = p_X^\rho$ and $p_Y^\sigma = p_Y^\rho$, and hence the same supports $s_X, s_Y$. Going through the proof for the pure equality state $|\psi\rangle\langle\psi|$, and in particular adopting the phase conventions made for $\psi$ we find that $U_{ij} = c$ for all $i \in s_X$ and $j \in s_Y$. But then, if we apply $U$ to any other unit vector $\psi'$ in the support of $\rho$ we find that $U\psi'$ is constant on its support $s_Y$. Hence $\psi'$ equals $\psi$ up to a phase, the support of $\rho$ is one-dimensional, and $\rho$ must be pure.

An alternative proof of the necessity of purity, at least for the Shannon case $\alpha = \beta = 1$, is via inequality[12]

$$H(p_X^\rho) + H(p_Y^\rho) \geq \log\frac{1}{c^2} + H(\rho). \tag{28}$$

Clearly, for equality states the correction term, the von Neumann entropy $H(\rho)$, has to vanish, i.e., the state must be pure. □

An immediate consequence of Theorem IV.1 is that for most overlap matrices no equality states exist, because $1/c^2$ is not an integer. Since the rows of a unitary matrix must be normalized, this integer is at most $d$, in which case we must have a Hadamard matrix. When $1/c^2 < d$ one can build examples with equality by first solving a unitary matrix completion problem, starting from the known $s_x \times s_Y$ block. One then has to modify the matrix by unitary rotations on the complementary blocks so that all matrix elements become $\leq c$. The lowest-dimensional example is $2 = 1/c^2 < d = 3$, and the overlap matrix

$$U = \begin{bmatrix} a & a & 0 \\ b & -b & a \\ -b & b & a \end{bmatrix} \quad \text{with} \quad a = \frac{1}{\sqrt{2}} \quad \text{and} \quad b = \frac{1}{2}. \tag{29}$$

Some higher-dimensional examples can be generated by replacing the matrix elements $a$ and $b$ by $aU_1$ and $bU_2$, where $U_1, U_2$ are any Hadamard matrices of the same dimension.

By definition, Hadamard matrices have $d$ orthogonal equality states with supports $(|s_X|, |s_Y|) = (1, d)$ and $(d, 1)$, respectively. In prime dimension this is clearly the only possibility. However, even if the dimension is composite there may be no more than this, as the example[33]

$$C_6 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & -\eta & -\eta^2 & \eta^2 & \eta \\ 1 & -\eta^{-1} & 1 & \eta^2 & -\eta^3 & \eta^2 \\ 1 & -\eta^{-2} & \eta^{-2} & -1 & \eta^2 & -\eta^2 \\ 1 & \eta^{-2} & -\eta^{-3} & \eta^{-2} & 1 & -\eta \\ 1 & \eta^{-1} & \eta^{-2} & -\eta^{-2} & -\eta^{-1} & -1 \end{bmatrix} \tag{30}$$

with $\eta = \frac{1-\sqrt{3}}{2} + \mathrm{i}\sqrt{\frac{\sqrt{3}}{2}}$, shows. Here one can mechanically check that none of the 300 $3 \times 2$-submatrices has the property that all elements become equal after multiplication of rows and columns with suitable phases. Hence from Theorem IV.1 it is clear that the point $(\log 3, \log 2)$ on the MU-line is not accessible for any state.

In the special case of a Fourier matrix (see the end of Sect. II for notations) we can get a complete description of the equality cases from Theorem IV.1, as has been observed in Theorem 4.(1) of [29] for the special case of a cyclic group. We will do the same for an arbitrary finite abelian group $J$. It turns out that the equality states are then directly linked to the subgroups of $J$ and its dual $K$. The subgroups always come in pairs, i.e., when $L \subset J$ is a subgroup, so is its annihilator[34]

$$L^\perp = \{k \in K \,|\, \forall j \in L \ \zeta(j,k) = 1\} \subset K. \tag{31}$$

The basic result about annihilators is that $(L^\perp)^\perp = L$ for every subgroup, so there is a ono-to-one correspondence between the subgroups of $J$ and $K$, under which $L_1 \subset L_2 \Leftrightarrow L_1^\perp \supset L_2^\perp$. For any non-empty set $L \subset J$, we denote by $\chi_L$ the $\ell^2$-normalized indicator function, i.e., $\chi_L(j) = |L|^{-1/2}$ for $j \in L$ and $\chi_L(j) = 0$ otherwise.

**Corollary IV.2.** *Let $J$ be a finite abelian group, with Fourier matrix $U$, and $L \subset J$ a subgroup. Then*

$$U\chi_L = \chi_{L^\perp}, \tag{32}$$

*and the vectors of the form $\chi'(j') = \zeta(j',k)\,\chi_L(j'-j)$, where $j \in J/L$ and $k \in K/L^\perp$ are an orthonormal basis so that each $|\chi'\rangle\langle\chi'|$ is an equality state. Moreover, all equality states are of this form.*

Note that in the formula for $\chi'$ we can take arbitrary $j \in J$ and $k \in K$, but two such choices $(j_1, k_1)$ and $(j_2, k_2)$ define the same function $\chi'$ when $j_1 - j_2 \in L$ and $k_1 - k_2 \in L^\perp$. This observation is expressed by taking $j, k$ in the respective quotients.

We remark that, by the fundamental structure theorem of finite abelian groups, every such group is a cartesian product of cyclic groups, and has subgroups of every order which divides $d$ (see Thm. 4.3 in [35]). Hence the equality points on the MU line are *all* points $(\log d_1, \log d_2)$ with $d_1 d_2 = d$.

*Proof.* Let $|\psi\rangle\langle\psi|$ be an equality state. The Theorem then says that for $j \in s_X$, and $k \in s_Y$ we must have $\zeta(j,k) = \mu(k)\nu(j)$ for suitable phase-valued functions $\mu : s_Y \to \mathbb{C}$ and $\nu : s_X \to \mathbb{C}$. Now we can apply translations as in the construction of $\chi'$ in the Corollary to get an equality state with $0 \in s_X$ and $0 \in s_Y$, from which we get $\mu(k)\nu(0) = 1$ and $\mu(0)\nu(j) = 1$, so that the functions $\mu, \nu$ are actually constant. After applying an overall phase factor we can assume without loss of generality, that $\zeta(j,k) = 1$ for $j \in s_X$, and $k \in s_Y$, and that $\psi = \chi_{s_X}$. In terms of annihilators this is expressed equivalently by $s_Y \subset s_X^\perp$ or $s_X \subset s_Y^\perp$.

When $k \in s_X^\perp$ we still have $\zeta(j,k) = 1$ for $j \in s_X$. But then $(U\psi)(k) = (U\psi)(0) > 0$ and we must also have $k \in s_Y$. It follows that $s_X^\perp \subset s_Y$. Combined with the already established reverse inclusion we get that $s_Y = s_X^\perp$ and, symmetrically $s_X = s_Y^\perp$. Note that since any set of the form $A^\perp$ is automatically a subgroup, we have shown that we can take $s_X = L$, $s_Y = L^\perp$ for some subgroup $L \subset J$.

We have so far only shown that $U\chi_L$ is constant on $L^\perp$, namely equal to $\sqrt{|L|/|J|}$, coming from the summation of $|L|$ terms equal to $|L|^{-1/2}$, and observing the overall normalization factor $|L|^{-1/2}$

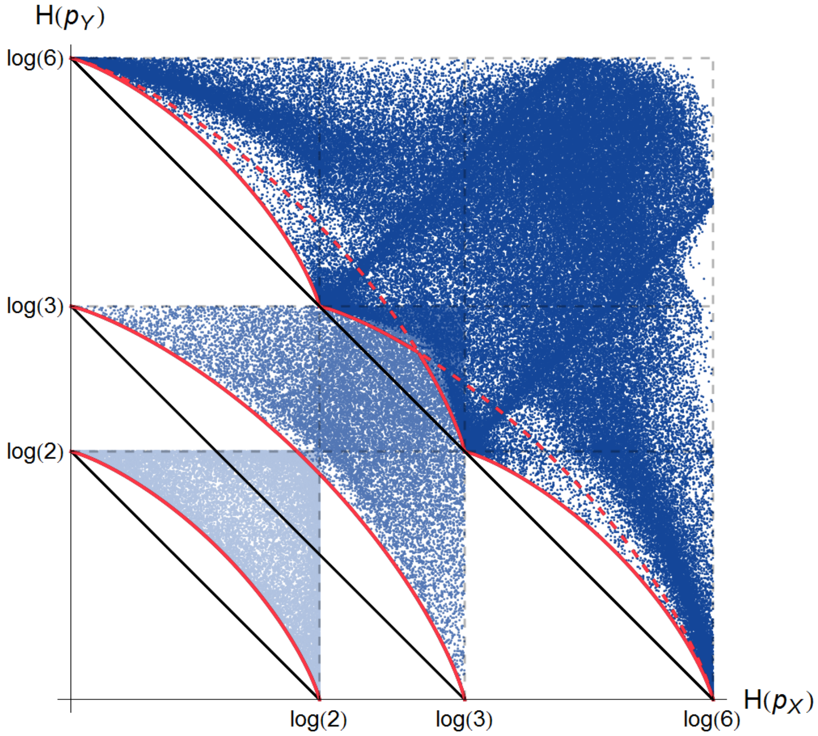


FIG. 3: Numerical sampling of the entropy diagram for dimensions $d = 2$ (light shading), $d = 3$ (medium shading) and $d = 6$ (dark shading) for Fourier-related observables and Shannon entropies. By Theorem IV.1 the number of equality states corresponds to the number of divisors of the respective dimension. The optimal bounds (solid curves) are obtained by applying Conjecture V.6 and Conjecture V.7 presented in Sect. V E.

of the Fourier matrix. We also have to show that $\sum_{j\in L}\zeta(j,k)=0$ whenever $k\notin L^\perp$. However, in that case $k$ induces a non-constant complex homomorphism on $L$, so it suffices to show that such functions add up to 0 on any finite abelian group. However, this is immediately obvious for cyclic groups, and hence follows for arbitrary groups by the structure theorem. So we conclude that $U\chi_L$ is proportional to $\chi_{L^\perp}$, and since $U$ is unitary, it must be equal, and $|L^\perp|\,|L|=|J|$.

Finally, let us count the translates $\chi'$ for a given subgroup. Clearly, they are orthogonal to $\chi_L$ whenever either $j+L\cap L=\emptyset$ or $k+L^\perp\cap L\perp=\emptyset$. In other words, by taking one representative $g$ from each class in $G/H$ and also one $k$ from each class in $K/L^\perp$ we get an orthogonal family. This has $(|J|/|L|)\,(|K|/|L^\perp|)=|J|$, i.e., is an orthonormal basis. □

For a product of abelian groups the Fourier matrix is the tensor product of the Fourier matrices of the factors. Moreover one gets many equality states by tensoring, i.e., by taking subgroups of the form $L_1\times L_2\subset J_1\times J_2$. This additive structure is quite apparent from FIG. 3). It is therefore useful to note that this is also true without assuming the group structure. This is shown by the following result.

**Corollary IV.3.** *Let $U_1,U_2$ be unitary operators of dimension $d_1$ and $d_2$, respectively. Suppose that for each unitary operator there exist an equality state $\sigma^1_{\rm eq}$ and $\sigma^2_{\rm eq}$ as characterized by Theorem IV.1. Then the state $\sigma_{\rm eq}=\sigma^1_{\rm eq}\otimes\sigma^2_{\rm eq}$ is an equality state for the unitary operator $U_1\otimes U_2$.*

*Proof.* First, note that $\max_{j,k}|(U_1\otimes U_2)_{jk}|=\max_{j,k}|U_{1,jk}|\,\max_{j,k}|U_{2,jk}|$. The MU relation then implies that, for any state $\sigma$ on a $d_1\,d_2$-dimensional Hilbert space,

$$H_\alpha(p^\sigma_X)+H_\beta(p^\sigma_Y)\geq -2\log\max_{j,k}|(U_1\otimes U_2)_{jk}| = -2\log\max_{j,k}|U_{1,jk}|\,\max_{j,k}|U_{2,jk}|\ . \tag{33}$$

In particular, for the state $\sigma_{\rm eq}=\sigma^1_{\rm eq}\otimes\sigma^2_{\rm eq}$, we have

$$\begin{aligned}H_\alpha(p_X^{\sigma_{\rm eq}})+H_\beta(p_Y^{\sigma_{\rm eq}}) &= H_\alpha(p_X^{\sigma^1_{\rm eq}})+H_\alpha(p_X^{\sigma^2_{\rm eq}})+H_\beta(p_Y^{\sigma^1_{\rm eq}})+H_\beta(p_Y^{\sigma^2_{\rm eq}})\\ &= -2\log\max_{j,k}|U_{1,jk}|\,\max_{j,k}|U_{2,jk}|\ .\end{aligned} \tag{34}$$

Hence, $\sigma_{\rm eq}$ is an equality state for $U_1\otimes U_2$. □

This Corollary should not be taken to suggest that *only* products will be equality states. For example, take the Fourier matrix of any abelian group of the form $J\times J$, which is the tensor product of two copies of the Fourier matrix of $J$. Then each subgroup $L$ with $|J|$ elements generates a basis of equality states for the point $(\log|J|,\log|J|)$. These are tensor product states for the subgroup $L=\{(j,0)|j\in J\}=J\times\{0\}$. But for $H=\{(j,j)|j\in J\}$ we get a maximally entangled equality state. Again, the basic idea of this example generalizes to more general settings. If $U_1$ is any Hadamard matrix and $\overline{U_1}$ its complex conjugate, the maximally entangled vector $\psi=d^{-1/2}\sum_j|jj\rangle$ is invariant under $U=U_1\otimes\overline{U_1}$. Hence both $\psi$ and $U\psi=\psi$ belong to the equidistribution on $d$ points, and $|\psi\rangle\langle\psi|$ is an equality state with entropies $(\log d,\log d)$, just like $|\phi\rangle\langle\phi|$ with $\phi=d^{-1/2}\sum_j|1j\rangle$.

Perhaps one of the more surprising aspects of Theorem IV.1 is that neither the characterization of the equality states nor indeed the value of the lower bound depends on $\alpha,\beta$. Hence we have

**Corollary IV.4.** *Let $\sigma_{\rm eq}$ be an equality state, i.e. it saturates the uncertainty relation for some $\alpha,\beta>\frac{1}{2}$ satisfying the duality relation. Then $\sigma_{\rm eq}$ is also an equality state for all other pairs $(\alpha,\beta)$ that satisfy the duality relation, including $(\alpha,\beta)=(1/2,\infty)$, $(\infty,1/2)$.*

The boundary cases for the inequality are proved by taking the limits on $(\alpha,\beta)$, and since the lower bound is independent of these, equality carries over. However, additional states may then also

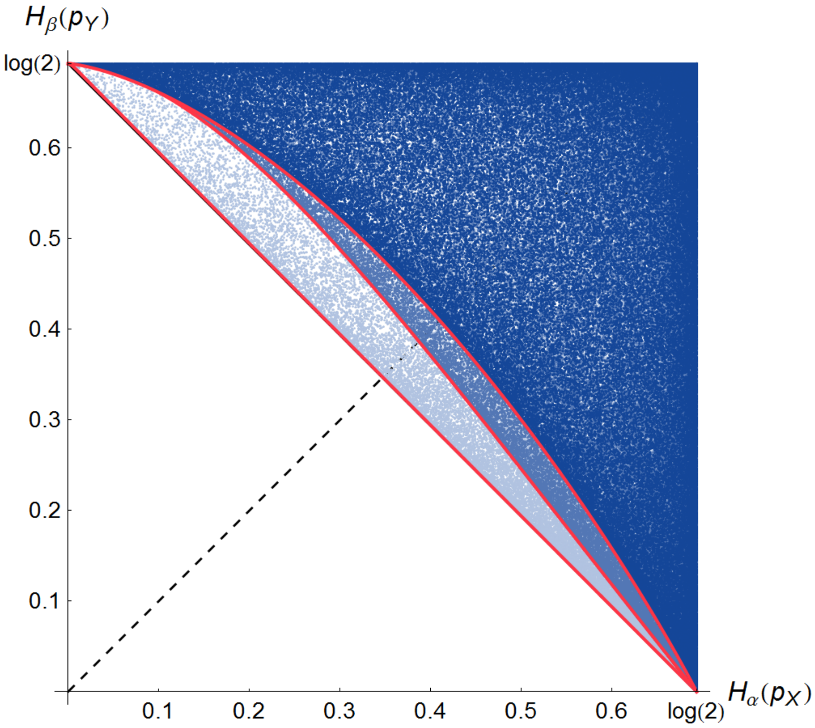


FIG. 4: Typical entropy diagram for Hadamard related observables in prime dimension for different values of $\alpha, \beta$ satisfying the duality relation (2): $\alpha = 1/2$ (light shading), $\alpha = 0.6$ (medium shading) and $\alpha = 0.75$ (dark shading). The MU bound is optimal if and only if $\alpha = 1/2$.

satisfy equality. Indeed, Theorem IV.1 does not hold in this case. As a counterexample consider an arbitrary Hadamard matrix $U$. Without loss of generality we can take it dephased, i.e., with all entries in the first row and column equal to $1/\sqrt{d}$. Consider then some arbitrary state $\psi \in \mathbb{R}^d_+$ with real and positive components to find

$$\max_k |(U\psi)_k|^2 \geq |(\tilde{U}\psi)_1|^2 = \frac{1}{d}\left(\sum_k \psi_k\right)^2 . \tag{35}$$

Taking the logarithm and using the definitions (6) this is equivalent to

$$\log d \geq H_{\frac{1}{2}}(p_X^\psi) + H_\infty(p_Y^\psi), \tag{36}$$

which is $\geq \log d$ by the MU inequality. Hence all such states are equality states, and we can continuously interpolate between $H_{\frac{1}{2}} = 0$ and $H_{\frac{1}{2}} = \log d$. Thus the MU bound coincides with the optimal bound (see FIG. 4) and there is a continuum of equality states in contrast to Theorem IV.1.

Another feature is true only in the boundary case, namely that for *every* $U$ there is an equality state. To see this, let us consider an eigenstate $x_j$ of $X$, for which $H_{1/2}(p_X^{x_j}) = 0$. But at the same time we have

$$\min_j H_\infty(p_Y^{x_j}) = \min_j(-\log\max_k |\langle y_k|x_j\rangle|^2) = -2\log c . \tag{37}$$

One could summarize this by saying that in the boundary case $\{\alpha, \beta\} = \{1/2, \infty\}$ the MU bound is just too good to allow a useful characterization of equality.

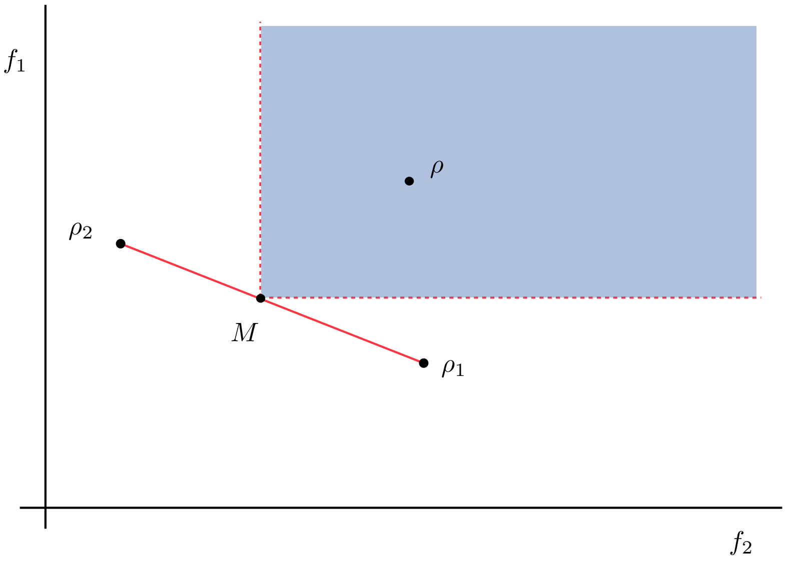


FIG. 5: Consequences of concavity for the set of entropy pairs.

## V. CHARACTERISATION OF THE CURVE OF MINIMAL ENTROPY PAIRS

Due to the study of equality in the previous section it is clear that the MU bound is, in almost all cases, not optimal, i.e. it does not coincide with the curve of minimal entropy pairs. To characterize this optimal bound is the aim of this section. We establish three general results that hold for arbitrary dimension: First, we prove that the curve of minimal entropies can be parametrized by pure states. Second, we show that for all real-valued unitary operators we can restrict the problem to real states. And last, we establish a necessary criterion for the Fourier case which all optimal states must satisfy thereby being able to characterize a whole class of potentially optimal states. Additionally, we provide a complete characterisation of the optimal bound for the simplest case of two-dimensional state space, $d = 2$. For $d = 3$ there is an analytic expression[8], which is well-confirmed by numerics, although not proved. However, for higher dimensions the optimal bound remains unknown. Nevertheless, we present random samples that suggest a number of conjectures, which, if true, vastly simplify the characterisation of the optimal bound.

### A. Sufficiency of pure states

In this section we show that the optimal bound can be parametrized by pure states. At a first glance, this result may seem not too surprising since the situation is clear when minimizing only one concave functional $f(\rho)$ over all states: In this case one can immediately restrict to pure states, since one of the convex components $\rho'$ of $\rho$ must always give a value $f(\rho') \leq f(\rho)$. However, the situation is not so simple when we consider a pair of concave functions, and the image of the state space under a two-component mapping $f = (f_1, f_2)$ as in (7). The direct consequence of concavity is then that for, say $\rho = (\rho_1 + \rho_2)/2$, the point $f(\rho)$ lies above the midpoint $M = \big(f(\rho_1) + f(\rho_2)\big)/2$ in the coordinatewise ordering, i.e., $f_i(\rho) \geq \big(f_i(\rho_1) + f_i(\rho_2)\big)/2$ for $i = 1, 2$ (see FIG. 5). We therefore cannot conclude that the set $\{f(\rho)\}$ is convex: the midpoint $M$ is not in general in the set. Indeed this is clearly shown by the entropy diagrams, from which it is also clear that the complement is not convex either, except in simple cases.

For the same reasons it is not obvious that it is sufficient to restrict to pure states. This is highlighted by looking at the problem a bit more generally, considering the pairs of probability distributions in two bases.

**Proposition V.1.** *Consider two orthonormal bases $X, Y$ in a Hilbert space and let $p_X^\rho, p_Y^\rho$ denote the respective probability distributions in the state $\rho$. Then*

- *If $d = 2$, then for every state $\rho$ there is pure state $\sigma$ such that $p_X^\rho = p_X^\sigma$ and $p_Y^\rho = p_Y^\sigma$.*
- *If $d \leq 3$, then for every $\rho$ we can find a convex decomposition $\rho = \sum_i \lambda_i \sigma_i$ into pure states $\sigma_i$ with $p_X^\rho = p_X^{\sigma_i}$ for all $i$.*

*For larger dimensions both statements fail.*

Thus, for $d = 2$ the range $\{f(\rho)\}$ is already exhausted by pure states, and for $d = 3$ the monotone closed uncertainty diagram can be computed just with pure states. For if $f(\rho)$ is any point in the diagram, we can decompose into the $\sigma_i$, without any increase of $f_1$, so by concavity we know one of the pure components has smaller $f_2$. However, this proof strategy will fail for $d \geq 4$.

*Proof.* (1) For $d = 2$, the set of quantum states $\rho$ with the same distribution $p_X^\rho$ is the intersection of the Bloch ball with a hyperplane. Intersecting with the hyperplane for $p_Y^\rho$ we get a line, which also intersects the Bloch sphere, i.e., there is a pure state with the same distributions.

(not 1) The example uses Fourier transform in $d = 3$. Two density operators have the same position distribution iff their diagonals coincide and the same momentum distribution iff the sums $\sum_x \langle x|\rho|x + y\rangle$ coincide for all $y$. Now consider a diagonal matrix with diagonal entries $(1, 1, 0)/2$. A pure state with this diagonal will have just one non-zero phase in the 1-2 matrix element, so the sum with $y = 1$ will be non-zero other than for the mixed state.

(2) Let us consider the convex subset $K(p)$ of states with $p_X^\rho = p$. We have to show that for $d = 3$ all extreme points of this set are, in fact, pure. Our method will also show that this fails for $d \geq 4$.

First observe by just conjugating with a positive diagonal operator from right and left we get an isomorphism of $K(p)$ and $K(q)$, as long as $p, q$ have the same support (of size $d$). So we may as well take $p$ to be uniform, for which we write $K(1)$ (Normalization factors are irrelevant here).

Let us sort the potential extreme points by rank. Full rank is not possible, since then *any* vector with uniform distribution could be subtracted with a positive weight. Rank 1 is uninteresting, because it is of the form we want to exclude. This takes care of $d = 2$ and leaves only the rank 2 case for $d = 3$.

So let us consider the case of rank 2 for general $d$. Let $\phi_1, \phi_2$ be two linearly independent vectors in the range of the density operator $\rho = |\phi_1\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2|$. The condition that $\rho$ has uniform position distribution means that $|\phi_1(x)|^2 + |\phi_1(x)|^2 = 1$ for all $x$. In other words, the pair $\Phi(x) = \big(\phi_1(x), \phi_2(x)\big) \in \mathbb{C}^2$ is a unit vector for every $x$ . Then we ask whether there is any non-zero vector $\Psi \in \mathbb{C}^d$ of the form $\Psi(x) = \overline{\alpha_1}\phi_1(x) + \overline{\alpha_2}\phi_2(x)$ such that $|\Psi(x)| = 1$ for all $x$. This would be a convex component of $\rho$ with even distribution, so we could further decompose $\rho$.

We can read this as a scalar product $|\langle\alpha, \Phi(x)\rangle|^2$. Think of the $\Phi(x)$ and of $\alpha$ as represented on the Bloch sphere, where the geodesic distance is just a function of the above scalar product. So our question reduces to: Given $d$ vectors on the sphere, can we find one further vector which has the same distance from each of them?

Now for $d = 2$ this is obvious, and for $d = 3$ it works just like in the planar geometry of triangles: The locus of all points which have the same distance from $\Phi(1)$ and $\Phi(2)$ is a great circle bisecting their connecting geodesic at a right angle. Intersect with the bisector for $\Phi(2)$ and $\Phi(3)$, which gives a point which has the same distance from all three points. Therefore, for $d = 3$, there are no extreme points of rank 2, hence all are of rank 1 as claimed.

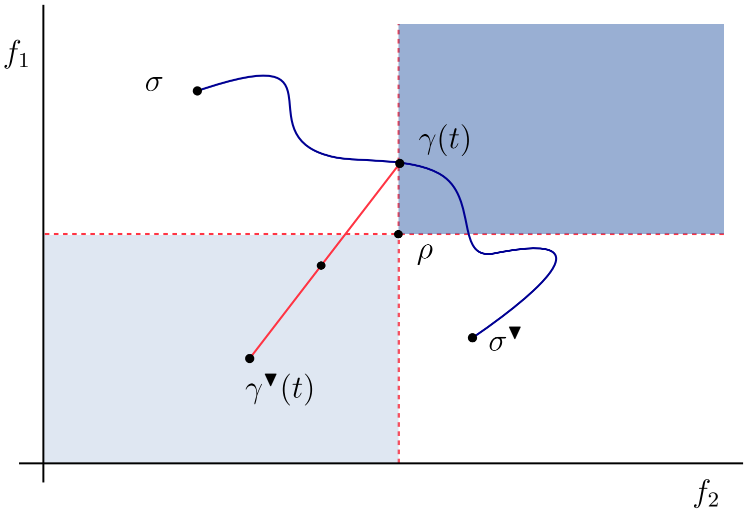


FIG. 6: States appearing in the proof of Theorem V.2 as mapped to the two entropies plane.

For higher $d$ it is easy to find $d$ points, which do not lie on a circle, i.e., there is no point equidistant from all of them. Hence there are extreme points of $K(1)$ of rank 2.

□

Surprisingly however, pure states can be shown to saturate all uncertainty diagrams, practically without assumptions on $X, Y, \alpha, \beta$.

**Theorem V.2.** *Let $f_1, f_2$ be continuous concave functionals on the state space, define the order relation $\sqsubseteq$ as after equation (7). Then for every state $\rho$ there is a pure state $\sigma$ such that $\sigma \sqsubseteq \rho$.*

*Proof.* The plan of the proof is to show that for every non-pure $\rho$ we can find another state $\sigma$ of strictly smaller rank such that $\sigma \sqsubseteq \rho$. Then we can successively lower the rank, arriving finally at a pure state.

Consider the face $F$ of the state space generated by $\rho$. Its topological boundary $\partial F$ consists precisely of the possible convex components of $\rho$ of lower rank, and is connected. For each point $\sigma \in \partial F$ there is a unique "antipode" $\sigma^\blacktriangledown$. It is defined as

$$\sigma^\blacktriangledown = \frac{1}{\lambda}\Big(\rho - (1-\lambda)\sigma\Big) \tag{38}$$

for the smallest $\lambda$ for which the right hand side is positive semidefinite. It is clearly a state of reduced rank, i.e., $\sigma^\blacktriangledown \in \partial F$. We note that the required weight $\lambda$ cannot be 0 or 1.

We need not consider the case that $\sigma \sqsubseteq \rho$, since otherwise we have found the desired element. Therefore, by exchanging the functions $f_1$ and $f_2$ if necessary, we may assume that $f_1(\sigma) > f_1(\rho)$. We cannot also have $f_1(\sigma^\blacktriangledown) \geq f_1(\rho)$. Indeed, this would lead to the contradiction

$$f_1(\rho) \geq (1-\lambda)f_1(\sigma) + \lambda f_1(\sigma^\blacktriangledown) > f_1(\rho). \tag{39}$$

Now consider a continuous curve $[0,1] \ni t \mapsto \gamma(t) \in \partial F$ connecting $\sigma$ and $\sigma^\blacktriangledown$, i.e., such that $\gamma(0) = \sigma$ and $\gamma(1) = \sigma^\blacktriangledown$ (see FIG. 6). Since $f_1$ was assumed to be continuous the previous argument shows that, for some $t$, $f_1\big(\gamma(t)\big) = f_1(\rho)$.

If $f_2\big(\gamma(t)\big) \leq f_2(\rho)$ we have found the desired element $\gamma(t) \sqsubseteq \rho$. The non-trivial case to consider is therefore $f_2\big(\gamma(t)\big) > f_2(\rho)$, or $\rho \sqsubseteq \gamma(t)$. Let $\lambda \in (0,1)$ be the weight so that $\rho =$

$(1-\lambda)\gamma(t)+\lambda\gamma(t)^{\blacktriangledown}$. Then by concavity, for $i=1,2$,

$$\begin{aligned} f_i(\rho) &\geq (1-\lambda)f_i\big(\gamma(t)\big)+\lambda f_i(\gamma(t)^{\blacktriangledown}) \\ &\geq (1-\lambda)f_i(\rho)+\lambda f_i\big(\gamma(t)^{\blacktriangledown}\big) \\ \text{i.e.,}\quad f_i(\rho) &\geq f_i\big(\gamma(t)^{\blacktriangledown}\big). \end{aligned} \tag{40}$$

Therefore $\gamma(t)^{\blacktriangledown} \sqsubseteq \rho$.

□

### B. Sufficiency of real states for real unitary matrices

From the previous section we know that for all unitary operators the complete optimal bound can be parametrized by pure states. Now we show that if the unitary matrix linking the two observables is real-valued, then we can further restrict the set of states for the complete optimal bound to the set of real-valued vectors. In this whole subsection we fix the Hilbert space to be $\mathbb{C}^d$ with componentwise complex conjugation, so that the real vectors $\mathbb{R}^d \subset \mathbb{C}^d$ are naturally embedded.

**Theorem V.3.** *Let $f_1, f_2$ be continuous concave functionals on the state space and their inputs linked by a real unitary operator $U_{\text{real}}$. Also define the order relation $\sqsubseteq$ as after equation (7). Then for every state $\rho$ there is a pure and real state $\sigma$ such that $\sigma \sqsubseteq \rho$.*

*Proof.* The idea of the proof is to employ again the proof technique of Theorem V.2, i.e. decompose a state in two states with the desired property (in this case, real states) and use the concavity property of the functions.

Let $\psi \in \mathbb{C}^d$ be a pure state. Since we are interested in a decomposition into real states, it is natural to consider the decomposition

$$\psi = \sqrt{\lambda}v + \mathrm{i}\sqrt{1-\lambda}w \tag{41}$$

where $v, w \in \mathbb{R}^d$ are the normalized real and imaginary part of $\psi$ and $\lambda = |\operatorname{Re}(\psi)|^2$ ranges from 0 to 1. We are only interested in the case where neither $v \sqsubseteq \psi$ nor $w \sqsubseteq \psi$, otherwise the statement follows immediately. Furthermore, we assume without loss of generality that $f_1(v) > f_1(\psi)$. Similar to the proof in Theorem V.2 we cannot also have that $f_1(w) > f_1(\psi)$ because we would then find the contradiction

$$f_1(\psi) \geq \lambda f_1(v) + (1-\lambda)f_1(w) > f_1(w)\ . \tag{42}$$

Consider now the states

$$\varphi(t) := \mathrm{e}^{\mathrm{i}t}\psi \tag{43}$$

and their normalized real and imaginary part

$$\begin{aligned} \gamma(t) &:= \operatorname{Re}\big(\varphi(t)\big)/|\operatorname{Re}\big(\varphi(t)\big)|\ , \\ \sigma(t) &:= \operatorname{Im}\big(\varphi(t)\big)/|\operatorname{Im}\big(\varphi(t)\big)| \end{aligned} \tag{44}$$

such that

$$\varphi(t) = \sqrt{\mu(t)}\gamma(t) + \mathrm{i}\sqrt{1-\mu(t)}\sigma(t)\ , \tag{45}$$

where $\mu(t) = ||\gamma(t)||$. Note that $f_i(\varphi(t)) = f_i(\psi)$ for all $t \in (0, 2\pi)$. Also note that for a real-valued unitary operator the probability distributions $p_X^{\varphi(t)}$ and $p_Y^{\varphi(t)}$ have the same form

$$p_{X/Y}^{\varphi(t)} = \mu(t) p_{X/Y}^{\gamma(t)} + \big(1 - \mu(t)\big) p_{X/Y}^{\sigma(t)} \; . \tag{46}$$

Due to continuity we know that there exists $t_0$ such that either $\gamma(t_0) \sqsubseteq \psi$, from which we obtain the desired statement, or $\psi \sqsubseteq \gamma(t_0)$. Using the concavity of the functions $f_i$, the latter then implies

$$\begin{aligned} f_i(\psi) = f_i\big(\varphi(t_0)\big) &\geq \mu(t_0) f_i\big(\gamma(t_0)\big) + \big(1 - \mu(t_0)\big) f_i\big(\sigma(t_0)\big) \\ &\geq \mu(t_0) f_i(\psi) + \big(1 - \mu(t_0)\big) f_i\big(\sigma(t_0)\big) \; , \end{aligned} \tag{47}$$

from which obtain $f_i\big(\sigma(t_0)\big) \leq f_i(\psi)$, or equivalently $\sigma(t_0) \sqsubseteq \psi$. □

### C. Variatonal method

So far we characterized the optimal bound by the order relation $\sqsubseteq$. Equivalently, we may also consider an optimisation problem as mentioned in (8): Given some fixed value of $H_\beta(p_Y^\rho) = \delta$ the optimal bound $\gamma$ is described by minimising $H_\alpha(p_X^\rho)$, i.e.

$$\gamma(\delta) = \min_\rho \{ H_\alpha(p_X^\rho) | H_\beta(p_Y^\rho) = \delta \} \; , \tag{48}$$

where $\delta$ ranges from 0 to $\log d$. However, performing this optimisation is in general quite difficult, especially because a nice characterisation of the constant entropy set $\{\rho | H_\beta(p_Y^\rho) = \delta\}$ is not known. Instead, we restrict to optimising over a subset of this constant entropy set, namely states with varied phases. Clearly, this method will not yield a sufficient criterion for a state to be optimal. However, it provides us with a necessary criterion which allows us to identify a whole class of candidates of optimal states.

More concretely, using Theorem V.2 we consider pure states $\varphi \in \mathbb{C}^d$ and denote the components of the phase-varied state in $Y$ basis by

$$\psi_j = \varphi_j \exp\left(\frac{2\pi \mathrm{i}}{d} \, \theta_j\right) \tag{49}$$

for some phases $\theta_j$. Varying these phases does not change the probability distribution, $p_Y^\psi = p_Y^\varphi$, and hence the phase varied states form a subset of the constant entropy set. For observables linked by Fourier transformation, we can optimize $H_\alpha(p_X^\psi)$ over these states to find the following extremality criterion:

**Lemma V.4.** *Let the two observables $X$ and $Y$ be linked by the Fourier matrix* (9) *and let $\psi$ denote an optimal state of this setup. Furthermore, let $\hat{\psi}$ denote the Fourier transform of $\psi$. Then $\psi$ satisfies*

$$\mathrm{Im}\left( \psi_k \sum_{j=1}^{d} \frac{\partial H_\alpha(p_X^\psi)}{\partial |\hat{\psi}_j|^2} \overline{\hat{\psi}_j} \exp\left(\frac{2\pi \mathrm{i} jk}{d}\right) \right) = 0 \quad \forall k. \tag{50}$$

*Proof.* In order to optimize $H_\alpha(p_X^\psi)$ we compute

$$\left.\frac{\partial H_\alpha(p_X^\psi)}{\partial \theta_k}\right|_{\theta=0} = \sum_{j=1}^{d} \left.\frac{\partial H_\alpha(p_X^\varphi)}{\partial |\hat{\psi}_j|^2} \frac{\partial |\hat{\psi}_j|^2}{\partial \theta_k}\right|_{\theta=0} \overset{!}{=} 0 \; . \tag{51}$$

With $\omega := \exp\left(\frac{2\pi \mathrm{i}}{d}\right)$ the Fourier transform of $\psi$ is defined as $\hat{\psi}_j := \frac{1}{\sqrt{d}} \sum_{m=1}^{d} \psi_m \omega^{jm}$ and, hence,

$$|\hat{\psi}_j|^2 = \frac{1}{d} \sum_{m,n=1}^{d} \varphi_m \overline{\varphi_n}\ \omega^{j(m-n)+\theta_m-\theta_n}\ . \tag{52}$$

Therefore we have

$$\begin{aligned} \left.\frac{\partial|\hat{\psi}_j|^2}{\partial\theta_k}\right|_{\theta=0} &= \frac{1}{d} \sum_{m,n=1}^{d} \varphi_m \overline{\varphi_n} \omega^{j(m-n)+\theta_m-\theta_n} \Bigg|_{\theta=0} \\ &= \frac{2\pi \mathrm{i}}{d^2} \operatorname{Im}\left(\varphi_k \overline{\hat{\varphi}_j} \omega^{jk}\right) \end{aligned} \tag{53}$$

Combining (51) and (53) we obtain the desired statement. □

Any optimal state must necessarily satisfy the above criterion. This allows us to characterize a whole class of potentially optimal states:

**Lemma V.5.** *Let $\varphi$ be a real-real symmetric state, i.e. a real state, $\varphi \in \mathbb{R}^d$, satisfying the symmetry condition*

$$\varphi(j) = \varphi(d-j) \quad \forall j = 1, ..., d-1 \tag{54}$$

*or, equivalently, a real state with real Fourier transform, $\hat{\varphi} \in \mathbb{R}^d$. Then $\varphi$ satisfies the extremality criterion* (50)*.*

*Proof.* We first note a simple, but important property of real-real symmetric states: If $\varphi$ is a real-real symmetric state and $\xi$ is a state with components $\xi_j = f(\varphi_j)$, where $f$ is any function taking real numbers to real numbers, then $\xi$ is also a real-real symmetric state. For example, the Fourier transform of any real-real symmetric state is also real-real symmetric.

Now $\varphi$ is assumed to be real-real symmetric. Hence, $\hat{\varphi}$ is real-real symmetric. Define

$$\xi_j := \frac{\partial H_\alpha(p_X^\psi)}{\partial|\hat{\varphi}_j|^2} \hat{\varphi}_j \tag{55}$$

and note that $\xi$ is also real-real symmetric. Importantly this implies that its Fourier transform, $\hat{\xi}$ is real. We therefore have for all $k$

$$\operatorname{Im}\left(\varphi_k \sum_{j=1}^{d} \frac{\partial H_\alpha(p_X^\psi)}{\partial|\hat{\varphi}_j|^2} \overline{\hat{\varphi}_j} \exp\left(\frac{2\pi \mathrm{i} jk}{d}\right)\right) = \operatorname{Im}\left(\sum_{j=1}^{d} \xi_j \exp\left(\frac{2\pi \mathrm{i} jk}{d}\right)\right) = \operatorname{Im}\left(\hat{\xi}\right) = 0\ , \tag{56}$$

which finishes the proof. □

### D. Simplest case: $d = 2$

The results we presented so far are not sufficient to provide a complete characterisation of the curve of minimal entropy pairs. In what follows we therefore restrict to small dimension in order to reduce the complexity of the problem.

More concretely, we investigate the simplest case, where the dimension of the Hilbert space is $d = 2$. In [26, 27] the authors characterized the curve of minimal entropy pairs for all unitary operators while restricting to the case of Shannon entropies. We now generalize their result to

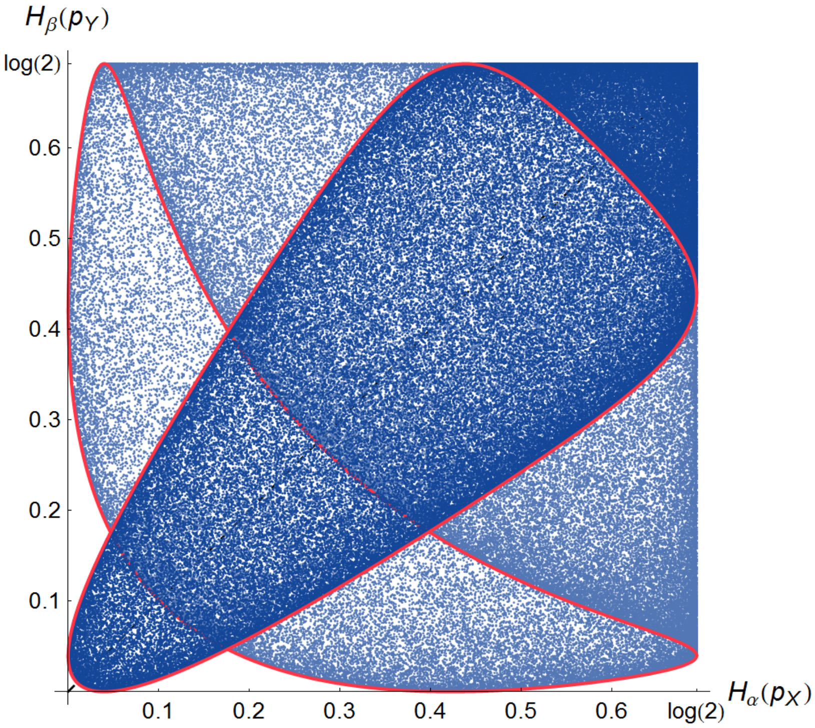


FIG. 7: The optimal bound can be completely characterized in the qubit case (solid curves). The plot illustrates two entropy diagrams for randomly chosen unitary operators and entropy-pairs with $\alpha = \beta = 10$ (light shading) and $\alpha = \beta = 8$ (dark shading).

arbitrary pairs of Rényi entropies: First we show that for each $2 \times 2$ unitary operator $U$ there is a real unitary operator $\tilde{U}$ with the same entropy diagram. Then from Theorem V.3 we can immediately infer that the lower bound can be parametrized by real states. More concretely, our aim is to show that any unitary operator, which we can always write in $\{x_i\}$ basis up to an (irrelevant) global phase as

$$U = \begin{pmatrix} \cos(\varphi) & \sin(\varphi)\mathrm{e}^{-\mathrm{i}\theta} \\ -\sin(\varphi)\mathrm{e}^{\mathrm{i}\theta} & \cos(\varphi) \end{pmatrix} , \tag{57}$$

is equivalent to the matrix

$$\tilde{U} = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix} . \tag{58}$$

Indeed, the entropy diagram does not change if we first modify the unitary operator to $U' = UV$ if $V$ is a unitary operator satisfying $Vx_i = \exp(\mathrm{i}\varphi_i)x_i$ for some phases $\varphi_i$ and all $i$, since then for any state $\rho$ there exists a state $\rho'$ that yields the same pair of entropies. To see this, let $\rho' = V^\dagger \rho V$ to find that

$$p_X^{\rho'}(i) = \langle x_i|\rho'|x_i\rangle = \langle x_i|V^\dagger \rho V|x_i\rangle = \langle x_i|\rho|x_i\rangle = p_X^{\rho}(i) \tag{59}$$

and

$$p_{Y'}^{\rho'}(j) = \langle y_j'|\rho'|y_j'\rangle = \langle y_j|VV^\dagger \rho VV^\dagger|y_j\rangle = \langle y_j|\rho|y_j\rangle = p_Y^{\rho}(j) \ . \tag{60}$$

Now consider the modification

$$U' = \begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi)\mathrm{e}^{\mathrm{i}\theta} & \cos(\varphi)\mathrm{e}^{\mathrm{i}\theta} \end{pmatrix} \tag{61}$$

obtained via the unitary operator

$$V_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} \tag{62}$$

However, $U'$ yields exactly the same probability distribution as $\tilde{U}$. Hence, by Theorem V.3 the curve of minimal entropy pairs can be parametrized by real states.

Since the real states form a one-parameter family it is not difficult to check that the states

$$\psi = \big(\cos(\xi), \sin(\xi)\big) \ , \tag{63}$$

where the range of $\xi$ is either $(0, \arccos(|U_{1,1}|)$ or $(\arccos(|U_{1,1}|), \pi/2)$ depending on whether $\arccos(|U_{1,1})| \in (\pi/4, 3\pi/4)$ or not, parametrize the curve of minimal entropy pairs for all unitary operators and all Rényi entropies. The problem is therefore completely solved in the simplest case $d = 2$ (see FIG. 7).

### E. Numerical sampling and conjectures

In the previous section we characterized the optimal bound in the special case of dimension $d = 2$. To the best of our knowledge the problem is unsolved for all other dimensions. Instead the authors of [8] provide a conjecture stating that the curve of minimal entropies is traced out by states of the form

$$\psi = (\sqrt{p_2}, \sqrt{p_2}, ..., , \sqrt{p_2}, \sqrt{p_1})^{\mathsf{T}} \tag{64}$$

with $p_1 + (d-1)p_2 = 1$ in the case of complex Hadamard matrices and Shannon entropies. Due to the results of [26, 27] it is clear that this conjecture is correct for $d = 2$. The conjecture also holds true in the case $d = 3$ if we trust the numerics presented in FIG. 3, where the solid curve directly corresponds to the states (64). However, for $d = 4$ we show that the conjecture already fails: For complex Hadamard matrices $c = 1/\sqrt{d}$ and, hence, according to our analysis of equality in the MU bound there must be three distinct equality points, whereas the conjectured states only yield two equality points (see FIG. 8).

However, we present two different conjectures which, if correct, explain how the bound in FIG. 3 and 8 can be obtained:

**Conjecture V.6.** *(Product states for matrices with product form)*
*Let the unitary operator $U$ linking the two observables be a matrix of the form $U = U_1 \otimes U_2$. Then for any state $\rho$ there exists a product state $\rho_1 \otimes \rho_2$ with the same pair of entropies.*

The consequence of this our first conjecture is that the curve of minimal entropies for product form unitary operators in some composite dimension $d = d_1 d_2$ is just comprised of tensor products of states that parametrize the curve in dimension $d_1$ and $d_2$, respectively. Indeed, from the additivity of the Rényi entropy it then directly follows that a state $\rho_d = \rho_{d_1} \otimes \rho_{d_2}$ is optimal with respect to the unitary operator $U = U_{d_1} \otimes U_{d_2}$ if and only if the marginals $\rho_{d_1}$ and $\rho_{d_2}$ are optimal with respect to the unitary operators $U_{d_1}$ and $U_{d_2}$, respectively. We note that this conjecture also agrees with our findings for the equality states, especially Corollary IV.3.

**Conjecture V.7.** *(Decomposition of the Fourier matrix)*
*Let the two observables be linked by the Fourier matrix $U_d^F$ of composite dimension $d = d_1 d_2$. Then the entropy diagram does not change if we replace $U_d^F$ by $U_{d_1}^F \otimes U_{d_2}^F$.*

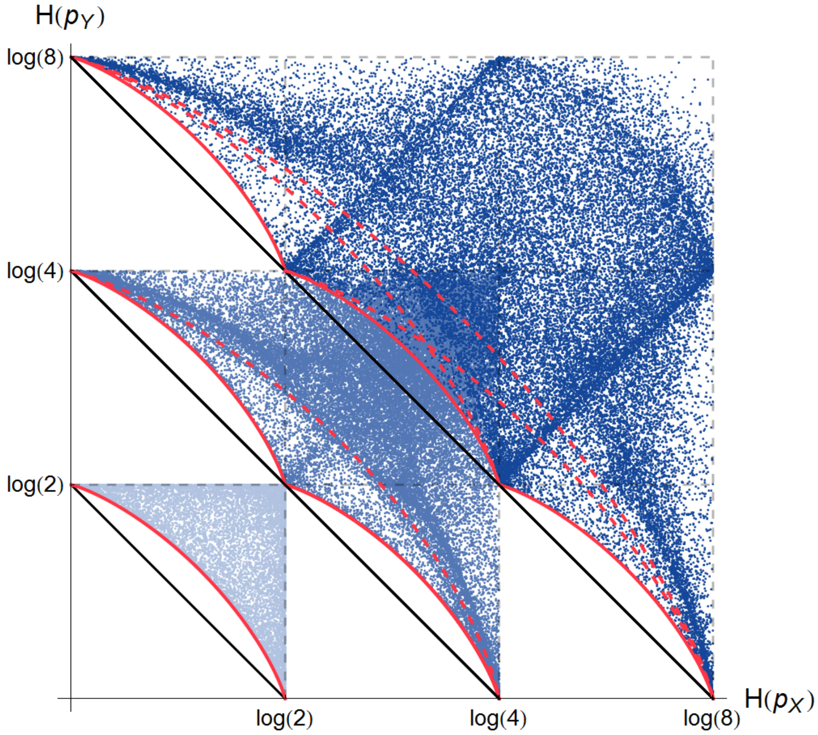


FIG. 8: Random sample of the entropy diagram for dimensions $d = 2$ (light shading), $d = 4$ (medium shading) and $d = 8$ (dark shading) for Fourier related observables and Shannon entropies. Our results falsify a previous conjecture by Englert *et al.* (dashed curves). Instead the optimal bounds are given by the solid curves, which are obtained by applying Conjecture V.6 and Conjecture V.7.

The consequence of this second conjecture is that, although the Fourier matrix can, in general, not be decomposed into a tensor product of Fourier matrices of smaller dimension, the entropy diagram (and hence the curve of minimal entropy pairs) does not change under this replacement. Hence, if this conjecture were true, we could apply Conjecture V.6 and characterize the curve of minimal entropy pairs by states of product form, where the marginals parametrize the optimal bound in the respective smaller dimension.

As an example let us consider Fourier related observables in dimension $d = 4$. Employing both conjectures we know that it suffices to consider only the problem of characterising the optimal bound for Fourier related observables in dimension $d = 2$. But for such observables we already characterized the bound completely (see Sect. V D) and, hence, the optimal bound in $d = 4$ is traced out by product states with marginals given by (63). Indeed, this result agrees with the random sample (FIG. 8). In FIG. 3 we also show other examples, where the numerics validate the two conjectures above.

Note that the above conjectures are statements about the case of composite dimension, effectively stating that for a large class of unitary operators one only needs to solve the problem in prime dimension. The prime-dimensional case, however, still remains a hard problem. But we can provide two further conjectures that, if correct, vastly reduce the complexity of calculating the optimal bound in these instances:

**Conjecture V.8.** *(Independence of the optimal states of $(\alpha, \beta)$)*
*If $\rho$ is an optimal state for any unitary operator and any $\alpha, \beta > \frac{1}{2}$ satisfying the duality relation* (2)*, then $\rho$ is also an optimal state for all other dual pairs.*

This conjecture can be seen as an extension of Corollary IV.4. Note that we again excluded

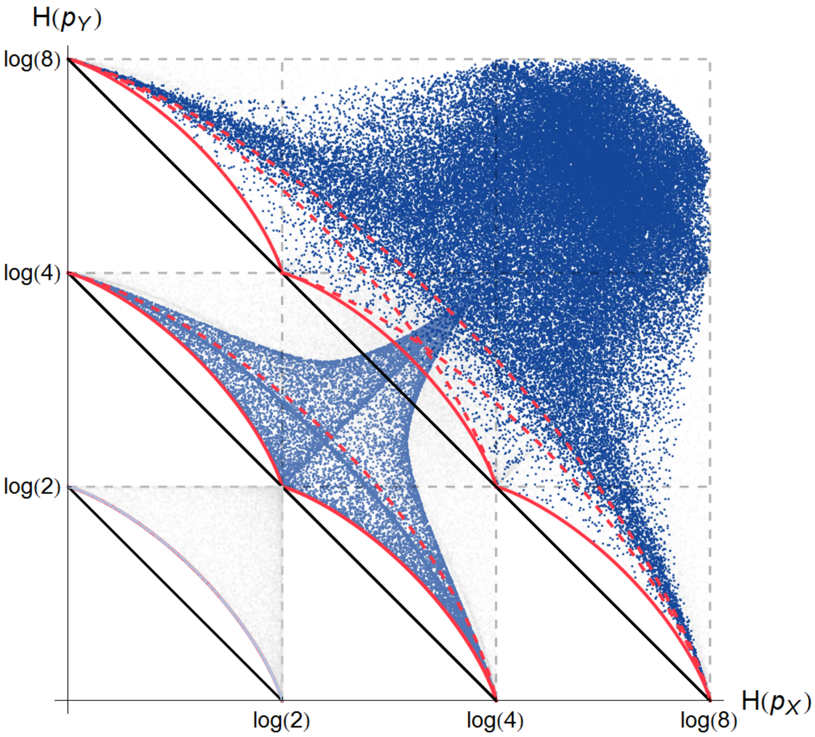


FIG. 9: Random sample of the entropy diagram for real-real symmetric states in dimensions $d = 2$ (light shading), $d = 4$ (medium shading) and $d = 8$ (dark shading) for Fourier related observables and Shannon entropies. Restricting to real-real symmetric states does not yield the complete entropy diagram (grey), but seems to be sufficient to characterize the optimal bound.

the extremal case $\{\alpha, \beta\} = \{1/2, \infty\}$ for the same reasons as explained in Sect. IV. In FIG. 4 the optimal bounds, although differently shaped, are traced out be the same states which supports Conjecture V.8.

The last conjecture only considers the case of observables linked by the Fourier matrix.

**Conjecture V.9.** *(Sufficiency of real-real symmetric states for Fourier)*
*If $\rho$ is an optimal state for the Fourier case, then there is a real-real symmetric state $\sigma$ as given by* (54) *with the same entropy pair.*

According to this conjecture it is sufficient to analyse the problem only for real-real symmetric states, which yields a huge simplification in both analytical and numerical treatments of the problem. As an example consider Fourier related observables in dimension $d = 3$. If Conjecture V.9 were correct, we already knew a characterisation of the optimal bound, since the real-real symmetric states in this case form a one-parameter family and therefore trace out the desired curve. Indeed, for $d = 3$ the real-real symmetric states coincide with the states conjectured by [8] which, as mentioned above, trace out the bound if we trust numerics. FIG. 9 also suggests the validity of Conjecture V.9.

Furthermore, we note that real-real symmetric states are closed under the tensor product, in the sense that any tensor product of two real-real symmetric states is again a real-real symmetric state. Hence, Conjecture V.6 and Conjecture V.9 agree with each other.

## VI. CONCLUSION AND OUTLOOK

We investigated the curve of minimal entropies that completely describes the entropic uncertainty tradeoff between two observables. We showed that the lower bound on the sum of two entropies as given by the Maassen-Uffink uncertainty relation is not optimal in almost all cases and hence does not correspond to the curve of minimal entropies. To show this, we presented a novel proof of the MU bound that allowed us to analyse the case of equality in the uncertainty relation.

In order to characterize the curve of minimal entropies, we provided three main results: First, we showed that the optimal bound can be traced out by pure states. Second, the optimal bound for real-valued unitary operators can be traced out by real-valued pure states. And last, we presented an extremality criterion, which any optimal state must satisfy. Numerical and analytical results for the case of small dimension suggest a number of conjectures that, if true, lead to a drastic reduction of the optimisation space. The optimal lower bound could then be computed.

### Acknowledgements

The work in Singapore is funded by the Singapore Ministry of Education (partly through the Academic Research Fund Tier 3 MOE2012-T3-1-009) and the National Research Foundation of Singapore. F.F. acknowledges support from LUH GRK 1463 and the Japan Society for the Promotion of Science (JSPS) by KAKENHI grant No. 24-02793. R.S. acknowledges support from the BMBF funded network Q.com-Q.

# CHAPTER 8

## Continuous systems: a modified Heisenberg algebra

In this chapter we will investigate preparation uncertainty relations of a simple model for physics in the regime of a hypothetical Planck scale. Various theories of quantum gravity propose the existence of a so called *minimal length* that sets an ultimate lower scale to physics. Regarded form the quantum perspective of quantum-gravity, this minimal length should manifests itself as a lower bound on uncertainty of position. This circumstance is sometimes called *the generalized uncertainty principle* (GUP).

In the following section we will consider a class of modifications of the Heisenberg algebra, which was suggested, among others, by string theory. We derive a list of sufficient criteria, those modifications have to obey, in order to lead to a minimal length/ position uncertainty. In opposition to other works from this field, we do not insist on self adjoint position operators, rather than that we will show that a position measurement, acting on physical states, is well described by an inherent noisy POVM, which naturally leads to a lower bound on any preparation uncertainty relation. We will investigate this minimal uncertainty in terms of variances, Shannon entropies and min entropies. Thereby, we will show that the notion of a minimal uncertainty state strongly depends on the underlying uncertainty measure, which shines a doubtful light on some heuristic applications of the GUP appearing in literature.

From a purely mathematical perspective, the resulting picture of 'Planck scale quantum mechanics' has many coincidences to the well established field of signal processing. Here, classical uncertainty relations have been extensively investigated by Landau, Slepian, and Pollak [SP61, LP61, LP62]. Within this analogy, states that obey a UV cutt-off correspond to band limited functions and an apparent discreteness of space-time to the well known effect of Nyquist-sampling.

## 8.1 [ACF$^+$16]

*Optimal uncertainty relations in a modified Heisenberg algebra*

- **Authors:** Kais Abdelkhalek, Wissam Chemissany, Leander Fiedler, Gianpiero Mangano, René Schwonnek
- **Published in:** Physical Review D 94, 123505 (2016)
- **DOI:** 10.1103/PhysRevD.94.123505
- **Presented version:** The presented version is identical to arXiv:1607.00081v2, the literature and the appendix is placed at the end of this thesis.
- **Contributions:** Main contributions to this work and the process of writing were done by Kais Abdelkhalek and René Schwonnek. The proof of Theorem 1 was contributed by Leander Fiedler. Gianpiero Mangano, author of the seminal paper [KMM95], contributed advise and the examples in Sec. III A.
- **Main results:**
  - sufficient conditions for the emergence of a minimal length
  - computation of minimal variance and variance based uncertainty relations
  - numerical computation of / lower bounds on the minimal Shannon entropy
  - minimal min entropy

# Optimal uncertainty relations in a modified Heisenberg algebra

Kais Abdelkhalek,[1, *] Wissam Chemissany,[1, 2, †] Leander Fiedler,[1, ‡] Gianpiero Mangano,[3, §] and René Schwonnek[1, ¶]

[1]Institut für Theoretische Physik, Leibniz Universität Hannover, Hannover, Niedersachsen, Germany
[2]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[3]INFN, Sezione di Napoli, Complesso Univ. Monte S. Angelo, Napoli, Italy

Various theories that aim at unifying gravity with quantum mechanics suggest modifications of the Heisenberg algebra for position and momentum. From the perspective of quantum mechanics, such modifications lead to new uncertainty relations which are thought (but *not* proven) to imply the existence of a minimal observable length. Here we prove this statement in a framework of sufficient physical and structural assumptions. Moreover, we present a general method that allows to formulate optimal and state-independent variance-based uncertainty relations. In addition, instead of variances, we make use of entropies as a measure of uncertainty and provide uncertainty relations in terms of min- and Shannon entropies. We compute the corresponding entropic minimal lengths and find that the minimal length in terms of min-entropy is exactly one bit.

## I. INTRODUCTION

A considerable amount of efforts has been devoted to reconcile gravity with quantum mechanics, but the conventional field theoretic avenues for quantizing general relativity have suffered issues in renormalisability. Several theories such as string theory have suggested that the sought-after quantum gravity has to be effectively cut off in the ultraviolet, leading to the notion of *minimal length* [1–3] (see also [4] and [5] and references therein). In other words, the gravitational effects become significantly important upon probing physics at an energy scale as large as the Planck scale. Such a nontrivial premise of the minimal position uncertainty has been corroborated by string theoretic arguments [1, 6], leading to the so-called generalized uncertainty principle.

There had been a consensus within the high energy physics community that such a minimal length has a quantum mechanical origin which should effectively be formulated in the form of a non-zero minimal uncertainty for a position measurement. In its simplest version, this can be obtained by explicitly constructing position and momentum operators $\mathbf{x}$ and $\mathbf{p}$ that satisfy a deformed Heisenberg algebra

$$[\mathbf{x}, \mathbf{p}] = i\hbar f(\mathbf{p}), \tag{1}$$

where the precise form of the modification $f(\mathbf{p})$ depends on which theory and approach is used [1, 6–8], see also e.g. [9–11] for deformations in configuration space. Recently, a deformation of (1) where the r.h.s. is assumed to be a stochastic gaussian variable has been also considered [12]. Large parts of related literature focuses on the case where $f(\mathbf{p}) = 1 + \beta\mathbf{p}^2$ and uncertainty is measured in terms of variances. This specific modification corresponds to the first term in a Taylor expansion in $\mathbf{p}^2$ and reflects the expected simplest deviation from the standard case. In this work we allow modifications that satisfy a number of physically well-motivated assumptions (see Section II) and are otherwise completely general.

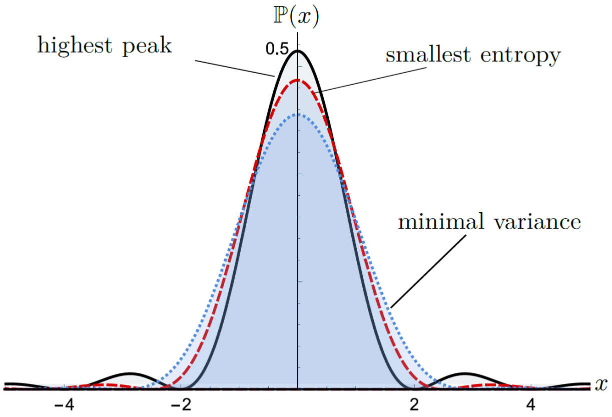


Fig. 1: Probability distributions of a position measurement on quantum states suspected to a UV cut-off (see (36)). Which one is the "sharpest"? All three distributions attain a minimal length in their own sense: the blue, dotted one attains minimal length in terms of variances as typically considered in the literature. The black, solid distribution has the highest peak and the red, dashed distribution has the smallest entropy. The latter two distributions have infinite variance, although they appear to be localised to some extent.

Despite the substantial understanding that had been gained in previous approaches, there still exist conceptual shortcomings in the study of the origin of minimal lengths. Firstly, the term "minimal length" should not be interpreted as an actual geometric length. Instead, minimal length refers to the perception that, in a modified algebra, the probability distributions obtained in a position measurement cannot become arbitrarily sharp. Hence, minimal length should rather be dubbed and always be understood as "minimal position uncertainty".

Secondly, if minimal length is to be understood as an

[*] kais.abdelkhalek@itp.uni-hannover.de
[†] wissam.chemissany@itp.uni-hannover.de,wissamch@mit.edu
[‡] leander.fiedler@itp.uni-hannover.de
[§] mangano@na.infn.it
[¶] rene.schwonnek@itp.uni-hannover.de

immediate consequence of modifying the Heisenberg algebra, it is important to show that *all* possible pairs of operators $\mathbf{x}$ and $\mathbf{p}$ that satisfy this algebra lead to a non-zero minimal position uncertainty. If such a statement is not correct in all its generality, what assumptions are needed besides modifying the algebra to prove minimal length? In the standard case this question is answered by the Stone-von Neumann theorem, in the present context the situation is not at all clear. Most work related to the study of minimal length focused only on showing the mere existence of such operators neglecting such uniqueness considerations[70].

Thirdly, no general method to compute optimal and state-independent uncertainty relations for a given modification $f(\mathbf{p})$ has been developed so far. While such relations directly yield minimal length, they are of scientific interest on their own, since they express the influence of the modification on all states. For example, experimental proposals like [13] aim at observing a modified uncertainty relation in an uncertainty regime where minimal length cannot be attained.

As a last point, characterizing minimal length in terms of variances is at least controversial: on one hand variances characterise well the uncertainty for most unimodal distributions, especially if they are Gaussian. On the other hand, variances of multimodal distributions are known to show strange and unwanted behaviour if interpreted as a measure of uncertainty (as is done for minimal lengths). There is no reason why a position distribution should in general be unimodal, especially since non-zero minimal length immediately implies that Gaussian states are not part of the considered Hilbert space. Hence, in most cases variances are not a good candidate to capture the notion of minimal length as can also be seen from Fig. 1. In Section IV we discuss this in more detail.

Which measure to use instead is far from unique and depends on the operational task that shall be accomplished. Entropies as a measure of uncertainty have been proven tremendously useful in various fields, such as quantum information theory (see e.g. [14]), quantum thermodynamics (see [15] for a survey), or quantum gravity (for recent work and references see [16]). For example, the uncertainty principle has been made operationally precise in the form of entropic uncertainty relations which for instance are an essential part of security proofs of quantum cryptographic protocols (e.g. [17]) and are still focus of much investigation [18–20], see also the reviews [21, 22]. It thus seems beneficial to formulate entropic uncertainty relations in the context of modified Heisenberg algebras thereby introducing the concept of *entropic minimal length.* We investigate its implications compared to those obtained by its variance-based counterpart, see [23].

In this work we develop the underlying quantum mechanical setting in which one can study the *direct* consequences of modifying the Heisenberg algebra in view of the existence of non-zero minimal lengths. Then, temporarily complying with the consensus to formulate minimal length in terms of variances, we provide a general framework from which optimal and state-independent uncertainty relations and minimal lengths can be calculated efficiently. Here, the term "optimal" refers to Pareto optimality, which originates in the theory of optimisation [24]. We will also argue why a typical approach that invokes equality in the Robertson-Kennard relation (15) does not provide an optimal uncertainty relation. Lastly, we compute and discuss implications and advantages of an entropic formulation of minimal length. In particular, we introduce minimal length in terms of the Shannon entropy (or differential entropy in the continuous setting) and show that both the minimal length value and the corresponding minimizing states are not equivalent to those obtained for the "standard" minimal length in terms of variances. We also discuss how a further feature appears when using entropies, namely a maximal entropy in momentum space. Finally, we compute minimal length in terms of the min-entropy, which quantifies the maximum probability of correctly predicting the outcome of a position measurement. Intriguingly, we find an intimate connection between variance-based and min-entropy based minimal length: for scenarios with normalised variance-based minimal length, the mininal length in terms of min-entropy is also normalised, meaning that the best possible localisation of space is exactly one bit.

## II. REPRESENTATION OF THE MODIFIED HEISENBERG ALGEBRA

Let us consider the position operator $\mathbf{x}$, i.e. the multiplication operator on the Hilbert space $\mathcal{H} := \mathcal{L}^2(\mathbb{R})$. In this section we characterise properties of momentum operators $\mathbf{p}$ that satisfy the modified Heisenberg algebra (1). After briefly summarising and unifying previous constructions that showed the *existence* of operators $\mathbf{p}$ that lead to minimal lengths effects, we present our main result of this section that proves the *uniqueness* of such constructions.

More concretely, we aim at characterising a linear, self-adjoint operator $\mathbf{p}$ with dense domain in a closed subspace $\mathcal{P}$ of a Hilbert space $\mathcal{H}$ with spectrum coinciding with $\mathbb{R}$ that satisfies the modified Heisenberg algebra[71] (we set $\hbar = 1$ in the following)

$$[\mathbf{x}, \mathbf{p}] = if(\mathbf{p}) \ , \tag{2}$$

where $f : \mathbb{R} \to \mathbb{R}$ satisfies

(i) $f(0) = 1$ ,

(ii) $f(p) = f(-p)$ for all $p \in \mathbb{R}$ ,

(iii) $f(p)$ is convex on $\mathbb{R}^+$, i.e. $\forall p, p' \geq 0$: $f(\lambda p + (1-\lambda)p') \leq \lambda f(p) + (1-\lambda) f(p')$.

Assumption (i) ensures that for small momentum we retrieve the original unmodified Heisenberg algebra, while

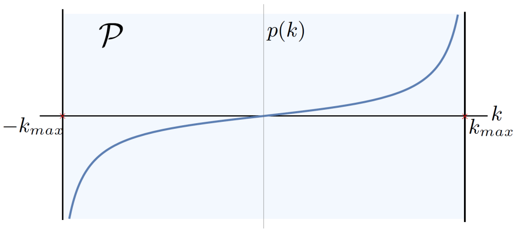


Fig. 2: Theorem 1 gives sufficient conditions for representing the operator $\mathbf{p}$ as a function $p(\mathbf{k})$, where $\mathbf{k}$ denotes the unmodified momentum operator (here shown for the modification $f(p) = 1 + \beta p^2$). We show that the support of $\mathbf{k}$ must be restricted to an interval $[-k_{max}, k_{max}]$, i.e. modifying the algebra directly leads to UV cut-off and, hence, minimal length.

assumption (ii) translates to the statement that momentum should not have a preferred direction. The last assumption (iii) is a generalisation of modifications that were considered previously in the literature [3, 5] and implies that higher momentum leads to stronger effects of the modification.

Since $f(\mathbf{p})$ is adimensional, it depends on momentum via the product $\sqrt{\beta}\,\mathbf{p}$ , with $\beta$ a constant with dimension of inverse squared momentum (or inverse squared mass in natural units), which sets the scale where deviations with respect to the standard picture are important[72]. In natural units $\beta$ is naturally expected to be of the order of $m_{Pl}^{-2}$, with $m_{Pl}$ the Planck mass, but we consider this scale as a free parameter.

Previous works aimed at explicitly constructing operators $\mathbf{p}$ that satisfy the algebra (2) and lead to a non-trivial minimal length. Before discussing subtleties arising in these approaches, we briefly review these constructions which may be unified as follows: consider the *unmodified* momentum operator $\mathbf{k}$ on $\mathcal{H}$ such that $\mathbf{x}$ and $\mathbf{k}$ satisfy the standard commutation relation

$$[\mathbf{x}, \mathbf{k}] = i\mathbb{I} \ , \tag{3}$$

i.e. the momentum operator is given by

$$\mathbf{k} = \mathcal{F}^\dagger \mathbf{x} \mathcal{F} \ , \tag{4}$$

with $\mathcal{F}$ the Fourier transform acting on states $\phi \in \mathcal{H}$ via

$$(\mathcal{F}\phi)(k) = \frac{1}{\sqrt{2\pi}} \int e^{ikx} \phi(x) dx \ . \tag{5}$$

We can then deform the spectrum of $\mathbf{k}$ until we find a linear operator $\mathbf{p} = p(\mathbf{k})$ that satisfies (2) (see Fig. 2). More concretely, by functional calculus we can evaluate the commutator

$$[\mathbf{x}, \mathbf{p}] = \left[ i\frac{d}{dk}, p(k) \right] = i\frac{d}{dk} p(k) \ , \tag{6}$$

to find that (2) translates to the differential equation

$$\frac{d}{dk} p(k) = f(p(k)) \ . \tag{7}$$

By the implicit function theorem we obtain the solution

$$k(p) = \int_{p_0}^{p} dp' \frac{1}{f(p')} \ , \tag{8}$$

where we set $p_0 = 0$, such that the momentum operators $\mathbf{p}$ and $\mathbf{k}$ yield the same physics in the small-momentum regime. For all cases where this integral is finite in the limit $p \to \infty$, this implies the existence of a momentum cut-off,

$$k_{\text{max}} := \int_{0}^{\infty} dp' \frac{1}{f(p')} \ . \tag{9}$$

This argument, commonly found in related literature [3, 25], shows that for states with support in the interval $[-k_{\text{max}}, k_{\text{max}}]$ there exist operators $\mathbf{x}$ and $\mathbf{p}$ satisfying the modified commutation relation (2). Moreover, the operator $\mathbf{p}$ is just defined on a proper subspace $\mathcal{P} = \mathcal{L}^2([-k_{\text{max}}, k_{\text{max}}])$ of the Hilbert space $\mathcal{H}$ (Fig. 3). In particular, states with vanishing position uncertainty which have, by (5), a broad momentum distribution are no longer contained in $\mathcal{P}$, hence implying the existence of a non-trivial minimal length. In this sense the existence of a momentum cut-off, sometimes also referred to as a UV cut-off, directly implies the existence of a non-trivial minimal length. Conversely, if $\mathcal{P} = \mathcal{H}$, i.e. there is no momentum cut-off, there are states with vanishing position uncertainty as is the case for the unmodified Heisenberg algebra.

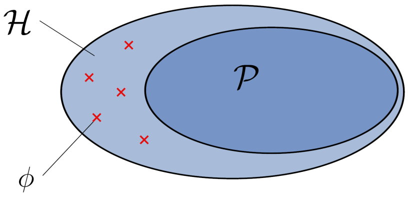


Fig. 3: For an unmodified algebra, minimal uncertainty is obtained for states $\phi$ from a state space $\mathcal{H}$. In the case of the usual Heisenberg algebra the corresponding minimal length becomes trivial, that is, all states in $\mathcal{H}$ are physical. If a modification of the algebra directly leads to a restricted state space $\mathcal{P}$ that does not contain such states, one obtains non-trivial minimal length as a direct consequence of modifying the algebra.

However, in order to interpret the momentum cut-off and the corresponding minimal length as a direct consequence of modifying the algebra, it is not sufficient to just show the existence of operators that allow for non-trivial minimal lengths as above. Instead, one needs to show that *all* operators $\mathbf{x}$ and $\mathbf{p}$ satisfying (2) lead to this effect. In this section we show that this is indeed the case if, given $\mathbf{x}$ is the standard position operator, we additionally require that the spectral projections of the canonical momentum operator $\mathbf{k} = \mathcal{F}^\dagger \mathbf{x} \mathcal{F}$ and the modified momentum operator $\mathbf{p}$ are close to each other in the regime of small momenta. This assumption immediately implies that, in this regime, the probability distributions induced by $\mathbf{k}$ and $\mathbf{p}$ are almost the same, which agrees with the intuition that *observable effects of the modified algebra only occur for high momenta.* More precisely, we require that there exists $\epsilon_0 > 0$ such that for all $\epsilon \in (0, \epsilon_0)$ there is an $\epsilon' > 0$ and $\delta > 0$ with $\delta \sim \mathcal{O}(\epsilon^3)$ such that

$$\|E_{\mathbf{k}}([-\epsilon, \epsilon]) - E_{\mathbf{p}}([-\epsilon', \epsilon'])\| < \delta \ , \qquad (10)$$

where $E_{\mathbf{k}}$ and $E_{\mathbf{p}}$ denote the spectral projections of $\mathbf{k}$ and $\mathbf{p}$, respectively. With this assumption we show the following theorem (see Appendix for the proof):

**Theorem 1.** *Let* $\mathbf{x}$ *be the position operator and* $\mathbf{k} = \mathcal{F}^\dagger \mathbf{x} \mathcal{F}$ *be the unmodified momentum operator as defined above. Denote by* $\mathbf{p}$ *a modified momentum operator on a Hilbert space* $\mathcal{P}$*, i.e.* $\mathbf{x}$ *and* $\mathbf{p}$ *satisfy the modified Heisenberg algebra* (2) *for all states in* $\mathcal{P}$*. If additionally* (10) *is satisfied, then*

- *there exists a* momentum cut-off*, i.e. there is an interval*

$$I = [-k_{\max}, k_{\max}] \subseteq \mathbb{R} \ , \qquad (11)$$

*with* $k_{\max}$ *as in* (9) *such that* $\mathcal{P} = \mathcal{L}^2(I)$*,*

- *there is a function* $p : I \to \mathbb{R}$ *such that for all states in* $\mathcal{P}$

$$\mathbf{p} = p(\mathbf{k}) \ . \qquad (12)$$

*Hence,* $\mathbf{p}$ *is indeed a function of* $\mathbf{k}$*, which legitimates the standard construction after eq.* (5)*.*

In other words, the scope of Theorem 1 can be summarised as follows: assume an experimenter who, on a length scale that is above Planck length, can agree on a clear notion of what the position $\mathbf{x}$ and the unmodified momentum $\mathbf{k}$, i.e. a particular representation of the Heisenberg algebra, should be. If he extrapolates his notion of $\mathbf{x}$ down to lower scales, and assumes that a modified algebra has a consistent limit to what he observed on higher scales, he obtains by Theorem 1 a *unique* notion of what $\mathbf{p}$ is in this situation. Theorem 1 can therefore be seen as the reason why the aforementioned construction is indeed meaningful: the construction describes *all* possible modified momentum operators $\mathbf{p}$. Importantly, it proves the existence of a UV cut-off and a corresponding non-trivial minimal length as a direct consequence of modifying the underlying algebra.

Note that Theorem 1 builds on the natural assumption that measurement probabilities should be similar when measuring the modified momentum operator $\mathbf{p}$ or the unmodified momentum operator $\mathbf{k}$ in the regime of small momentum. This assumption is essential since it provides a means to characterise the action of $\mathbf{x}$ on states in $\mathcal{P}$: as the position operator $\mathbf{x}$ induces shifts in $\mathbf{k}$-momentum space, knowing that $\mathbf{p}$ and $\mathbf{k}$ are not too different in the small-momentum regime implies that $\mathbf{x}$ also induces shifts in $\mathbf{p}$-momentum space (up to some arbitrarily small error). By how much $\mathbf{x}$ is shifting a state in $\mathbf{k}$- or $\mathbf{p}$-space is governed by the respective commutation relation. Hence, while $\mathbf{x}$ induces constant shifts in $\mathbf{k}$-space, the strength of shifting in $\mathbf{p}$-space is monotonically increasing with higher momentum as is the modification $f$ (Assumption (iii)). This leads to normalizability constraints: for high enough momentum the shift becomes too large for the corresponding states to be normalizable. The cut-off parameter $k_{\max}$ is exactly the momentum value for which the states cannot be normalised anymore.

The self-adjoint position operator $\mathbf{x}$ has a domain dense in $\mathcal{H}$. Theorem 1 shows however that, if the modification $f(\mathbf{p})$ is such that $k_{\max}$ is finite, the relevant Hilbert space $\mathcal{P}$ is strictly smaller than $\mathcal{H}$. So, how can the measurements of position be described if acted on states in $\mathcal{P}$? This is in particular interesting since the position operator $\mathbf{x}$ restricted to $\mathcal{P}$ is *not* self-adjoint anymore. Nevertheless, a position measurement still has a well-understood description, known as a POVM (positive operator valued measure). POVMs describe the most general form of a quantum measurement (see e.g. [14]). An explicit construction of the position operator as a POVM on a restricted state space can be found in [26].

## III. OPTIMAL UNCERTAINTY RELATIONS IN TERMS OF VARIANCES

The concept of minimal length expresses the fact that for all states position measurements will in general not produce arbitrarily sharp outcome distributions. This is typically quantified by computing the minimal variance of this distribution

$$l_{\min}^2 = \min_{\psi \in \mathcal{P}} \Delta \mathbf{x} \ , \qquad (13)$$

where

$$\Delta \mathbf{x} = \langle \psi | \mathbf{x}^2 | \psi \rangle - \langle \psi | \mathbf{x} | \psi \rangle^2 . \qquad (14)$$

As such, minimal length is intimately related to the concept of uncertainty relations, which place constraints on how sharp the distribution for some observable $A$ can be, given the sharpness of the distribution of another, say $B$. The standard example of such an uncertainty relation is the one due to Robertson and Kennard [27, 28], i.e.

$$\Delta A \Delta B \geq \frac{1}{4} |\langle \psi | [A, B] | \psi \rangle|^2 \ . \qquad (15)$$

A naive approach to compute the minimal length is to impose equality in (15) and then search for minimising states within the corresponding subset of states. In this section we will remark that this approach will fail in most cases due to the state dependence of the lower bound in (15). Instead, we provide a general framework to obtain optimal and state-independent uncertainty relations for a modified Heisenberg algebra in the first part of this section. As a side product this allows to directly compute the corresponding minimal length. Then, in the second part of the section, we exemplarily apply this framework to the modification $f(p) = 1 + \beta p^2$, since here all differential equations can be solved analytically. We obtain the same uncertainty relation as in [3], but now with a proof of its optimality. We also apply our framework to other modifications, i.e $f(p) = \cosh(\sqrt{\beta} p)$ and $f(p) = 1 + \beta p^2 + \beta^2 p^4/4$, employing numerical tools (see also [29] for a treatment of higher order modifications $f(\mathbf{p})$).

### A. A general method for finding uncertainty relations and minimal length

Uncertainty relations allow to lower bound the uncertainty of one measurement, given the uncertainty of another measurement. Having this in mind, a good way to generally think about uncertainty is in terms of diagrams, as shown in Fig. 4 [3, 18, 20]. Here, the blue shaded region indicates the set $\mathcal{U}$ of all tuples $(\Delta\mathbf{p}, \Delta\mathbf{x})$ that can be obtained by measuring both $\mathbf{p}$ and $\mathbf{x}$ on the same state $\psi$, where $\psi$ is taken from $\mathcal{P}$. In the following we will refer to $\mathcal{U}$ as the *uncertainty region.*

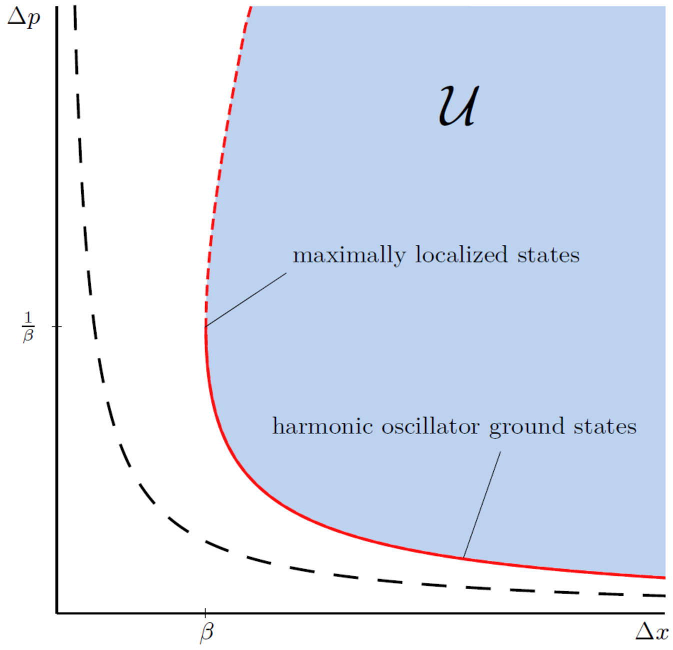


Fig. 4: The boundary of $\mathcal{U}$ for the modified (with $f(p) = 1 + \beta p^2$) and standard (long dashed black curve) Heisenberg algebra. We show the point corresponding to the maximally localized state (26), for which $\Delta\mathbf{x} = 1/\Delta\mathbf{p} = \beta$. The tradeoff curve (solid red line) $\gamma_{\mathcal{U}}(\lambda)$ branch below this point corresponds to ground states of (deformed) harmonic oscillators (36) with, from right to left, increasing frequency $\omega \in ]0, \infty[$. The upper part of the curve (short dashed red line) is obtained by considering states $\psi(k) \propto \cos(\sqrt{\beta} k)^{\gamma_\lambda}$ with $\gamma_\lambda < 1$, which are not the ground state of a harmonic oscillator.

Uncertainty relations express the fact that there is no state such that both variances are getting simultaneously arbitrary small. If this is the case, the point $(0, 0)$ is not contained in $\mathcal{U}$ and the uncertainty diagram has some empty space around the origin.

However, we can still ask for the "smallest" points in $\mathcal{U}$, i. e. the points on the "lower left" boundary of $\mathcal{U}$, which are obtained by minimising one variance under the constraint that the other stays below some fixed threshold, and vice versa [18, 20]. In Fig.4 this trade-off curve is indicated by a solid red line and henceforth referred to as an *optimal* and *state-independent* uncertainty relation. Here, the term "optimal" means that for any attainable value for $\Delta\mathbf{p}$, or equivalently for $\Delta\mathbf{x}$, we can find a state in $\mathcal{P}$ such that the uncertainty relation is tight, i.e. that equality is attained. "State-independent" means that the uncertainty relation only depends on functions of the variances $\Delta\mathbf{p}$, $\Delta\mathbf{x}$ and constants, but not on any other quantities that depend on the state. Hence an optimal and state-independent uncertainty relation defines a trade-off curve such that for any attainable value of $\Delta\mathbf{p}$, we can directly conclude the *best* lower bound on $\Delta\mathbf{x}$ that will hold for *all* states in $\mathcal{P}$, and vice versa.

Importantly, minimal length as defined in (13) can be directly computed if such an uncertainty relation is known by simply optimising over all possible values of $\Delta\mathbf{p}$. For this and other operationally motivated reasons pointed out by David Deutsch in the 80's [30], optimal and state-independent uncertainty relations are the ones to look for. However, such uncertainty relations are usually also the hardest ones to obtain because they always involve a constrained optimization problem over the whole state space.

At this point it might be important to recall the often ignored fact that, in general, an optimal and state-independent uncertainty relation cannot be inferred from the relation (15), which in our case takes the form

$$\Delta\mathbf{x}\Delta\mathbf{p} \geq \frac{1}{4}|\langle\psi|f(\mathbf{p})|\psi\rangle|^2 \ . \qquad (16)$$

Here, the expectation value on the right hand side of (16) (or (15)) is generally *state-dependent*, such that evaluating the uncertainty relation for a particular state does not allow to directly conclude anything about the uncertainty of any other state. In particular, it is generally not true, in neither direction, that states, which are giving equality in (15), correspond to points on the boundary of an uncertainty region. This circumstance can be exemplary checked by considering any non-commuting pair of measurements, e.g. two angular momentum components [18]. This has also been pointed out by [31, 32] for the case of the smallest value of $\Delta\mathbf{x}$. In general, this makes the method to investigate equality in (15) and to infer minimal length, quite problematic.

However, there are at least two exceptions to the above criticism: the one is the usual Heisenberg algebra, where the right hand side of (15) is the same for every normalized state and thus state-independent. Optimality is granted by Gaussian states, for which it is well-known that they achieve equality in (15) in the whole parameter range of $\Delta\mathbf{p}$ and $\Delta\mathbf{x}$, respectively.

The other exception has been exploited in [3] for the case of a modified Heisenberg algebra with $f(p) = 1 + \beta p^2$. If we take the square root on both sides of (16), the right hand side only contains a constant, some factors and the second moment of $\mathbf{p}$ and thus one obtains a state-independent uncertainty relation

$$\sqrt{\Delta\mathbf{x}\Delta\mathbf{p}} \geq \frac{1}{2}\left(1 + \beta\Delta\mathbf{p} + \langle\mathbf{p}\rangle^2\right) \geq \frac{1}{2}\left(1 + \beta\Delta\mathbf{p}\right). \quad (17)$$

Whilst this was not spelled out in [3] directly, this bound is in fact optimal as we will show at the end of this section by a straightforward application of Theorem 2.

For a large class of modifications (see Subsection C in the appendix) we can set

$$g(p) := f(-\sqrt{|p|})\ , \quad (18)$$

and substitute this into (16). Then, in a similar spirit as above, we obtain

$$\Delta\mathbf{x} \geq \frac{g(\Delta\mathbf{p})^2}{4\Delta\mathbf{p}}\ , \quad (19)$$

which is state-independent but in general not optimal (see the blue line in Fig. 5).

To the best of our knowledge no universal method for obtaining an optimal, state-independent uncertainty relation for an arbitrary pair of observables $A$ and $B$ is known. However, it is possible to obtain lower bounds (and by this a state-independent uncertainty relation) on every uncertainty region by computing its convex hull (see [18, 33, 34] and Subsection B in the appendix). Such a bound will become optimal whenever $\mathcal{U}$ itself is convex. In this case an uncertainty relation can always be characterised by a function $u(\lambda)$ with $\lambda \in [0, 1]$ and a set of linear inequalities

$$\lambda\Delta\mathbf{x} + (1-\lambda)\Delta\mathbf{p} \geq u(\lambda)\ . \quad (20)$$

The function $u(\lambda)$ will give us a full description of the boundary of the convex hull of $\mathcal{U}$ (see Subsection B in the appendix and Fig.10). If needed, one can recover the trade-off curve, denoted by $\xi_{\mathcal{U}}(\lambda)$ in the following, by the formula

$$\xi_{\mathcal{U}}(\lambda) = (u(\lambda) + (1-\lambda)u'(\lambda), u(\lambda) - \lambda u'(\lambda))\ . \quad (21)$$

Note that, given a particular form of $u(\lambda)$, one can always find a substitution for $\lambda$ in (21), such that (21) has a form that only depends on $\Delta\mathbf{x}$ and $\Delta\mathbf{p}$.

The following theorem states that the ansatz above is already sufficient for providing an *optimal* uncertainty relation.

**Theorem 2.** *Let $\mathcal{U}$ be the uncertainty region of $\mathbf{x}$ and $\mathbf{p}$ satisfying a modified algebra with a modification $f(p)$ that obeys the assumptions described in section II. Then*

1. *the lower boundary, i.e. the trade-off curve, of $\mathcal{U}$ lies completely on the boundary of a convex set,*
2. *states corresponding to this trade-off curve always have expectation $\langle\mathbf{p}\rangle = 0$ and can be chosen to have $\langle\mathbf{x}\rangle = 0$,*
3. *these states are ground states of the modified harmonic oscillator*

$$H_\lambda = \lambda\mathbf{x}^2 + (1-\lambda)\mathbf{p}^2.$$

A proof and a mathematically more dedicated formulation of the statements 1 and 2 from Theorem 2 can be found in Subsection B in the appendix. However, statement 3 can be concluded directly from 1 and 2 using (20): from 1 we know that we can obtain the optimal bound $u(\lambda)$ by minimising the expression $\lambda\Delta\mathbf{x} + (1-\lambda)\Delta\mathbf{p}$ for fixed $\lambda$ over all states in $\mathcal{P}$. Using 2 we arrive at

$$u(\lambda) = \min_{\psi\in\mathcal{P}}\langle\psi|(\lambda\mathbf{x}^2 + (1-\lambda)\mathbf{p}^2)|\psi\rangle, \quad (22)$$

which is exactly the ground state energy of a harmonic oscillator in the modified algebra. Moreover, when we represent $\mathbf{p}$ and $\mathbf{x}$ in the domain of $\mathbf{k}$, we can state the following:

**Corollary 3.** *An optimal and state-independent uncertainty relation can be directly obtained by solving the ground state problem of the Schrödinger operator*

$$H_\lambda = -\lambda\partial_k^2 + (1-\lambda)p(k)^2\ , \quad (23)$$

*with Dirichlet boundary conditions at $\pm k_{max}$ and a symmetric, convex and positive potential $p(k)^2$. This uncertainty relation saturates* (20) *and can be found by solving*

$$H_\lambda\psi_\lambda(k) = u(\lambda)\psi_\lambda(k)\ , \quad (24)$$

*where $u(\lambda)$ is given by the ground state energy of* (23)*.*

Fortunately, these kind of problems have been the content of many extensive studies (see for example [35, 36]) since the early days of quantum mechanics. Indeed, well-established numerical methods and analytical solutions for several particular instances of $p(k)$ are available.

Asking for an optimal bound in expression (20) for the special case $\lambda = 1$ directly translates into characterising the minimal length $l^2_{\mathrm{min}}$ in terms of variances. By Corollary 3 this turns into the task of finding the state $\psi(k)$ and the minimal value $u(1) = l^2_{min}$ such that the differential equation

$$-\partial_k^2\psi(k) = l^2_{\mathrm{min}}\psi(k)\ , \quad (25)$$

holds with the boundary condition $\psi(\pm k_{\mathrm{max}}) = 0$. But this is just the ground state problem of a particle in a box with length $2k_{\mathrm{max}}$ and this is solved by

$$\psi(k) = \frac{1}{\sqrt{k_{\mathrm{max}}}}\cos\left(\frac{\pi}{2}\frac{k}{k_{\mathrm{max}}}\right)\ , \quad (26)$$

so that

$$l^2_{\mathrm{min}} = \frac{\pi^2}{4k^2_{\mathrm{max}}}. \quad (27)$$

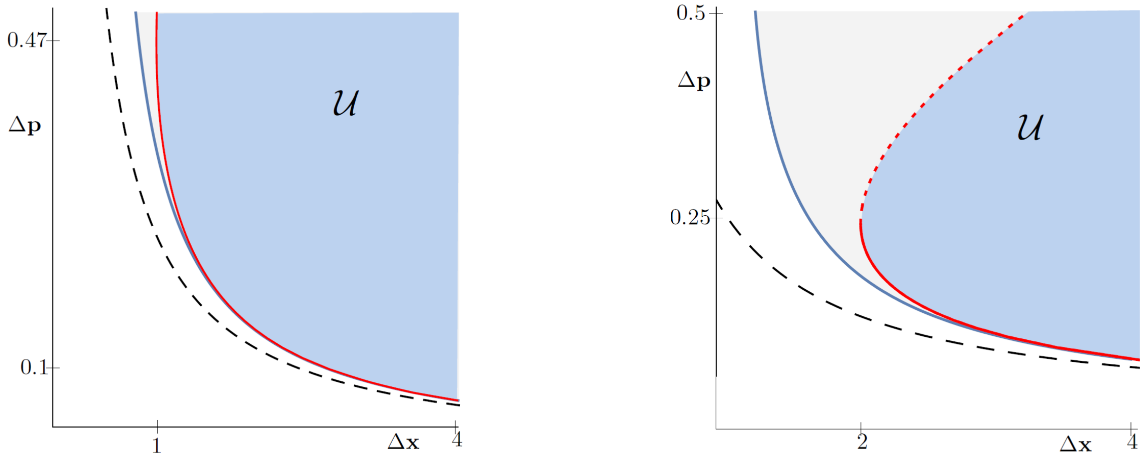


Fig. 5: Uncertainty regions $\mathcal{U}$ for the modifications $f_1 = \cosh(\sqrt{\beta}p)$ (left figure) and $f_2 = 1+\beta p^2+\beta^2 p^4/4$ (right figure) bounded by the red lines. The black dashed lines show the standard Heisenberg bound for the position and unmodified momentum operator. The blue lines show the not optimal but state-independent bounds from (19).

Note that all these results are completely general and hold for all modifications that satisfy the conditions (i) - (iii) of section II. As such they not only generalise previous results obtained in [31, 32], but also allow to drastically improve earlier approaches as they provide a means to straightforwardly compute optimal uncertainty relations and minimal lengths.

To illustrate this point, let us consider the most studied modification

$$f(\mathbf{p}) = 1 + \beta \mathbf{p}^2, \tag{28}$$

for which a value for minimal length and its corresponding quantum state is already known, while the uniqueness of the construction of $\mathbf{x}$ and $\mathbf{p}$ has been left open. In [3] the authors also provide a state-independent uncertainty relation. Using our results we can directly prove that this uncertainty relation is actually optimal. Additionally, we show how, from such an uncertainty relation, one can easily retrieve the aforementioned results which were previously obtained in a much more mathematically involved manner. The purpose of this example is therefore to show the validity of our results and to provide a step-by-step recipe to compute uncertainty relations and minimal lengths by making use of the main results presented so far in this paper.
As a first step, note that $f(\mathbf{p})$ satisfies the requirements (i) - (iii). Hence we know by Theorem 1 that the modified momentum operator $\mathbf{p}$ must be a hermitian operator that satisfies the differential equation (7). This yields

$$p(k) = \frac{1}{\sqrt{\beta}} \tan(\sqrt{\beta} k) \ . \tag{29}$$

Also, by (9) we can compute the momentum cut-off,

$$k_{\max} = \frac{\pi}{2\sqrt{\beta}} \ . \tag{30}$$

The optimal state-independent uncertainty relation is characterised by the trade-off curve of the uncertainty region $\mathcal{U}$, the exact form of which depends on the modification. By Theorem 2 we know that the states $\psi_\lambda$ parametrising this trade-off curve are ground states of the modified harmonic oscillator with Hamiltonian $\mathbf{H}_\lambda = \lambda \mathbf{x}^2 + (1-\lambda)\mathbf{p}^2$, i.e.

$$\mathbf{H}_\lambda \psi_\lambda = u(\lambda)\psi_\lambda \ . \tag{31}$$

We rewrite this condition by explicitly inserting the parameter $\beta$ to render all terms adimensional. By dividing by $\lambda$ we then have

$$\left(\frac{1}{\beta}\mathbf{x}^2 + \frac{1-\lambda}{\lambda}\beta\,\mathbf{p}^2\right)\psi_\lambda = \frac{u(\lambda)}{\lambda}\psi_\lambda \equiv \gamma(\lambda)\psi_\lambda \ . \tag{32}$$

The ground states of these Hamiltonians correspond to vectors in the kernel of the annihilation operator

$$\mathbf{a}_\lambda = \mathbf{x}/\sqrt{\beta} + i\gamma_\lambda\sqrt{\beta}\,\mathbf{p} \ , \tag{33}$$

since $\mathbf{H}_\lambda$ always[73] satisfies

$$\frac{1}{\lambda}\mathbf{H}_\lambda = \frac{1}{\beta}\mathbf{x}^2 + (\gamma_\lambda^2 - \gamma_\lambda)\beta\,\mathbf{p}^2 = \mathbf{a}_\lambda^\dagger \mathbf{a}_\lambda + \gamma_\lambda \mathbb{I} \ , \tag{34}$$

when choosing $\gamma_\lambda$ such that $\gamma_\lambda^2 - \gamma_\lambda = (1-\lambda)/\lambda$. Hence, we have $\mathbf{a}_\lambda \psi_\lambda = 0$, which translates into the differential equation

$$\partial_k \psi_\lambda(k) + \gamma_\lambda \sqrt{\beta} \tan(\sqrt{\beta}k)\psi_\lambda(k) = 0 \ , \tag{35}$$

with the solution

$$\psi_\lambda(k) = \left(\frac{\beta}{\pi}\right)^{1/4} \left(\frac{\Gamma(1+\gamma_\lambda)}{\Gamma(1/2+\gamma_\lambda)}\right)^{1/2} \cos(\sqrt{\beta}k)^{\gamma_\lambda} \ , \tag{36}$$

where

$$\gamma_\lambda = \frac{1}{2}\left(1 + \sqrt{1 + 4\frac{1-\lambda}{\lambda}}\right) \ . \tag{37}$$

These states parametrise the complete trade-off curve of the uncertainty region $\mathcal{U}$ as depicted in Fig. 4 and therefore yield an optimal state-independent uncertainty relation for the modification $f(\mathbf{p}) = 1 + \beta\mathbf{p}^2$. More concretely, we can evaluate the variances $\Delta\mathbf{x}$ and $\Delta\mathbf{p}$ for these states

$$\Delta\mathbf{x} = \beta\frac{\gamma_\lambda^2}{2\gamma_\lambda - 1} \ , \quad \Delta\mathbf{p} = \frac{1}{\beta}\frac{1}{2\gamma_\lambda - 1} \ , \qquad (38)$$

where we invoked Theorem 2 to set $\langle\mathbf{x}\rangle = \langle\mathbf{p}\rangle = 0$.

Notice that for $\lambda \to 0$ (i.e. $\gamma_\lambda \to \infty$), the state becomes a plane wave, while $\lambda = 1$ ($\gamma_\lambda = 1$) corresponds to the maximally localized state

$$\psi(k) = \sqrt{\frac{2\sqrt{\beta}}{\pi}}\cos(\sqrt{\beta}k) \ , \qquad (39)$$

for which $\Delta\mathbf{x} = l_{\text{min}}^2 = \beta$, compare with (26) and (27). The results (36) and (37) as well as the trade-off curve coincide with those obtained in [3] (compare with Eq. (69) in that paper) with the identification

$$\frac{1-\lambda}{\lambda} = \frac{1}{(\beta m\omega)^2} \ . \qquad (40)$$

See also [37], where the harmonic oscillator problem in presence of a minimal length uncertainty relation is also solved in arbitrary dimensions. However, our findings greatly simplify and extend the derivation of these results, while proving uniqueness properties and the optimality of the state-independent uncertainty relation, and allowing to treat any modification $f$ that satisfies (I) - (III).

The above discussion allows for interesting physics to become visible: when looking at the trade-off curve traced out by the states $\psi(k)$ (see Fig. 4) the position variance decreases with increasing momentum variance - exactly up to the point where the frequency of the harmonic oscillator diverges, $\lambda = \gamma_\lambda = 1$. At this point the state reaches the maximal possible localisation in space, the endpoint of the solid red line in Fig. 4. Actually, using (38) it is easy to check that the states (36) still saturate the generalized uncertainty principle bound even for $\gamma_\lambda < 1$ but they *do not correspond to the ground state of a harmonic oscillator*, see (40), but rather can be formally seen as eigenstates of a quadratic potential with *an imaginary frequency* $\omega$. This regime corresponds to the upper branch in Fig. 4 (dashed red line). When $\gamma_\lambda$ decreases, both $\Delta\mathbf{x}$ and $\Delta\mathbf{p}$ grow and diverge in the limit $\gamma_\lambda \to 1/2$. Yet, the states with any $\gamma_\lambda > -1/2$ are normalizable, so that we can associate to them an entropy in both momentum and position space, as we will see in the next section.

## IV. ENTROPIC BOUNDS

In this section we introduce, compute and discuss implications of an entropic formulation of uncertainty and minimal length. Here, one might be tempted to ask why using variances is not always a good choice to quantify the uncertainty of two measurements, especially since we dedicated the whole last section to exactly this setting. The answer is that the emphasis of the previous section lies in the formulation of optimal and state-independent uncertainty relations which best describe minimal uncertainties and are always superior to statements about minimal length only or state-dependent uncertainty relations. The concept of optimal and state-independent uncertainty relations is however, completely independent of the chosen uncertainty measure: one can formulate such relations using variances as done in the last section, or compute so-called entropic uncertainty relations as suggested by David Deutsch in his seminal paper [30], which will be the content of this section. Before introducing entropies, let us first clarify why variances as measure of uncertainty are problematic.

In [30] David Deutsch argued that variances suffer from the fact that they depend on the specific ordering and labeling of measurement outcomes. To illustrate this point consider a fair coin that yields “heads” or “tails” with equal probability. To be able to quantify the uncertainty about the outcome of one toin coss in terms of variances, one needs to artificially associate real numbers to heads or tails. In other words, variances depend on the choice of the labels of the possible outcomes - to the extent that we can choose this “measure of uncertainty” to become arbitrarily small or large, while intuitively our uncertainty is the same, independent of the labeling.

As another example that is related to this problem let us consider a spin measurement on a spin-1 particle [21, 38]. Let us assume that we only know that the outcomes $\{-1, +1\}$ occur with equal probability $p_{-1} = p_{+1} = 1/4$, whereas with highest probability $p_0 = 1/2$ we obtain outcome 0. Now imagine that we get *additional information* about the source telling us that we never obtain outcome zero. Our state of knowledge changes and so does the probability distribution, which now is given by $p_{-1} = p_{+1} = 1/2$ and $p_0 = 0$. Here a “good” measure of uncertainty should mirror our decrease of uncertainty by not increasing during this process. However, variances do not satisfy this minimal requirement: in fact, the variance in the above example will increase by getting further information and it is easy to construct similar examples when we consider continuous observables as well.

For unbounded observables yet another problem arises: namely the variance of a random variable can diverge although the corresponding probability distribution seems to be “located” in some sense. Prominent examples for this are Cauchy distributions and Lévy distributions. Moreover, this effect also occurs for two of the distributions shown in Fig. 1. Here, the black and the red curve correspond to distributions which appear to be “localised” even though their variances diverge.

All these examples are a consequence of the variance depending on the outcomes and not only on the underlying probability distribution. In finite dimensions a well-known alternative to variances, that does not suffer from this drawbacks, are entropies [22, 30, 39]. The most prominent one is the Shannon entropy[74] of a discrete probability distribution $w : \mathbb{Z} \to (0,1)$,

$$H(w) := -\sum_i w_i \log(w_i) \ , \tag{41}$$

which was introduced in the seminal work [40]. Later Alfréd Rényi introduced a whole family of entropies $H_\alpha$ [41], called Rényi-$\alpha$ entropies, which also do not suffer from the above drawbacks and contain the Shannon entropy in the limiting case $\alpha \to 1$. In this work we will consider only the Shannon entropy and the *min-entropy* $H_\infty$, which arises in the limit $\alpha \to \infty$

$$H_\infty(w) = -\log(\max_i w_i) \ . \tag{42}$$

Note that these two entropies are so far only defined in a finite-dimensional setting. In the following we will define and compute minimal length in terms of Shannon entropies and min-entropies for continuous variables.

### A. Shannon entropy

In [40] Shannon presented a generalisation of (41) for continuous variables with a probability density $w : \mathbb{R} \to \mathbb{R}$

$$h(w) = -\int dy \ w(y) \log\left(w(y)\right), \tag{43}$$

which is called the *differential entropy.* This quantity can be negative and even reach the value $-\infty$, which might, on a first view, appear to be an astonishing property for an uncertainty measure. We therefore give some clarification of its meaning. Consider a continuous valued observable given by a random variable $Y$ with outcomes $y$ on the whole real line. Now assume that an experimenter tries to measure this observable with a device that has a finite operating range, lets say in an interval $I$. Assume further that her measurement device only has a finite resolution, say $\varepsilon$, which means that the device can only decide whether the outcome of a measurement is in a particular interval of length $\varepsilon$ or not. Now the experimenter can divide the operating range $I$ into bins $\Omega^i_\varepsilon$ of length $\varepsilon$ and will thus, effectively, obtain a description of her measurement by a discrete random variable, say $Y^I_\varepsilon$ with approximately $|I|/\varepsilon$ different outcomes. Here computing entropies like (41) or (42) for this random variable will give her a good description of the information theoretic uncertainty.

The experimenter might then take a better device, i. e. one with a finer resolution and a larger operating range. In this case the entropy increases because the number of possible measurement results will increase. Moreover, in the limit $|I| \to \infty$ and $\varepsilon \to 0$ the entropy will reach infinity. Nevertheless, assuming that $Y$ is distributed by $w$, for (41), we can write down this limit as

$$\lim_{\substack{|I|\to\infty \\ \varepsilon\to 0}} H(Y^I_\varepsilon) = \lim_{\substack{|I|\to\infty \\ \varepsilon\to 0}} -\sum_i \mathbb{P}(Y, \Omega^i_\varepsilon) \log(\mathbb{P}(Y, \Omega^i_\varepsilon)) \ , \tag{44}$$

where $\mathbb{P}(Y, \Omega^i_\varepsilon)$ denotes the probability of measuring a result in the bin $\Omega^i_\varepsilon$. For small $\varepsilon$ and a bin with center $y_i$ we might approximate this probability by $\varepsilon w(y_i)$ and get

$$\begin{aligned} H(Y^I_\varepsilon) \ &\approx \ -\sum_i \varepsilon w(y_i) \log\left(w(y_i)\right) \\ &\quad - \ \sum_i \varepsilon w(y_i) \log(\varepsilon) \ , \end{aligned} \tag{45}$$

which gives

$$\lim_{\varepsilon\to 0} \lim_{|I|\to\infty} H(Y^I_\varepsilon) = h(w) + \infty \ . \tag{46}$$

Hence the quantity $h(w)$ can be understood as a deviation from infinity. However the limit, in (46), strongly depends on the experimenter choice of dividing the interval $I$ into equidistant bins. This choice corresponds to the assumption that, not knowing anything about $Y$, the probability of obtaining an outcome in any bin $\Omega^I_\varepsilon$ should be the same when sampling from a uniform distribution on $I$. This assumption might become controversial (see [42]) when taking the limit $|I|$ to infinity, because there is no notion of a uniform distribution on the whole real line. In the following, we will see that we will have to take care of such a choice of a reference measure when defining entropies for the modified momentum.

At this point we should also emphasise that, even if the absolute value $h(w)$ has only a rather indirect operational meaning, $h(w)$ is still a good quantity to judge if one distribution is "sharper" than another, and thus minimizing $h(w)$ will give us a good notion to characterise minimal length quantum states. Moreover, $h(w)$ can be used to compute other information theoretic quantities like the mutual information $I(A,B) = h(A) + h(B) - h(AB)$ which quantifies the correlation between two random variables $A$ and $B$. Here $h(A)$ denotes the Shannon entropy of the probability density of the random variable $A$. It was shown in [40], that the "$\infty$" term from (46) cancels out, such that $I(A,B)$ arises as a rigorous limit of a discrete quantity.

Let us now define and compute the corresponding minimal length in terms of the Shannon entropy. To this end, we consider the Shannon entropy of a probability distribution obtained by measuring $\mathbf{x}$ on a state $\psi \in \mathcal{P}$. Assuming that we are given $\psi$ as a function $\psi(k)$ of the coordinate $k$ we can obtain its representation as a function $\phi(x)$ of the coordinate $x$ by applying a Fourier transformation. In this case the amplitude $|\phi(x)|^2$ will correspond to a probability density on $\mathbb{R}$ normalized with respect to the measure '$dx$' and we set

$$h_{\mathbf{x}}(\phi) = -\int_{\mathbb{R}} dx |\phi(x)|^2 \log|\phi(x)|^2 \ , \tag{47}$$

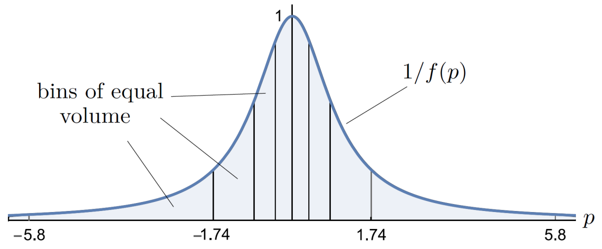


Fig. 6: Different binnings of measurement data lead to different uncertainty measures. Here we choose a binning such that all bins have the same probability in our definition of the Shannon entropy.

in order to define the Shannon entropy of a position measurement. In analogy to the "standard" (but problematic) definition of minimal length, minimal length *in terms of entropies* is now defined by the minimal entropy that a probability distribution obtained by a position measurement can have [75]

$$\Gamma_{\text{min}} = \min_{\psi \in \mathcal{P}} h_{\mathbf{x}}(\mathcal{F}(\psi)) \ . \tag{48}$$

Measuring the unmodified momentum $\mathbf{k}$ on $\psi \in \mathcal{P}$ will give us the probability distribution $|\psi(k)|^2$ such that we will define the unmodified momentum entropy as

$$h_{\mathbf{k}}(\psi) = -\int_I dk |\psi(k)|^2 \log |\psi(k)|^2 \ , \tag{49}$$

where the interval $I$ ranges from $-k_{\text{max}}$ to $+k_{\text{max}}$.

As mentioned in the previous subsection, the subtlety of choosing an appropriate reference measure emerges: when we represent $\psi$ as a function $\tilde{\psi}(p)$ of the coordinate $p$, the probability of measuring a certain value $p$ from an interval $(a, b)$ is given by

$$\mathbb{P}(p \in (a,b), \mathbf{p}, \psi) = \int_a^b \frac{dp}{f(p)} |\tilde{\psi}(p)|^2 \ , \tag{50}$$

which is no longer translation invariant, due to the scaling factor $1/f(p)$. Here the experimenter from the above example has to adapt to this when choosing bins: one choice would be to keep on taking bins with equal length. Another choice, the one we use in this work, is to take bins such that for all bins the probability obtained by measuring a function $\tilde{\psi}(p)$, which is constant on a particular bin, is the same (see Fig. 6). In this case a bin, with center $a$ and volume $\varepsilon$, will correspond to an interval $(p(-\varepsilon/2 + p^{-1}(a)), p(p^{-1}(a) + \varepsilon/2)$, where $p$ and $p^{-1}$ are obtained by representing $\mathbf{p}$ as a function of $\mathbf{k}$. By evaluating the limit of $\varepsilon \to 0$ we see that we therefore should define the entropy of a modified momentum measurement via[76]

$$h_{\mathbf{p}}(\tilde{\psi}) = -\int_{\mathbb{R}} \frac{dp}{f(p)} |\tilde{\psi}(p)|^2 \log |\tilde{\psi}(p)|^2 \ . \tag{51}$$

Notice that for $\mathbf{p}$ and $\mathbf{k}$ both choices will lead to the same definition of an entropy. Furthermore, $h_{\mathbf{p}}$ as defined above has the nice advantage that it arises from $h_{\mathbf{k}}$ by an integral substitution and thus does not change, i. e.

$$h_{\mathbf{p}}(\tilde{\psi}) = h_{\mathbf{k}}(\psi) \ . \tag{52}$$

Thus, the optimization of $h_{\mathbf{p}}$ over all states represented as functions of $p \in \mathbb{R}$ amounts to optimizing $h_{\mathbf{k}}$ over all functions of $k \in I$. Finally, $h_{\mathbf{p}}$ also depends on $f(p)$ only through the cut-off parameter $k_{\text{max}}$. All results that can be shown for an arbitrary $k_{\text{max}}$ are therefore valid for arbitrary modifications $f(p)$.

Having defined the entropic uncertainty measures, we can now compute the corresponding uncertainty relations. To this end, we consider all possible pairs $(h_{\mathbf{x}}(\phi), h_{\mathbf{k}}(\psi))$ which are attainable by $\psi(k) \in \mathcal{P}$ where $\phi(x) = \mathcal{F}[\psi](x)$. We recall that in the standard scenario, $\beta = 0$, both momentum and position entropies can become arbitrarily small or large if one consider sequences of states which converge (in the distribution sense) to the $\mathbf{x}$ or $\mathbf{p}$ eigenfunctions. In this case the "physical" states satisfies the Bialynicki-Birula (BB) bound [43, 44], obtained by the Babenko-Beckner inequality [45],

$$h_{\mathbf{k}}(\psi) + h_{\mathbf{x}}(\phi) \geq \log(\pi e) \ , \tag{53}$$

which is again saturated, as for the product of variances (see Theorem 2), by the ground states of harmonic oscillators with arbitrary frequency $\omega$.

For a modified algebra, the existence of a momentum cut-off is expected to imply a minimal value for $h_{\mathbf{x}}$. Before discussing its value, it is worth noticing that the representation of a modified algebra also implies a *maximal entropy $h_p$ in momentum space.* This maximal entropy is attained for any series of functions converging to the uniform distribution on $I = [-k_{\text{max}}, k_{\text{max}}]$, and reads

$$\max_{\psi \in \mathcal{P}} h_{\mathbf{k}}(\psi) = -\int_I \frac{1}{|I|} \log\left(\frac{1}{|I|}\right) dk = \log(2k_{\text{max}}). \tag{54}$$

In particular for the family of states from (36) this bound is obtained if we take the limit $\gamma_\lambda \to 0$, (see Fig. 8).

If we combine (54) with the (BB) bound we find that

$$h_{\mathbf{x}}(\phi) \geq 1 - \log\left(\frac{2k_{\text{max}}}{\pi}\right) , \tag{55}$$

implying a lower bound on the entropy $h_{\mathbf{x}}$. For a modification $f(\mathbf{p}) = 1 + \beta \mathbf{p}^2$ this reads

$$h_{\mathbf{k}}(\psi) \leq \log(\pi) - \frac{1}{2}\log(\beta) \tag{56}$$

and

$$h_{\mathbf{x}}(\phi) \geq 1 + \frac{1}{2}\log(\beta). \tag{57}$$

Yet, this bound is not optimal, because the (BB) bound, the dashed line in Fig. 7, becomes tight only on Gaussian functions, see [46]. Still however, all entropy pairs have to lie "above" this line.

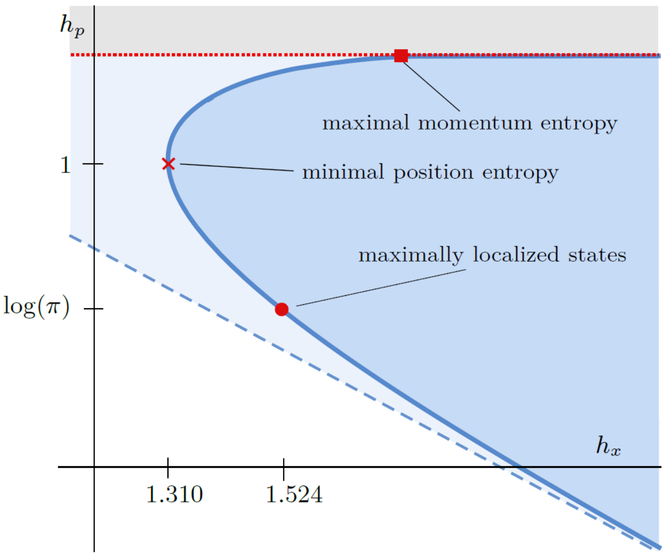


Fig. 7: Entropic uncertainty region in the $h_{\mathbf{x}} - h_{\mathbf{p}}$ plane for the modification $f(\mathbf{p}) = 1 + \beta \mathbf{p}^2$. The dashed line is the Bialynicki-Birula bound which is still valid in our setting but not longer optimal. The horizontal dotted line shows the upper bound on the entropy $h_{\mathbf{p}}$. The marked points correspond to the $h_{\mathbf{x}} - h_{\mathbf{p}}$ pairs for the maximally localized states, the state of minimal position entropy and the one with maximal momentum entropy, respectively. Physical states are conjectured to lie on the right of the solid line and below the dotted horizontal line.

In analogy with the standard result for the entropy bound, and motivated by our results of the previous section about the optimal uncertainty relation in terms of variances, we conjecture that the analogue of the (BB) curve corresponds to the states $\psi_\lambda(k) \propto \cos(\sqrt{\beta}k)^{\gamma_\lambda}$, which we saw for $\gamma_\lambda \geq 1$ represents the ground state of the deformed harmonic oscillator, while for $\gamma_\lambda < 1$ they still saturate the optimal bound in terms of variances, but can be seen as eigenstates of an imaginary frequency oscillator. This curve is shown as the convex solid line in Fig. 7.

We were unable to obtain an analytic form for the corresponding values of $h_{\mathbf{x}}$, which have been computed numerically. On the other hand, we can give a simple expression for $h_{\mathbf{k}}$ in terms of special functions. To this end let us observe that the states (36) can be also written as

$$\psi_\lambda(k) = \frac{1}{\sqrt{\kappa(\gamma)}} \exp\left(-\gamma \int_0^k p(k')dk'\right), \tag{58}$$

where we omit the argument $\gamma \equiv \gamma_\lambda$ for readability and

$$\kappa(\gamma) = \sqrt{\frac{\pi}{\beta}} \frac{\Gamma(1/2+\gamma)}{\Gamma(1+\gamma)}. \tag{59}$$

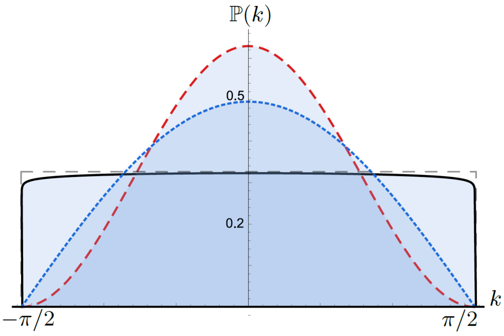


Fig. 8: Minimal length states (36) for $\beta = 1$ and $\gamma_\lambda = 1$ (red, dashed, variance), $\gamma_\lambda = 1/2$ (blue, dotted, Shannon entropy) and $\gamma_\lambda \approx 0$ (black, solid, min-entropy) as in Fig. 1 but in $k$-representation. These minimal length states differ from the states with maximal momentum uncertainty in the case of variances and Shannon entropy. For min-entropies however the minimal length state also yields maximal possible momentum uncertainty.

Thus, their entropy $h_{\mathbf{k}}$ reads

$$h_{\mathbf{k}} = \int_I dk \frac{1}{\kappa(\gamma)} \exp\left(-2\gamma \int_0^k p(k')dk'\right) \times \\ \times \left(2\gamma \int_0^k p(k')dk' + \log(\kappa(\gamma))\right), \tag{60}$$

or

$$h_{\mathbf{k}} = \log(\kappa(\gamma)) - \frac{1}{\kappa(\gamma)} \gamma \frac{d}{d\gamma} \kappa(\gamma). \tag{61}$$

Using (59), we finally find

$$h_{\mathbf{p}} = h_{\mathbf{k}} = \log\left(\sqrt{\frac{\pi}{\beta}} \frac{\Gamma(1/2+\gamma)}{\Gamma(1+\gamma)}\right) + \\ + \gamma \mathcal{N}(\gamma) - \gamma \mathcal{N}\left(\gamma - \frac{1}{2}\right), \tag{62}$$

with $\mathcal{N}(\gamma)$ the harmonic numbers.

This new boundary curve, shown in Fig. 7, can be divided into three parts, corresponding to different properties of the optimal states (58) (we fix here $\beta = 1$):

- for large positive values of $h_{\mathbf{x}}$ and large negative $h_{\mathbf{p}}$ the curve asymptotically reaches the (BB) bound, as expected (the role played by the cut-off $\beta$ can be neglected in this regime). As $h_{\mathbf{x}}$ decreases the curve starts bending and leaves the (BB) straight line. The solid circle shown in Fig. 7 denotes the states of maximal localization in terms of variances studied in the previous section, for which

$$h_{\mathbf{k}} = \log(2\pi/e) \text{ and } h_{\mathbf{x}} \simeq 1.374. \tag{63}$$

Till this point the optimal states are the $\psi_\lambda(k)$ with $\gamma_\lambda \geq 1$, i.e. the ground states of the deformed harmonic oscillators;

- as $\gamma$ falls below unity, the value of $h_{\mathbf{x}}$ continues to decrease, until $\gamma = 1/2$, which corresponds to the cross in Fig. 7. We have for this state

$$h_{\mathbf{k}} = 1 \text{ and } h_{\mathbf{x}} \simeq 1.310, \tag{64}$$

which represents the state of minimal entropy in position. We see that *minimal length in terms of entropy is not equivalent to minimal length in terms of variances.* Indeed, all optimal states in this branch of the curve have finite $\mathbf{x}$ and $\mathbf{p}$ variances, with the exception of the point $\gamma = 1/2$, see the previous section. The minimal position entropy can thus, be attained by considering a sequence of such states with $\gamma \to 1/2$. In this limit the variances of both $\mathbf{x}$ and $\mathbf{p}$ diverge (see Fig. 4), while their entropies stay finite;

- for even smaller values of $\gamma < 1/2$, $h_{\mathbf{x}}$ increases and so does $h_{\mathbf{k}}$ until it reaches its maximal value, corresponding to a constant wave function in $I$ ($\gamma = 0$), up to an arbitrary $k$ dependent phase. This state is shown in the entropy plane as the filled square in Fig. 7

$$h_{\mathbf{k}} = \log(\pi) \text{ and } h_{\mathbf{x}} \simeq 1.524, \tag{65}$$

This part of the curve, $1/2 > \gamma > 0$, corresponds to normalizable wave functions but with infinite variances for both $\mathbf{x}$ and $\mathbf{p}$. Thus, it is an open boundary for "physical" states, if by so we mean states which are in the domain of position and momentum operators.

The solid line and the part of the horizontal line $h_{\mathbf{k}} = \log(\pi)$ starting from the filled square bounds a convex region. Our conjecture is that this is in fact, the region in which all entropy pairs for states from $\mathcal{P}$ have to lie in. We cannot present here a proof of this, but we have performed a numerical scan of pairs $h_{\mathbf{x}} - h_{\mathbf{k}}$ corresponding to a random sample of 100000 states built from superposing low excited states (see Eq. (69) in next subsection).

### B. Min-entropy

Considering min-entropies in order to quantify the uncertainty of a measurement is meaningful for several reasons: on one hand $H_\infty$ (see (42)) sets a lower bound on all other entropies within the Rényi-$\alpha$ family, i.e. $H_\infty \leq H_\alpha$ for all $\alpha \in \mathbb{R}_+$. On the other hand it has a direct operational interpretation in the following sense: consider again an experimenter who now samples a (discrete) random variable $Y^I_\varepsilon$ and assume that she tries to guess the outcome of a particular sample. In this case the quantity $\exp\left(-H_\infty(Y^I_\varepsilon)\right)$ gives the highest guessing probability she can attain when doing so. In the same spirit, the min-entropy $h_\infty(w)$, i.e. the continuous counterpart/analogue of $H_\infty$, of a probability density $w$ is given as

$$h_\infty(w) = -\log\left(\operatorname*{ess\,sup}_x |w(x)|\right). \tag{66}$$

Here the essential supremum $\operatorname{ess\,sup}_x |w(x)|$ (66) is needed to correctly deal with sets of measure zero. In particular, when regarding a partitioning of an interval into bins in the limit of vanishing bin size, the essential supremum arises as the natural limit of a supremum over finite bins. Operationally, this quantity can therefore be understood as follows: consider we choose a partitioning into bins $\Omega_i$ of size $|\Omega_i|$ in a measurement scenario in which the outcomes are distributed according to a probability density $w(x)$. If the experimenter was to guess the bin in which the next measurement outcome will occur, the success probability $\mathbb{P}_i = \int_{\Omega_i} w(x)$ is upper bounded by $\mathbb{P}_i \leq c^*|\Omega_i|$ with $c^*$ some constant that may depend on the binning sizes, but is independent of the bin $i$. The min-entropy characterises the smallest constant $c^*$ for all possible binning sizes which is exactly given by $\operatorname{ess\,sup}_x |w(x)|$.

Using the framework developed in this paper we are able to directly compute the minimal length in terms of min-entropy

$$\Gamma^\infty_{\min} := \inf_{\psi\in\mathcal{P}} \left[-\log\left(\operatorname*{ess\,sup}_x |\phi(x)|^2\right)\right] \tag{67}$$

$$= -\log\left(\sup_{\psi\in\mathcal{P}} \operatorname*{ess\,sup}_x |\phi(x)|^2\right), \tag{68}$$

as follows: consider the “particle in a box” basis for the interval $[-k_{\max}, k_{\max}]$, i.e.

$$\psi_n(k) = \frac{1}{\sqrt{k_{\max}}} \sin\left[\frac{\pi n}{2k_{\max}}(k - k_{\max})\right], \tag{69}$$

with its Fourier transform

$$\phi_n(x) = \left(\frac{\pi n^2 k_{\max}}{2}\right)^{1/2} \frac{\sin(k_{\max}x - \frac{\pi n}{2})}{k^2_{\max}x^2 - \frac{\pi^2 n^2}{4}} \mathrm{e}^{-\frac{\pi n}{2}i}. \tag{70}$$

We can then decompose any $\psi \in \mathcal{P}$ by $\psi(k) = \sum_n \alpha_n \psi_n(k)$ with a square summable sequence $\alpha_n$. In the same way, its Fourier transform $\phi = \mathcal{F}[\psi]$ reads $\phi(x) = \sum_n \alpha_n \phi_n(x)$. We therefore have

$$\Gamma^\infty_{\min} = -\log\left(\sup_{\alpha:||\alpha||_{l^2}=1} \sup_x \big|\sum_n \alpha_n \phi_n(x)\big|^2\right), \tag{71}$$

where we used the fact that the $\phi_n(x)$ are smooth to replace the essential supremum with the ordinary supremum. The term $|\sum_n \alpha_n \phi_n(x)|^2$ can be rewritten as a scalar product $|\langle \alpha_n|\phi_n(x)\rangle_{l^2}|^2$ which, due to the Cauchy-Schwarz inequality, is maximised by $\langle \phi_n(x)|\phi_n(x)\rangle^2$. Also note that by choosing the phase of $\psi$ appropriately we can without loss of generality set the maximum of the $\phi_n(x)$ to be at $x = 0$. Some algebra shows that

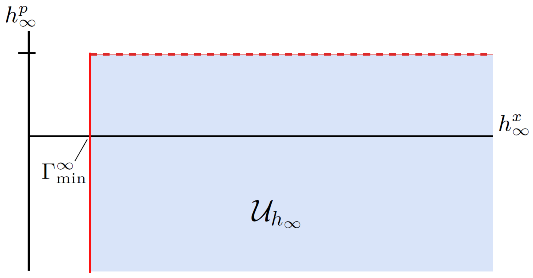


Fig. 9: Entropic uncertainty region for min-entropies $h_\infty$. The minimal length $\Gamma_{\text{min}}^\infty$ is analytically computed in (72). As for Shannon entropies, the min-entropy of momentum attains an upper bound (dashed line), $\max h_\infty^p = \log(2k_{\text{max}})$, for the uniform distribution on the interval $I$. Numerical evaluation indicates that any entropy pair satisfying these two constraints can be attained. Hence the uncertainty region $\mathcal{U}_{h_\infty}$ and the corresponding trade-off curve are given by the shaded region and the solid line, respectively.

$\sum_n |\phi_n(0)|^2 = k_{\text{max}}/\pi$, proving that the minimal length in terms of min-entropy is given by (see Fig. 9)

$$\Gamma_{\text{min}}^\infty = -\log(k_{\text{max}}/\pi) \ . \tag{72}$$

In particular, whenever the variance-based minimal length is normalised, i.e. $l_{\text{min}}^2 = 1$, the minimal length in terms of min-entropy is given bys

$$\Gamma_{\text{min}}^\infty = \log 2 \equiv 1 \,[\text{bit}] \ , \tag{73}$$

i.e. the *minimal length is exactly one bit.*

## V. CONCLUDING REMARKS AND FUTURE DIRECTIONS

Minimal length is to be understood as the minimal possible uncertainty about position measurements due to a modification of the Heisenberg algebra. Nevertheless, two questions remained to be settled: first, under what assumptions does a modification actually *imply* minimal length? This line of thought is in contrast to previous attempts where only the existence of operators was shown which both show minimal length and satisfy the algebra. Despite its importance this question has not been answered so far to the best of our knowledge. One of our main results, Theorem 1, clarifies this issue and proves under physically well-motivated assumptions the uniqueness of such operators. As such it justifies results obtained in previous literature.

Second, one should operationally motivate the choice of uncertainty measure used to define minimal length, since such measures are far from unique. The choice to use variances is in this context not at all an obvious one. Instead entropic measures are known to have an operational interpretation while not suffering from a number of severe deficiencies that variances show if interpreted as a measure of uncertainty. We therefore introduce and show implications of an entropic formulation of minimal length.

Our main results can be summarized as follows: the states which correspond to the maximal possible localization in the $x$ space in terms of variance $\Delta\mathbf{x}$ and of the Shannon entropy $h_\mathbf{x}$ or min-entropy $h_\infty^x$, are different, showing that the physical notion of minimal length itself depends on the particular choice adapted to its operational meaning. The min-entropy lower bounds all other Rényi-$\alpha$ entropies; likewise the minimal length in terms of min-entropy is a lower bound to all other entropic minimal lengths and turns out to be exactly one bit. On the other side, entropic bounds also show another novel feature, namely the presence of an upper bound on the entropy in $p$ space $h_\mathbf{p}$ or $h_\infty^p$ for "physical states". In other words deformations of the standard Heisenberg algebra leading to a minimal length lead to a lower limit on the information we can get on a given state in terms of its momentum distribution. This is not the case if we use variances to quantify this information, since the value of $\Delta p$ can be arbitrarily large.

We have established a framework that allows to compute optimal and state-independent uncertainty relations for modified Heisenberg algebras (see e.g. Theorem 2 and Corollary 3). Optimal and state-independent uncertainty relations directly yield minimal lengths, but contain much more information as they describe the uncertainties for all quantum states in the modified algebra.

One of the natural generalizations would be to extend our setting to higher dimensions. We hope that the study of the entropic uncertainties in three and four dimensions may shed some light on the plausible connection that might exist between previous limits such as bound on information storage (holographic bound) [47–49], bounds on information scrambling/chaos [50], bounds on quantum evolution [51–53] (see also the book [54]) and quantum computation/complexity [55–57].

It might be also interesting to apply the general treatment of the geometry of the Heisenberg algebra in the context of Aharonov's reformulation of quantum theory in terms of modular variables, see [58], in the case we have considered of a deformed Heisenberg algebra.

Finally, our new entropic bounds on $h_\mathbf{x}$ and $h_\mathbf{p}$ are directly applicable to the entropic steering inequalities formulated in [59] and thus lead to new limitations on the amount of entanglement that can be shared between two distant parties governed by a modified Heisenberg algebra.

It is worth noting that one of our main results, Theorem 1, directly links the study of quantum physics in a modified algebra to the study of classical information processing of bandlimited analog signals. Quantum states are then replaced by the complex current in a wire, the considered observables change from position and momentum to time and frequency (again linked by Fourier transformation). A momentum cut-off due to a modified algebra can therefore be understood as a frequency (or "band") limitation of the complex current. We already exploited this analogy to some extent by considering operationally more relevant uncertainty measures as was first done in the well-studied field [60–62] of classical information processing. By considering the Nyquist sampling theorem steps into this direction have been taken by [63]. However, we strongly believe that one can obtain many more fundamental insights in the field of modified algebras by just transferring results and concepts from classical information theory.

**Acknowledgments**

The authors are grateful to Saurya Das, Reinhard F. Werner, Tobias Osborne, and Inken Siemon for discussions related to this work. They especially thank Lars Dammeier for perpetually emphasising the importance of entropic uncertainty relations. They also thank Achim Kempf for reading a preliminary version of the manuscript and valuable comments. Special thanks also to Michael Hall for helpful comments on the definition of length in terms of entropies. This work was supported by the European grants DQSIM, QFTCMPS and SIQS and by the cluster of excellence EXC 201 Quantum Engineering and Space-Time Research. GM acknowledges the support of INFN, I.S. TASP. RS acknowledges the support from the BMBF project Q.com-q.

# Conclusions

In order to draw a conclusion to this thesis we can have a second look at our initial three questions:

### (1.) How can we quantify measurement/preparation uncertainty?

The choice of the 'correct' error measure for quantifying measurement uncertainty has been the content of several debates within the last decade (most prominently [Oza13] vs. [BLW13]). In this thesis we adapted the position of [BLW14b] and identify measurement uncertainty with the task of finding good approximate joint measurements.

In chapter Ch. 2 we provided a framework that allows to construct such error measures in a wide range of situations. Thereby, the underlying construction can be seen as a direct generalization of the error measures from [BLW14b]. The core of our framework are cost functions. They are regarded as an open input to the mathematical construction of an error measure. In a particular situation, this open input can be taken as a tool to model the nature of the underlying physics in order to get a meaningful measurement uncertainty relation.

We used this tool in chapter Ch. 5 for deducing measurement uncertainty relations for information theoretic applications. We considered error measures based on the discrete metric. This metric is the natural choice for comparing two measurements with coinciding but unstructured outcome sets. However, our construction is not limited to this, since, by a suitable choice of the cost function, measurements with different outcome sets can be compared, as well. We had to take care of this, in the definition of entropic measurement uncertainty relations.

Previous literature on preparation uncertainty relations was concentrated on taking variances and Shannon entropies as deviation measure. In this thesis, i.e. in chapter Ch. 3, we additionally introduced a generalized notion of deviation based on a general cost function. This construction yields variances and Shannon entropies as particular examples, obtained by taking the square euclidean distance or the self-information as cost function.

### (2.) How can we compute measurement/preparation uncertainty relations?

In chapter Ch. 2 we proved the convexity of measurement uncertainty regions. This implies that all uncertainty relations can be described by a collection of linear uncertainty relations. For measurements on finite Hilbert spaces and with finite outcome sets, those linear relations can be computed by semidefinite programming. Hence, we can conclude that, in a finite setting, measurement uncertainty relations are efficiently computable.

However, this changes for continuous outcome sets. Here, the methods developed for finite observables are not applicable any more. Hence, computing optimal entropic measurement uncertainty relations remains an open problem. The best we achieved in this case is to provide lower bounds in terms of uncertainty relations based on the discrete metric and, for sharp measurements, by linear entropic preparation uncertainty relations.

For finite Hilbert spaces and deviation measures based on cost functions with finite support, the uncertainty region consists of a union of finitely many joint numerical ranges. Hence, in this case, preparation uncertainty relations can be computed efficiently, as well.

For preparation uncertainty in terms of variances an algorithmic method is provided, which allows to satisfactorily compute linear uncertainty relations. However, variance based uncertainty regions do not necessarily have to be convex. Hence, non-linear improvements are possible and an open topic. The algorithmic method also works for general POVMs. This was used in [SDW7] to improve an existing entanglement-detection scheme with respect to local noise. Here the new method allowed us to compute the influence of local noise sources, i.e. of noisy detectors, within a fully quantum mechanical framework which then leads to more sensitive confidence intervals for a detection of entanglement.

We provided algorithmic methods for computing preparation uncertainty in terms of entropies, too. Unfortunately, these methods produce, in finite runtime, only upper bounds on the optimal uncertainty and do not provide a gap estimate. Hence, they should better not be used in critical applications like security proofs or entanglement detection schemes. For the special, but very common, case of two sharp measurements, we proved an additivity theorem for the global uncertainty of local measurements. In this case only local bounds have to be computed. An interesting application of this result are multi qubit systems, here all local bounds can be computed easily.

### (3.) Can we find connections between measurement and preparation unceratinty relations?

By using the constructions provided in this thesis, measurement and preparation uncertainty can both be based on a cost function. Therefore, it makes sense to compare a measurement and a preparation uncertainty relation for a common cost function. Within these terms, an answer to the above question is given by Thm. 3.4 in chapter Ch. 3.

Here we proved that the statement

$$\text{preparation uncertainty} \preceq \text{measurement uncertainty}$$

holds for linear measurement uncertainty relations between sharp measurements. Interestingly, the analogous statement will fail if we compare the corresponding non-linear uncertainty relations, here we expect no clear ordering. An analogous statement will also fail for non-sharp measurements.

In chapter Ch. 5, we had a closer look at this statement for information theoretic quantities. For the special case of two sharp measurements $A$ and $B$ and equal weights, we can assign a second meaning to the maximal overlap, $c^* = \max_{ij} |\langle \phi_i^A | \phi_j^B \rangle|$ which was previously only used as indicator for the presence of preparation uncertainty relations: He we have the bound

$$\varepsilon^{dm}(A|A') + \varepsilon^{dm}(B|B') \geq 2(1 - c^*),$$

for errors based on the discrete metric, and

$$\varepsilon^{info}(A|\mathtt{A}') + \varepsilon^{info}(B|\mathtt{B}') \geq -2\log(c^*),$$

for errors based on the self-information.

# Literature

[AAHB16] A. A. Abbott, P.-L. Alzieu, M. J. W. Hall, and C. Branciard. Tight state-independent uncertainty relations for qubits. *Mathematics*, 4(1), 2016. arXiv:1512.02383.

[AB16] A. A. Abbott and C. Branciard. Noise and disturbance of qubit measurements: An information-theoretic characterization. *Phys. Rev. A*, 94:062110, 2016. arXiv:1607.00261.

[App98] D. M. Appleby. Concept of experimental accuracy and simultaneous measurements of position and momentum. *Int. J. Theor. Phys*, 37:1491–1509, 1998.

[Bal69] L. Ballentine. The uncertainty principle and the statistical interpretation of quantum mechanics. *Can. J. Phys.*, 47(21):2417–2419, 1969.

[BB15] T. Bullock and P. Busch. Measurement uncertainty relations: characterising optimal error bounds for qubits. 2015. arXiv:1512.00104.

[BE67] M. Bunge (Editor). Quantum theory and reality. 1967.

[BGT17] A. Barchielli, M. Gregoratti, and A. Toigo. Measurement uncertainty relations for position and momentum: Relative entropy formulation. *Entropy*, 19(301), 20017.

[BGT18] A. Barchielli, M. Gregoratti, and A. Toigo. Measurement uncertainty relations for discrete observables: Relative entropy formulation. *Comm. Math. Phys.*, 357(3):1253–1304, 2018.

[BHL07] P. Busch, T. Heinonen, and P. Lahti. Heisenberg's uncertainty principle. *Physics Reports*, 452(6):155 – 176, 2007.

[BHOW14] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. Wilde. Noise and disturbance in quantum measurements: An information-theoretic characterisation. *Phys. Rev. Lett.*, 112:050401, 2014. arXiv:1310.6603.

[BKW16] P. Busch, J. Kiukas, and R.F. Werner. Sharp uncertainty relations for number and angle. 2016. arXiv:1601.03843.

[BLPY16] P. Busch, P. Lahti, J.-P. Pellonpää, and K. Ylinen. *Quantum measurement.* Springer International Publishing, 2016.

[BLW13] P. Busch, P. Lahti, and R. F. Werner. Proof of Heisenberg's error-disturbance relation. *Phy. Rev. Lett.*, 111:160405, 2013. arXiv:1306.1565.

[BLW14a] P. Busch, P. Lahti, and R. F. Werner. Heisenberg uncertainty for qubit measurements. *Phys. Rev. A*, 89:012129, 2014.

[BLW14b] P. Busch, P. Lahti, and R. F. Werner. Measurement uncertainty relations. *J. Math. Phys.*, 55:042111, 2014. arXiv:1312.4392.

[BR18] S. Boughn and M. Reginatto. Another look through heisenberg's microscope. *European Journal of Physics*, 39(3):035402, 2018.

[Bus85] P. Busch. Indeterminacy relations and simultaneous measurements in quantum theory. *International Journal of Theoretical Physics*, 24(1):63–92, 1985.

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 204.

[CBTW17] Patrick Coles, Mario Berta, Marco Tomamichel, and Stephanie Wehner. Entropic uncertainty relations and their applications. *Rev. Mod. Phys.*, 89, 2017. and arXiv:1511.04857.

[CF05] N. Cerf and J. Fiurasek. Optical quantum cloning - a review. *Progress in Optics*, 49:455, 2005. arXiv:quant-ph/0512172.

[CSUG17] A. C. Costa Sprotte, R. Uola, and O. Gühne. Steering criteria from general entropic uncertainty relations. 2017. arXiv:1710.04541.

[DBKMR05] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of operations research*, 134(1):19–67, 2005.

[Deu83] D. Deutsch. Uncertainty in quantum measurements. *Phys. Rev. Lett.*, 50:631–633, 1983.

[dGMSS18] H. de Guise, L. Maccone, B. C. Sanders, and N. Shukla. State-independent preparation uncertainty relations. 2018. arXiv:1804.06794.

[DSW15] L. Dammeier, R. Schwonnek, and R.F. Werner. Uncertainty relations for angular momentum. *New J. Phys.*, 9(17):093946, 2015. arXiv:1505.00049.

[Eul58] L. Euler. Elementa doctrine solidorum. *Novi comm. acad. scientiarum imperialis petropolitanae*, (4):72–108, 1758.

[Fad57] D. K. Fadeev. *Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschemas*. Deutscher Verlag der Wissenschafte, 1957. Arbeiten zur Informationstheorie I.

[FFB$^+$12] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B Scholz, M. Tomamichel, and R. F Werner. Continuous variable quantum key distribution: Finite-key analysis of composable security against coherent attacks. *Phys. Rev. Lett.*, 109:100502, 2012. arXiv:1112.2179.

[Fra15] L. Fraatz. Mmessunschärfe-Relation für diskrete Observablen. 2015. Bachelor thesis, Leibniz Universität Hannover.

[Gü04] O. Gühne. Detecting quantum entanglement: entanglement witnesses and uncertainty relations. 2004. PhD thesis, Universität Hannover.

[GB05] A. Gunawardana and W. Byrne. Convergence theorems for generalized alternating minimization procedures. *Journal of Machine Learning Research*, 6:2049–2073, 2005.

[GL04] O. Guehne and M. Lewenstein. Entropic uncertainty relations and entanglement. *Phys. Rev. A*, 70:022316, 2004. arXiv:quant-ph/0403219.

[GMR03] G. Ghirardi, L. Marinatto, and R. Romano. An optimal entropic uncertainty relation in a two-dimensional Hilbert space. *Phys. Lett. A*, 317:32–36, 2003.

[Has17] A. K. Hashagen. Universal asymmetric quantum cloning revisited. *Q. Inf. Comp.*, 17(9):0747–0778, 2017.

[Hei27] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys.*, 43:172–198, 1927.

[Hei44] W. Heisenberg. Physikalische Prinzipien der Quantentheorie. 1944.

[Hei84] W. Heisenberg. The actual content of quantum theoretical kinematics and mechanics. *Nasa technical memorandum*, page 77379, 1984.

[HKR15] T. Heinosaari, J. Kiukas, and D. Reitzner. Noise robustness of the incompatibility of quantum measurements. *Phys. Rev. A*, (92):022115, 2015.

[HO10] J. Hendrickx and A. Olshevsky. Matrix p-norms are NP-hard to approximate if $p \neq 1, 2, \infty$. *SIAM J. M. A. A.*, 31:2802–2812, 01 2010. arXiv:0908.1397.

[HPDR11] Q. He, S. Peng, D. Drummond, and M. Reid. Planar quantum squeezing and atom interferometry. *Phys. Rev. A.*, (84):022107, 2011. Master thesis, Leibniz Universität Hannover.

[HT03] H. F. Hofmann and S. Takeuchi. Violation of local uncertainty relations as a signature of entanglement. *Phys.Rev. A.*, 68:032103, 2003. arXiv:quant-ph/0212090.

[HU16] Jan Hilgevoord and Jos Uffink. The uncertainty principle. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition, 2016.

[Hua12] Y. Huang. Variance-based uncertainty relations. *Phys. Rev. A*, 86:024101, 2012. arXiv:1012.3105.

[HW18] A. K. Hashagen and M. M. Wolf. Universality and optimality in the information-disturbance tradeoff. 2018. arXiv:1802.09893.

[Jav16] R. Javadi. Messunschärfe für Qubitstrings. 2016. Bachelor thesis, Leibniz Universität Hannover.

[Kö17] K. K. König. Kapazitäten für das Verschließen von Information. 2017. Master thesis, Leibniz Universität Hannover.

[Kar84] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[Ken27] E. H. Kennard. Zur Quantenmechanik einfacher Bewegungstypen. *Z. Phys.*, 44:326–352, 1927.

[KM97] A. Kempf and G. Mangano. Minimal length uncertainty relation and ultraviolet regularization. *Physical Review D*, 55(12):7909, 1997. arXiv: hep-th/9612084.

[KMM95] A. Kempf, G. Mangano, and R. B. Mann. Hilbert space representation of the minimal length uncertainty relation. *Physical Review D*, 52(2):1108, 1995. arXiv: hep-th/9412167.

[Kus16] K. Kusmerik. Measurement uncertainty relations with finite operating range. 2016. Bachelor thesis, Leibniz Universität Hannover.

[KW14] S. Kechrimparis and S. Weigert. Heisenberg uncertainty relation for three canonical observables. *Phys. Rev. A*, (90):062118, 2014. arXiv:1407.0083.

[KW16] S. Kechrimparis and S. Weigert. Preparational uncertainty relations for n continuous variables. *Mathematics*, 4(3), 2016. arXiv:1606.09148.

[LP61] H. J. . Landau and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty ii. *Bell System Technical Journal*, 40(1):65–84, 1961.

[LP62] H. J. . Landau and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty iii. *Bell System Technical Journal*, 41(4):1295–1336, 1962.

[MC67] H. Margenau and L. Cohen. Probabilities in quantum mechanics. In *Quantum theory and reality*. Springer, 1967.

[MU88] H. Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Phys. Rev. Lett.*, 60:1103–1106, 1988.

[Oza04] M. Ozawa. Uncertainty relations for noise and disturbance in generalized quantum measurements. *Annals of Physics*, 311(2):350 – 416, 2004.

[Oza13] M. Ozawa. *Disproving Heisenberg's error-disturbance relation.* 2013. arXiv:1308.3540.

[Pop89] K. Popper. Quantum theory and the schism of physics. 1989.

[Ré61] A. Rényi. On measures of entropy and information. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, 1961.

[RMM17a] A. Riccardi, C. Macchiavello, and L. Maccone. Multipartite steering inequalities based on entropic uncertainty relationss. 2017. arXiv:1711.09707.

[RMM17b] A. Riccardi, C. Macchiavello, and L. Maccone. Tight entropic uncertainty relations for systems with dimension three to five. *Phys. Rev. A*, 95:032109, 2017. arXiv:1701.04304.

[Rob29] H. P. Robertson. The uncertainty principle. *Phys. Rev.*, 34:163–164, 1929.

[Roh00] J. Rohn. Computing the norm $\|A\|_{\infty,1}$, is NP-hard. *Lin. and Multilin. Alg.*, 47(3):195–204, 2000.

[RRW13] D. Reeb, D. Reitzner, and M. M. Wolf. *J. Phys. A: Math. Theor*, 46:462002, 2013.

[RSH16] J. M. Renes, V. B. Scholz, and S. Huber. Uncertainty relations: An operational approach to the error-disturbance tradeoff. *Quantum*, 1(20), 2016. arXiv:1612.02051.

[Sch30] E. Schrödinger. Zum Heisenbergschen Unschärfeprinzip. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse.*, (14):296–303, 1930.

[Sch35] E. Schrödinger. Die gegenwärtige Situation in der Quantenmechanik. *Naturwiss.*, 23(48):807–812, 1935.

[Sha48] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[SP61] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, fourier analysis and uncertainty i. *Bell System Technical Journal*, 40(1):43–63, 1961.

[SR95] J. Sánchez-Ruiz. Improved bounds in the entropic uncertainty and certainty relations for complementary observables. *Phys. Lett. A*, 201:125–131, 1995.

[SR98] J. Sánches-Ruiz. Optimal entropic uncertainty relation in two-dimensional Hilbert space. *Phys. Lett. A*, (244):189–195, 1998.

[Ste05] D. Steinberg. Computation of matrix norms with applications to robust optimization. *M.Sc. thesis, Technion—Israel Institute of Technology*, 2005.

[SW18a] R. Schwonnek and R. F. Werner. Properties of the wigner distribution for n arbitrary operators. 2018. arXiv:1802.08343.

[SW18b] R. Schwonnek and R. F. Werner. Wigner distributions for n arbitrary operators. 2018. arXiv:1802.08342.

[SZ18] K. Szymański and K. Zyczkowski. Geometric and algebraic origins of additive uncertainty relations. 2018. arXiv:1804.06191.

[Toe18] O. Toeplitz. Das algebraische Analogon zu einem Satze von Frejér. *Mathematische Zeitschrift*, (2):187–197, 1918.

[UBGP15] R. Uola, C. Budroni, O. Gühne, and J.-P. Pellonpää. One-to-one mapping between steering and joint measurability problems. *Phys. Rev. Lett.*, 115:230402, 2015. arXiv:1507.08633.

[VB96] L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996.

[Vil09] C. Villiani. *Optimal transport: old and new.* Springer, 2009.

[Wei11] S. Weis. Quantum convex support. *Linear Algebra and its Applications*, (435):3168–3188, 2011.

[Wer16] R. F. Werner. Uncertainty relations for general phase spaces. *Proceedings QCMC 2014*, 20016. arXiv:1604.00566.

[Wer98] R. F. Werner. Optimal cloning of pure states. *Phys. Rev. A*, 58:1827, 1998.

[Wer04] R. F. Werner. The uncertainty relation for joint measurement of position and momentum. *Q. Inf. Comp.*, 4:546–562, 2004.

[Wer17] R. F. Werner. Unschärfe von Heisenberg bis heute. In *Quanten 5.* Hirzel, 2017.

[Wey28] H. Weyl. *Gruppentheorie und Quantenmechanik.* Hirzel, 1928.

[Wil13] M. Wilde. *Quantum Information Theory.* Cambridge University Press, 2013. arXiv:1106.1445.

[WW10] S. Wehner and A. Winter. Entropic uncertainty relations – a survey. *New J. Phys.*, 12:025009, 2010. arXiv:0907.3704.

[XGM$^+$17] Y. Xiao, C Guo, F Meng, N. Jing, and M.-H. Yung. Incompatibility of observables as state-independent bound of uncertainty relations. 2017. arXiv:1706.05650.

[YO14] S. Yu and C. H. Oh. Optimal joint measurement of two observables of a qubit. 2014. arXiv:1402.3785.

# Literature to: [SRW16]

[1] Heisenberg, W. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschr. Phys.* **1**927, *43*, 172–198.

[2] Kennard, E. Zur Quantenmechanik einfacher Bewegungstypen. *Zeitschr. Phys.* **1**927, *44*, 326–352.

[3] Werner, R.F. The uncertainty relation for joint measurement of position and momentum. *Quant. Inform. Comput.* **2**004, *4*, 546–562. and arXiv:quant-ph/0405184.

[4] Ozawa, M. Uncertainty relations for joint measurements of noncommuting observables. *Phys. Lett. A* **2**004, *320*, 367–374.

[5] Busch, P.; Lahti, P.; Werner, R.F. Quantum root-mean-square error and measurement uncertainty relations. *Rev. Mod. Phys.* **2**014, *86*, 1261–1281. and arXiv:.

[6] Appleby, D.M. Concept of experimental accuracy and simultaneous measurements of position and momentum. *Int. J. Theor. Phys.* **1**998, *37*, 1491–1509.

[7] Busch, P.; Lahti, P.; Werner, R.F. Proof of Heisenberg's Error-Disturbance Relation. *Phys. Rev. Lett.* **2**013, *111*, 160405. and arXiv:1306.1565.

[8] Ozawa, M. Disproving Heisenberg's error-disturbance relation **2**013. arXiv:1308.3540.

[9] Busch, P.; Lahti, P.; Werner, R.F. Measurement uncertainty relations. *J. Math. Phys.* **2**014, *55*, 042111. and arXiv:1312.4392.

[10] Appleby, D.M. Quantum Errors and Disturbances: Response to Busch, Lahti and Werner. arXiv:1602.09002, 2016.

[11] Busch, P.; Heinosaari, T. Approximate joint measurement of qubit observables. *Quantum Inf. Comp.* **2**008, *8*, 0797–0818. and arXiv:0706.1415.

[12] Bullock, T.; Busch, P. Incompatibillity and Error Relations for Qubit Observables. arXiv:1402.6711, 2015.

[13] Busch, P.; Lahti, P.; Werner, R.F. Heisenberg uncertainty for qubit measurements. *Phys. Rev. A* **2**014, *89*, 012129.

[14] Dammeier, L.; Schwonnek, R.; Werner, R.F. Uncertainty Relations for Angular Momentum. *New J. Phys.* **2**015, *17*, 093046. and arXiv:1505.00049.

[15] Werner, R.F. Uncertainty relations for general phase spaces. *Front. Phys.* **2**016, *11*, 110305. proceedings of the QCMC 2014, and arXiv:1601.03843.

[16] Busch, P.; Kiukas, J.; Werner, R.F. Sharp uncertainty relations for number and angle. arxiv:1604.00566, 2016.

[17] Werner, R.F. Quantum harmonic analysis on phase space. *J. Math. Phys.* **1**984, *25*, 1404–1411.

[18] Vandenberghe, L.; Boyd, S. *Convex Optimization*; Cambridge UP, 2004.

[19] Vandenberghe, L.; Boyd, S. Semidefinite Programming. *SIAM Rev.* **1**996, *38*, 49–95.

[20] Grant, M.; Boyd, S. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. `http://cvxr.com/cvx`, 2014.

[21] Grant, M.; Boyd, S. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*; Blondel, V.; Boyd, S.; Kimura, H., Eds.; Lecture Notes in Control and Information Sciences, Springer-Verlag Limited, 2008; pp. 95–110. `http://stanford.edu/~boyd/graph_dcp.html`.

[22] Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, 1970.

[23] Nikaidô, H. On von Neumann's minimax theorem. *Pacific J. Math.* **1**954, *4*, 65–72.

[24] Sion, M. On general minimax theorems. *Pac. J. Math*, *8*, 171–176.

[25] Holevo, A.S. Statistical Decision Theory for Quantum Systems. *J. Multivariate Anal.* **1**973, *3*, 337.

[26] Yuen, H.P.; Kennedy, R.S.; Lax, M. Optimum testing of multiple hypotheses in quantum detection theory. *IEEE Trans. Inf. Theory*, *IT-21*.

[27] Villani, C. *Optimal Transport: Old and New*; Springer, 2009.

# Literature to: [SDW17]

[1] Earl Kennard. Zur Quantenmechanik einfacher Bewegungstypen. *Zeitschr. Phys.*, 44:326–352, 1927.

[2] Werner Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschr. Phys*, 43(3-4):172–198, 1927.

[3] For hermitian operators we define the variance of a state $\rho$ by $\Delta^2_\rho(A) = \langle A^2\rangle_\rho - \langle A\rangle^2_\rho = \operatorname{tr}\rho A^2 - (\operatorname{tr}\rho A)^2$.

[4] Holger F. Hofmann and Shigeki Takeuchi. Violation of local uncertainty relations as a signature of entanglement. *Phys. Rev. A.*, 68:032103, 2003. and arXiv:quant-ph/0212090.

[5] Otfried Gühne and Géza Tóth. Entanglement detection. *Phys. Rep.*, 474(1–6):1 – 75, 2009.

[6] Otfried Gühne. *Detecting quantum entanglement: entanglement witnesses and uncertainty relations*. PhD thesis, Universität Hannover, 2004.

[7] Lars Dammeier, René Schwonnek, and Reinhard F. Werner. Uncertainty relations for angular momentum. *New Journal of Physics*, 17(9):093046, 2015. and arXiv:1505.00049.

[8] Qiongyi He, Shi-Guo Peng, Peter D. Drummond, and Margaret D. Reid. Planar quantum squeezing and atom interferometry. *Phys. Rev. A*, 84:022107, Aug 2011. and arXiv:1101.0448.

[9] Alastair A. Abbott, Pierre-Louis Alzieu, Michael J. W. Hall, and Cyril Branciard. Tight state-independent uncertainty relations for qubits. *Mathematics*, 4(1), 2016. and arXiv:1512.02383.

[10] Paul Busch, Pekka Lahti, and Reinhard F. Werner. Measurement uncertainty relations. *J. Math. Phys.*, 55(4):042111, 2014. and arXiv:1312.4392.

[11] Reinhard F. Werner. Uncertainty relations for general phase spaces. *Frontiers of Physics*, 11(3):110305, 2016. and arXiv:1601.03843.

[12] Paul Busch, Jukka Kiukas, and Reinhard F. Werner. Sharp uncertainty relations for number and angle. 2016. and arXiv:1604.00566.

[13] Yichen Huang. Variance-based uncertainty relations. *Phys. Rev. A*, 86:024101, Aug 2012. and arXiv:1012.3105.

[14] Hans Maassen and Jos Uffink. Generalized entropic uncertainty relations. *Phys. Rev. Lett.*, 60:1103–1106, 1988.

[15] Kais Abdelkhalek, René Schwonnek, Hans Maassen, Fabian Furrer, Jörg Duhme, Philippe Raynal, Berthold-Georg Englert, and Reinhard F Werner. Optimality of entropic uncertainty relations. *Int. J. Quant. Inf.*, 13(06):1550045, 2015. and arXiv:1509.00398.

[16] Patrick Coles, Li Yu, and Michael Zwolak. Relative entropy derivation of the uncertainty principle with quantum side information. 2011. and arXiv:1105.4865.

[17] Jorge Sánchez-Ruiz. Improved bounds in the entropic uncertainty and certainty relations for complementary observables. *Phys. Lett. A*, 201:125–131, 1995.

[18] Jorge Sánches-Ruiz. Optimal entropic uncertainty relation in two-dimensional Hilbert space. *Phys. Lett. A*, 244:189–195, 1998.

[19] Gian Carlo Ghirardi, Luca Marinatto, and Raffaele Romano. An optimal entropic uncertainty relation in a two-dimensional Hilbert space. *Phys. Lett. A*, 317:32–36, 2003. and arXiv:quant-ph/0310120.

[20] Radosław Adamczak, Rafał Latała, Zbigniew Puchała, and Karol Życzkowski. Asymptotic entropic uncertainty relations. *J. Math. Phys.*, 57:032204, 2016. and arXiv:1412.7065.

[21] Patrick Coles and Fabian Furrer. State-dependent approach to entropic measurement–disturbance relations. *Phys. Lett. A*, 379:105–112, 2015. and arXiv:1311.7637.

[22] Patrick Coles, Roger Colbeck, Li Yu, and Michael Zwolak. Uncertainty relations from simple entropic properties. *Phys. Rev. Lett.*, 108:210405, 2012. and arXiv:1112.0543.

[23] Berthold-Georg Englert, Dagomir Kaszlikowski, Leong Chuan Kwek, and Wei Hui Chee. Wave-particle duality in multi-path interferometers: General concepts and three-path interferometers. *Int. J. Quant. Inf.*, 6:129–157, 2008. and arXiv:0710.0179.

[24] Patrick Coles, Mario Berta, Marco Tomamichel, and Stephanie Wehner. Entropic uncertainty relations and their applications. *Rev. Mod. Phys.*, 89, 2017. and arXiv:1511.04857.

[25] Alexey E. Rastegin. Rényi formulation of the entropic uncertainty principle for POVMs. *J. Phys. A*, 43:155302, 2010.

[26] Łukasz. Rudnicki, Zbigniew Puchała, and Karol Życzkowski. Strong majorization entropic uncertainty relations. *Phys. Rev. A*, 89, 2014. and arXiv:1402.0129.

[27] Zbigniew Puchała, Łukasz. Rudnicki, and Karol Życzkowski. Majorization entropic uncertainty relations. *J. Phys. A*, 46:272002, 2013. and arXiv:1304.7755.

[28] Shmuel Friedland, Vlad Gheorghiu, and Gilad Gour. Universal uncertainty relations. *Phys. Rev. Lett.*, 111(23):230401, 2013. and arXiv:1304.6351.

[29] Spiros Kechrimparis and Stefan Weigert. Heisenberg uncertainty relation for three canonical observables. *Phys. Rev. A*, 90:062118, Dec 2014. and arXiv:1407.0083.

[30] Spiros Kechrimparis and Stefan Weigert. Preparational uncertainty relations for n continuous variables. *Mathematics*, 4(3), 2016. and arXiv:1606.09148.

[31] Jędrzej Kaniewski, Marco Tomamichel, and Stephanie Wehner. Entropic uncertainty from effective anticommutators. *Phys. Rev. A*, 90:012332, 2014. and arXiv:1402.5722,.

[32] David Deutsch. Uncertainty in quantum measurements. *Phys. Rev. Lett.*, 50:631–633, Feb 1983.

[33] Howard Percy Robertson. The uncertainty principle. *Phys. Rev.*, 34:163–164, 1929.

[34] Lorenzo Maccone and Arun K. Pati. Stronger uncertainty relations for all incompatible observables. *Phys. Rev. Lett.*, 113:260401, 2014. and arXiv:1407.0338.

[35] René Schwonnek, David Reeb, and Reinhard F. Werner. Measurement uncertainty for finite quantum observables. *Mathematics*, 4(2), 2016. and arXiv:1604.00382.

[36] Supplementary material.

[37] Günther Ludwig. *Foundations of Quantum Mechanics I.* Springer Berlin Heidelberg, Berlin, Heidelberg, 1983.

[38] Teiko Heinosaari and Mário Ziman. *The mathematical language of quantum theory: from uncertainty to entanglement.* Cambridge University Press, 2011.

[39] Jian Ma, Xiaoguang Wang, Chang-Pu Sun, and Franco Nori. Quantum spin squeezing. *Phys. Rep.*, 509(2–3):89 – 165, 2011. and arXiv:1011.2978.

[40] Anders S. Sørensen, Luming Duan, Ignacio Cirac, and Peter Zoller. Many-particle entanglement with bose–einstein condensates. *Nature*, 409(6816):63–66, 2001. and arXiv:quant-ph/0006111.

[41] Anders S. Sørensen and Klaus Mølmer. Entanglement and extreme spin squeezing. *Phys. Rev. Lett.*, 86:4431–4434, May 2001. and arXiv:quant-ph/0011035.

[42] Bernd Lücke, Jan Peise, Giuseppe Vitagliano, Jan Arlt, Luis Santos, Géza Tóth, and Carsten Klempt. Detecting multiparticle entanglement of dicke states. *Phys. Rev. Lett.*, 112:155304, Apr 2014. and arXiv:1403.4542.

[43] Bernd Lücke, Manuel Scherer, Jens Kruse, Luca Pezzé, Frank Deuretzbacher, Phillip Hyllus, Oliver Topic, Jan Peise, Wolfgang Ertmer, Jan Arlt, Luis Santos, Augusto Smerzi, and Carsten Klempt. Twin matter waves for interferometry beyond the classical limit. *Science*, 2011. and arXiv:1204.4102.

[44] David Avis and Komei Fukuda. A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra. *Disc. Comput. Geom.*, 8:295–313, 1992.

# Literature to: [S18]

[1] J. Schneeloch, C. J. Broadbent, S. P. Walborn, E. G. Cavalcanti, and J. C. Howell. Einstein-Podolsky-Rosen steering inequalities from entropic uncertainty relations. *Physical Review A*, 87:062103, 2013. arXiv:1303.7432.

[2] A. C. Costa Sprotte, R. Uola, and O. Gühne. Steering criteria from general entropic uncertainty relations. 2017. arXiv:1710.04541.

[3] A. Riccardi, C. Macchiavello, and L. Maccone. Multipartite steering inequalities based on entropic uncertainty relationss. 2017. arXiv:1711.09707.

[4] Z.-A. Jia, Y.-C. Wu, and G.-C. Guo. Characterizing nonlocal correlations via universal uncertainty relations. *Phys. Rev. A*, 96:032122, 2017. arXiv:1705.08825.

[5] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner. The uncertainty principle in the presence of quantum memory. *Nature Phys.*, 2010. arXiv:0909.0950.

[6] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B Scholz, M. Tomamichel, and R. F Werner. Continuous variable quantum key distribution: Finite-key analysis of composable security against coherent attacks. *Phys. Rev. Lett.*, 109:100502, 2012. arXiv:1112.2179.

[7] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys.*, 43:172–198, 1927.

[8] E. H. Kennard. Zur Quantenmechanik einfacher Bewegungstypen. *Z. Phys.*, 44:326–352, 1927.

[9] H. P. Robertson. The uncertainty principle. *Phys. Rev.*, 34:163–164, 1929.

[10] R. Schwonnek, L. Dammeier, and R.F. Werner. State-independent uncertainty relations and entanglement detection in noisy systems. *Phys. Rev. Lett.*, 119:170404, 2017. arXiv:1705.10679.

[11] L. Dammeier, R. Schwonnek, and R.F. Werner. Uncertainty relations for angular momentum. *New J. Phys.*, 9(17):093946, 2015. arXiv:1505.00049.

[12] P. J. Coles and M. Piani. Improved entropic uncertainty relations and information exclusion relations. *Phys. Rev. A*, 89, 2014. arXiv:1307.4265.

[13] S. Wehner and A. Winter. Entropic uncertainty relations – a survey. *New J. Phys.*, 12:025009, 2010. arXiv:0907.3704.

[14] A. E. Rastegin. Rényi formulation of the entropic uncertainty principle for POVMs. *J. Phys. A*, 43:155302, 2010.

[15] Y. Xiao, C Guo, F Meng, N. Jing, and M.-H. Yung. Incompatibility of observables as state-independent bound of uncertainty relations. 2017. arXiv:1706.05650.

[16] P. Busch, P. Lahti, and R. F. Werner. Measurement uncertainty relations. *J. Math. Phys.*, 55:042111, 2014. arXiv:1312.4392.

[17] R. Schwonnek, D. Reeb, and R. F. Werner. Measurement uncertainty for finite quantum observables. *Mathematics*, 4(2):38, 2016. arXiv:1604.00382.

[18] J. M. Renes, V. B. Scholz, and S. Huber. Uncertainty relations: An operational approach to the error-disturbance tradeoff. *Quantum*, 1(20), 2016. arXiv:1612.02051.

[19] A. A. Abbott and C. Branciard. Noise and disturbance of qubit measurements: An information-theoretic characterization. *Phys. Rev. A*, 94:062110, 2016. arXiv:1607.00261.

[20] A. Barchielli, M. Gregoratti, and A. Toigo. Measurement uncertainty relations for discrete observables: Relative entropy formulation. *Comm. Math. Phys.*, (357):1253–1304, 2016. arXiv:1608.01986.

[21] H. Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Phys. Rev. Lett.*, 60:1103–1106, 1988.

[22] R. F. Werner. Quantum states with einstein-podolsky-rosen correlations admitting a hidden-variable model. *Phy. Rev. A*, 40(8):4277, 1989.

[23] S. Friedland, V. Gheorghiu, and G. Gour. Universal uncertainty relations. *Phys. Rev. Lett.*, 111(23):230401, 2013. arXiv:1304.6351.

[24] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47(10):777, 1935.

[25] E. Schrödinger. Die gegenwärtige Situation in der Quantenmechanik. *Naturwiss.*, 23(48):807–812, 1935.

[26] H.M. Wiseman, S. J. Jones, and A. C. Doherty. Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox. *Phys. Rev. Lett.*, 98(14):140402, 2007. arXiv:quant-ph/ 0612147.

[27] G. Sharma, C. Mukhopadhyay, S. Sazim, and A.K. Pati. Quantum uncertainty relation based on the mean deviation. and arXiv:1801.00994.

[28] D. Deutsch. Uncertainty in quantum measurements. *Phys. Rev. Lett.*, 50:631–633, 1983.

[29] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[30] A.N. Kolmogorov. On tables of random numbers. *Theoretical Computer Science*, 207(2):387–395, 1998.

[31] Hmolpedia. The Neumann-Shannon anecdote. `http://www.eoht.info/page/Neumann-Shannon+anecdote`.

[32] M. Tribus and E. C. Mc Irvine. Energy and information. *Sc. Am.*, 224:178–184, 1971.

[33] F. Rozpędek, J. Kaniewski, P. J. Coles, and S. Wehner. Quantum preparation uncertainty and lack of information. *New Journal of Physics*, 19(2):023038, 2016. arXiv:1606.05565.

[34] K. Abdelkhalek, R. Schwonnek, H. Maassen, F. Furrer, J. Duhme, P. Raynal, B.G. Englert, and R.F. Werner. Optimality of entropic uncertainty relations. *Int. J. Quant. Inf.*, 13(06):1550045, 2015. arXiv:1509.00398.

[35] H. Hadwiger. Minkowskische Addition und Subtraktion beliebiger Punktmengen und die Theoreme von Erhard Schmidt. *Math.Z.*, 53(3):210–218, 1950.

[36] H. F. Hofmann and S. Takeuchi. Violation of local uncertainty relations as a signature of entanglement. *Phys.Rev. A.*, 68:032103, 2003. arXiv:quant-ph/0212090.

[37] O. Gühne. *Detecting quantum entanglement: entanglement witnesses and uncertainty relations.* PhD thesis, Universität Hannover, 2004.

[38] B. Lücke, J. Peise, G. Vitagliano, J. Arlt, L. Santos, G. Tóth, and C. Klempt. Detecting multiparticle entanglement of Dicke states. *Phys. Rev. Lett.*, 112:155304, 2014. arXiv:1403.4542.

[39] O. Gühne and G. Tóth. Entanglement detection. *Phys. Rep.*, 474(1–6):1 – 75, 2009.

[40] O. Gühne and A. Costa. Private communication, 2017.

[41] I. Białynicki-Birula and J. Mycielski. Uncertainty relations for information entropy in wave mechanics. *Communications in Mathematical Physics*, 44(2):129–132, 1975.

[42] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, pages 159–182, 1975.

[43] I. I. Hirschman. A note on entropy. *American Journal of Mathematics*, 79(1):152–156, 1957.

[44] H. Maassen. The discrete entropic uncertainty relation. Talk given in Leyden University. Slides of a later version available from the author's website, 2007.

[45] H. Maassen. *Discrete entropic uncertainty relation.* Springer, 1990. 'Quantum Probability and Applications V' (Proceedings Heidelberg 1988),Lecture Notes in Mathematics 1442.

[46] A. Rényi. On measures of entropy and information. *Fourth Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, 1961.

[47] J. Hendrickx and A. Olshevsky. Matrix p-norms are NP-hard to approximate if $p \neq 1, 2, \infty$. *SIAM J. M. A. A.*, 31:2802–2812, 01 2010. arXiv:0908.1397.

[48] J. Rohn. Computing the norm $\|A\|_{\infty,1}$, is NP-hard. *Lin. and Multilin. Alg.*, 47(3):195–204, 2000.

[49] K. Drakakis and B. A. Pearlmutter. On the calculation of the $l_2 \to l_1$ induced matrix norm. *Int. J. Alg.*, 3(5):231–240, 2009.

[50] M. Riesz. Sur les maxima des formes bilinéaires et sur les fonctionnelles linéaires. *Acta Mathematica*, 49(3-4):465–497, 1926.

[51] O. G. Thorin. Convexity theorems generalizing those of M. Riesz and Hadamard with some applications. 1948.

[52] S. Golden. Lower bounds for the Helmholz function. *Phys. Rev*, 137:B1127–B1128, 1965.

[53] C. J. Thompson. inequality with applications in statistical mechanics. *J. Math. Phys.*, 6(11):1812–1813, 1965.

[54] P. J. Forrester and C. J. Thompson. The Golden-Thompson inequality: Historical aspects and random matrix applications. *J. Math. Phys.*, 55(2):023503, 2014. arXiv:1408.2008.

[55] T. Tao. The Golden-Thompson inequality| What's new?, 2010. `https://terrytao.wordpress.com/2010/07/15/the-golden-thompson-inequality/`.

[56] R. Frank and E. Lieb. Entropy and the Uncertainty Principle. *Ann. l'Ins. Henri Poincare*, 13(8):1711–1717, 2012. arXiv:1109.1209.

[57] E.H. Lieb. Convex trace functions and the Wigner-Yanase-Dyson conjecture. *Adv. in Math.*, 11:267–288, 1973.

[58] M. Berta D. Sutter and M. Tomamichel. Multivariate trace inequalitie. *Com. Mat. Phys.*, 352(1):37–58, 2016. arXiv:1604.03023.

[59] M. Lemm. On multivariate trace inequalities of Sutter, Berta and Tomamichel. *J. Mat. Phys.*, 59:012204, 2018. arXiv:1708.04836.

[60] F. Hansen. Multivariate extensions of the Golden- Thompson inequality. *An. Func. An.*, 6(4):301–310, 2015. arXiv:1406.5686.

[61] W. Grey and G. Sinnamon. Product operators on mixed norm spaces. *Lin. and Non. Lin. A.*, 2(2):189–197, 2016. arXiv:1602.0879.

[62] G. H. Hardy, J.E. Littlewood, and G. Polya. *Inequalities.* Cambridge University Press, 1934.

[63] M. A. Ballester and S. Wehner. Entropic uncertainty relations and locking: tight bounds for mutually unbiased bases. *Phys. Rev. A*, 75, 2007. arXiv:quant-ph/0606244.

[64] A. Winter. Weak locking capacity of quantum channels can be much larger than private capacity. *Journal of Cryptology*, 30(1):1–21, 2017. arXiv:1403.6361.

[65] J. Sánches-Ruiz. Optimal entropic uncertainty relation in two-dimensional Hilbert space. *Phys. Lett. A*, 244:189–195, 1998.

[66] J. Sánchez-Ruiz. Improved bounds in the entropic uncertainty and certainty relations for complementary observables. *Phys. Lett. A*, 201:125–131, 1995.

[67] A. Riccardi, C. Macchiavello, and L. Maccone. Tight entropic uncertainty relations for systems with dimension three to five. *Phys. Rev. A*, 95:032109, 2017. arXiv:1701.04304.

[68] T. Simnacher and N. Wyderka. Private communication, 2017.

# Literature to: [ASM$^+$15]

[1] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Z. Phys.*, 43:172–198, 1927.

[2] E. H. Kennard. Zur Quantenmechanik einfacher Bewegungstypen. *Z. Phys.*, 44:326–352, 1927.

[3] H. Weyl. *Gruppentheorie und Quantenmechanik*. Hirzel, Leipzig, 1928.

[4] H. P. Robertson. The uncertainty principle. *Phys. Rev.*, 34:163–164, 1929.

[5] P. Busch, P. Lahti, and R. F. Werner. Measurement uncertainty relations. *J. Math. Phys.*, 55:042111, 2014. arXiv:1312.4392.

[6] P. J. Coles and F. Furrer. State-dependent approach to entropic measurement–disturbance relations. *Phys. Lett. A*, 379:105–112, 2015. arXiv:1311.7637.

[7] J. M. Renes and V. B. Scholz. Operationally-motivated uncertainty relations for joint measurability and the error-disturbance tradeoff. 2014. arXiv:1402.6711.

[8] B.-G. Englert, D. Kaszlikowski, L. C. Kwek, and W. H. Chee. Wave-particle duality in multi-path interferometers: General concepts and three-path interferometers. *Int. J. Quant. Inf.*, 6:129–157, 2008. arXiv:0710.0179.

[9] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner. Tight finite-key analysis for quantum cryptography. *Nature Commun.*, 3:634, 2012. arXiv:1103.4130.

[10] D. Deutsch. Uncertainty in quantum measurements. *Phys. Rev. Lett.*, 50:631–633, 1983.

[11] H. Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Phys. Rev. Lett.*, 60:1103–1106, 1988.

[12] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner. The uncertainty principle in the presence of quantum memory. *Nature Phys.*, 2010. arXiv:0909.0950.

[13] I. D. Ivanovic. An inequality for the sum of entropies of unbiased quantum measurements. *J. Phys. A*, 25:L363, 1992.

[14] J. Sánchez-Ruiz. Improved bounds in the entropic uncertainty and certainty relations for complementary observables. *Phys. Lett. A*, 201:125–131, 1995.

[15] M. A. Ballester and S. Wehner. Entropic uncertainty relations and locking: tight bounds for mutually unbiased bases. *Phys. Rev. A*, 75, 2007. arXiv:quant-ph/0606244.

[16] Entropic uncertainty relation for mutually unbiased bases. 79.

[17] S. Wehner and A. Winter. Entropic uncertainty relations – a survey. *New J. Phys.*, 12:025009, 2010. arXiv:0907.3704.

[18] R. Adamczak, R. Latała, Z. Puchała, and K. Życzkowski. Asymptotic entropic uncertainty relations. 2014. arXiv:1412.7065.

[19] Z. Puchała, Ł. Rudnicki, and K. Życzkowski. Majorization entropic uncertainty relations. *J. Phys. A*, 46:272002, 2013. arXiv:1304.7755.

[20] Z. Puchała, Ł. Rudnicki, K. Chabuda, M. Paraniak, and K. Życzkowski. Certainty relations, mutual entanglement and non-displacable manifolds. 2015. arXiv:1506.07709.

[21] Ł. Rudnicki, Z. Puchała, and K. Życzkowski. Strong majorization entropic uncertainty relations. *Phys. Rev. A*, 89, 2014. arXiv:1402.0129.

[22] M. Krishna and K. R. Parthasarathy. An entropic uncertainty principle for quantum measurements. *Ind. J. Stat. A*, 64 No.3:842–852, 2001. arXiv:quant-ph/0110025.

[23] A. E. Rastegin. Rényi formulation of the entropic uncertainty principle for POVMs. *J. Phys. A*, 43:155302, 2010.

[24] P. J. Coles and M. Piani. Improved entropic uncertainty relations and information exclusion relations. *Phys. Rev. A*, 89, 2014. arXiv:1307.4265.

[25] S. Zozor, G. M. Bosyk, and M. Portesi. General entropy-like uncertainty relations in finite dimensions. *J. Phys. A*, 47:495302, 2014. arXiv:1311.5602.

[26] J. Sánches-Ruiz. Optimal entropic uncertainty relation in two-dimensional Hilbert space. *Phys. Lett. A*, 244:189–195, 1998.

[27] G. Ghirardi, L. Marinatto, and R. Romano. An optimal entropic uncertainty relation in a two-dimensional Hilbert space. *Phys. Lett. A*, 317:32–36, 2003.

[28] H. Maassen. The discrete entropic uncertainty relation. Talk given in Leyden University. Slides of a later version available from the author's website, 2007.

[29] P. J. Coles, L. Yu, and M. Zwolak. Relative entropy derivation of the uncertainty principle with quantum side information. 2011. arXiv:1105.4865.

[30] E. Phragmén and E. Lindelöf. Sur une extension d'un principe classique de l'analyse et sur quelques propriétés des fonctions monogènes dans le voisinage d'un point singulier. *Acta Math.*, 31:381–406, 1908.

[31] E. Hopf. A remark on linear elliptic differential equations of second order. *Proc. Amer. Math. Soc.*, 34(3):791–793, 1952.

[32] M. H. Protter and H. F. Weinberger. *Maximum principles in differential equations.* Springer, New York, 1984.

[33] K. Życzkowski. Complex hadamard matrices. Online catalogue, http://chaos.if.uj.edu.pl/˜karol/hadamard/, 2003.

[34] W. Rudin. *Fourier Analysis on Groups.* John Wiley & Sons, 1962.

[35] J. A. Gallian. *Contemporary abstract algebra.* Brooks / Cole, cengage learning, 2006.

# Literature to: [ACF$^+$16]

[1] D. Amati, M. Ciafaloni, and G. Veneziano. Can space-time be probed below the string size? *Physics Letters B*, 216:41, 1989.

[2] L. J. Garay. Quantum gravity and minimum length. *International Journal of Modern Physics A*, 10(02):145–165, 1995. arXiv: gr-qc/9403008.

[3] A. Kempf, G. Mangano, and R. B. Mann. Hilbert space representation of the minimal length uncertainty relation. *Physical Review D*, 52(2):1108, 1995. arXiv: hep-th/9412167.

[4] L. N. Chang, Z. Lewis, D. Minic, and T. Takeuchi. On the minimal length uncertainty relation and the foundations of string theory. *Advances in High Energy Physics*, 2011:30, 2011. arXiv: 1106.0068.

[5] S. Hossenfelder. Minimal length scale scenarios for quantum gravity. *Living Reviews in Relativity*, 16(2):90, 2013. arXiv: 1203.6191.

[6] K. Konishi, G. Paffuti, and P. Provero. Minimum physical length and the generalized uncertainty principle in string theory. *Physics Letters B*, 234:276, 1990.

[7] A. Kempf. Uncertainty relation in quantum mechanics with quantum group symmetry. *Journal of Mathematical Physics*, 35:4483–4496, 1994. arXiv: hep-th/9311147.

[8] A. Connes. *Noncommutative Geometry.* Academic Press, 1995.

[9] J. E. Moyal. Quantum mechanics as a statistical theory. *Mathematical Proceedings of the Cambridge Philosophical Society*, 45:99–124, 1949.

[10] H. J. Groenewold. On the principles of elementary quantum mechanics. *Physica*, 12:405–460, 1946.

[11] S. Doplicher, K. Fredenhagen, and J. E. Roberts. The Quantum structure of space-time at the Planck scale and quantum fields. *Communications in Mathematical Physics*, 172:187–220, 1995. arXiv: hep-th/0303037.

[12] G. Mangano, F. Lizzi, and A. Porzio. Inconstant Planck's constant. *International Journal of Modern Physics A*, 30(34):1550209, 2015. arXiv: 1509.02107.

[13] I. Pikovski, M. R. Vanner, M. Aspelmeyer, M. S. Kim, and C. Brukner. Probing Planck-scale physics with quantum optics. *Nature Physics*, 8:393–397, 2012. arXiv: 1111.1979.

[14] M. Wilde. *Quantum Information Theory.* Cambridge University Press, 2013. arXiv: 1106.1445.

[15] J. Goold, M. Huber, A. Riera, L. del Rio, and P. Skrzypczyk. The role of quantum information in thermodynamics — a topical review. 2015. arXiv: 1505.07835v2.

[16] S. M. Carroll and G. N. Remmen. What is the entropy in entropic gravity? *Physical Review D*, 93,:124052, 2016. arXiv: 1601.07558.

[17] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B. Scholz, M. Tomamichel, and R. F. Werner. Continuous variable quantum key distribution: Finite-key analysis of composable security against coherent attacks. *Physical Review Letters*, 109:100502, 2012. arXiv: 1112.2179v3.

[18] L. Dammeier, R. Schwonnek, and R. F. Werner. Uncertainty relations for angular momentum. *New Journal of Physics*, 17(9):093046, 2015. arXiv: 1505.00049.

[19] P. Busch, P. Lahti, and R. F. Werner. Measurement uncertainty relations. *Journal of Mathematical Physics*, 55:042111, 2014. arXiv: 1312.4392.

[20] K. Abdelkhalek, R. Schwonnek, H. Maassen, F. Furrer, J. Duhme, P. Raynal, B. G. Englert, and R. F. Werner. Optimality of entropic uncertainty relations. *International Journal of Quantum Information*, 13(06):1550045, 2015. arXiv: 1509.00398.

[21] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner. Entropic uncertainty relations and their applications. 2015. arXiv: 1511.04857.

[22] S. Wehner and A. Winter. Entropic uncertainty relations – a survey. *New Journal of Physics*, 12(2):025009, 2010. arXiv: 0907.3704.

[23] J. Pye, W. Donnelly, and A. Kempf. Locality and entanglement in bandlimited quantum field theory. *Physical Review D*, 92(10):105022, 2015. arXiv: 1508.05953.

[24] Altannar Chinchuluun, Panos M. Pardalos, Athanasios Migdalas, and Leonidas Pitsoulis. *Pareto Optimality, Game Theory And Equilibria*. Springer New York, 2008.

[25] R. Brout, Cl. Gabriel, M. Lubo, and Ph. Spindel. Minimal length uncertainty principle and the trans-planckian problem of black hole physics. *Physical Review D*, 59(4), jan 1999.

[26] R. F. Werner. Dilations of symmetric operators shifted by a unitary group. *Journal of Functional Analysis*, 92(1):166–176, 1990.

[27] H. P. Robertson. The uncertainty principle. *Physical Review*, 34:163–164, 1929.

[28] E. H. Kennard. Zur Quantenmechanik einfacher Bewegungstypen. *Zeitschrift für Physik*, 44:326–352, 1927.

[29] M. Bojowald and A. Kempf. Generalized uncertainty principles and localization of a particle in discrete space. *Physical Review D*, 86:085017, 2012. arXiv: 1112.0994.

[30] D. Deutsch. Uncertainty in quantum measurements. *Physical Review Letters*, 50(9):631, 1983.

[31] S. Detournay, C. Gabriel, and P. Spindel. About maximally localized states in quantum mechanics. *Physical Review D*, 66:125004, 2002. arXiv: hep-th/0210128.

[32] G. C. Dorsch and J. A. Noguera. Maximally localized states in quantum mechanics with a modified commutator relation to all orders. *International Journal of Modern Physics A*, 27(21):1250113, 2012. arXiv: 1106.2737.

[33] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[34] R. Schwonnek, D. Reeb, and R. F. Werner. Measurement uncertainty for finite quantum observables. *Mathematics*, 4(2):21, 2016. arXiv: 1604.00382.

[35] M. Reed and B. Simon. *Functional Analysis*, volume 1 of *Methods of Modern Mathematical Physics.* 1970.

[36] G. Pöschl and E. Teller. Bemerkungen zur Quantenmechanik des anharmonischen Oszillators. *Zeitschrift für Physik*, 83(3-4):143–151, 1933.

[37] Lay Nam Chang, Djordje Minic, Naotoshi Okamura, and Tatsu Takeuchi. Exact solution of the harmonic oscillator in arbitrary dimensions with minimal length uncertainty relations. *Physical Review D*, 65:125027, jun 2002. arXiv: hept-th/0111181.

[38] I. Białynicki-Birula and Ł. Rudnicki. Entropic uncertainty relations in quantum physics. In *Statistical Complexity*, pages 1–34. 2011.

[39] H. Maassen and J. Uffink. Generalized entropic uncertainty relations. *Physical Review Letters*, 60(12):1103, 1988.

[40] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[41] A. Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561, Berkeley, Calif., 1961.

[42] E. T. Jaynes. Information theory and statistical mechanics. *Brandeis University Summer Institute Lectures in Theoretical Physics, Statistical Physics*, 3:181–218, 1963.

[43] I. Białynicki-Birula and J. Mycielski. Uncertainty relations for information entropy in wave mechanics. *Communications in Mathematical Physics*, 44(2):129–132, 1975.

[44] Iwo Białynicki-Birula. Entropic uncertainty relations. *Physics Letters A*, 103(5):253–254, 1984.

[45] W. Beckner. Inequalities in Fourier analysis. *Annals of Mathematics*, pages 159–182, 1975.

[46] E. Lieb. Gaussian kernels have only Gaussian maximizers. *Inventiones Mathematicae*, 102:179–208, 1990.

[47] J. D. Bekenstein. Universal upper bound on the entropy-to-energy ratio for bounded systems. *Physical Review D*, 23(2):287, 1981.

[48] G. 't Hooft. Dimensional reduction in quantum gravity. 2009. arXiv: gr-qc/9310026.

[49] L. Susskind. The world as a hologram. *Journal of Mathematical Physics*, 36(11):6377–6396, 1995. arXiv: hep-th/9409089.

[50] J. Maldacena, S. H. Shenker, and D. Stanford. A bound on chaos. 2015. arXiv: 1503.01409.

[51] J. Anandan and Y. Aharonov. Geometry of quantum evolution. *Physical Review Letters*, 65(14):1697, 1990.

[52] Y. Aharonov and D. Bohm. Time in the quantum theory and the uncertainty relation for time and energy. *Physical Review*, 122(5):1649, 1961.

[53] N. Margolus and L. B. Levitin. The maximum speed of dynamical evolution. *Physica D: Nonlinear Phenomena*, 120(1):188–195, 1998. arXiv: quant-ph/9710043.

[54] Yakir Aharonov and Daniel Rohrlich. *Quantum paradoxes: quantum theory for the perplexed*. John Wiley & Sons, 2008.

[55] S. Lloyd. Ultimate physical limits to computation. *Nature*, 406(6799):1047–1054, 2000. arXiv: quant-ph/9908043.

[56] A. R. Brown, D. A. Roberts, L. Susskind, B. Swingle, and Y. Zhao. Complexity, action, and black holes. *Physical Review D*, 93,:086006, 2015. arXiv: 1512.04993.

[57] S. Lloyd. The quantum geometric limit. 2012. arXiv: 1206.6559.

[58] L. Freidel, R. G. Leigh, and D. Minic. Quantum spaces are modular. 2016. arXiv: 1606.01829.

[59] J. Schneeloch, C. J. Broadbent, S. P. Walborn, E. G. Cavalcanti, and J. C. Howell. Einstein-Podolsky-Rosen steering inequalities from entropic uncertainty relations. *Physical Review A*, 87:062103, 2013. arXiv: 1303.7432.

[60] D. Slepian and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty I. *Bell System Technical Journal*, 40(1):43–63, 1961.

[61] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty II. *Bell System Technical Journal*, 40(1):65–84, 1961.

[62] H. J. Landau and H. O. Pollak. Prolate spheroidal wave functions, Fourier analysis and uncertainty III. *Bell System Technical Journal*, 41(4):1295–1336, 1962.

[63] A. Kempf and R. Martin. Information theory, spectral geometry, and quantum gravity. *Physical Review Letters*, 100:021304, 2008. arXiv: 0708.0062.

[64] N. I. Akhiezer and I. M. Glazman. *Theory of linear operators in Hilbert space.* Dover Publishing Inc., 1963.

[65] K. Schmüdgen. *Unbounded self-adjoint operators on Hilbert space.* Springer Science & Business Media, 2012.

[66] B. Simon. *Functional Integration and Quantum Physics.* AMS, 1979.

[67] A. Kempf. On nonlocality, lattices and internal symmetries. *Europhysics Letters*, 40:257–261, 1997. arXiv: hep-th/9706213.

[68] A. Kempf and G. Mangano. Minimal length uncertainty relation and ultraviolet regularization. *Physical Review D*, 55(12):7909, 1997. arXiv: hep-th/9612084.

[69] Michael J. W. Hall. Universal geometric approach to uncertainty, entropy, and information. *Phys. Rev. A*, 59:2602–2615, Apr 1999. arXiv: physics/9903045.

# Appendices

## Appendix to: Measurement Uncertainty for Finite Quantum Observables

# Appendix: Optimal Transport

## A.1 Kantorovich duality

In this appendix we collect the basic theory of optimal transport adapted to the finite setting at hand. This eliminates all the topological and measure theoretic fine points that can be found, e.g., in Villani's book [21], which we also recommend for extended proofs of the statements in our summary. We slightly generalize the setting from the cost functions used in the main text of this paper: We allow the two variables on which the cost function depends to range over different sets. This might actually be useful for comparing observables, which then need not have the same outcome sets. Which outcomes are considered to be close or the same must be specified in terms of the cost function. We introduce this generalization here less for the sake of applications rather than for a simplification of the proofs, in particular for the book-keeping of paths in the proof of Lemma 2.

The basic setting is that of two finite sets $X$ and $Y$, and a arbitrary function $c : X \times Y \to \mathbb{R}$, called the *cost function*. The task is to optimize the transport of some distribution of stuff on $X$, described by a distribution function $p : X \to \mathbb{R}_+$, to a final distribution $q : Y \to \mathbb{R}_+$ on $Y$ when the transportation of one unit of stuff from the point $x$ to the point $y$ costs $c(x, y)$. In the first such scenario ever considered, namely by Gaspar Monge, the "stuff" was earth, the distribution $p$ a hill, and $q$ a fortress. Villani [21] likes to phrase the scenario in terms of bread produced at bakeries $x \in X$ to be delivered to cafés $y \in Y$. This makes plain that optimal transport is sometimes considered a branch of mathematical economics, and indeed Leonid Kantorovich, who created much of the theory, received a Nobel prize in economics. In our case the "stuff" will be probability.

A *transport plan* (or *coupling*) will be a probability distribution $\gamma : X \times Y \to \mathbb{R}_+$, which encodes how much stuff is moved from any $x$ to any $y$. Since all of $p$ is to be moved, $\sum_y \gamma(x, y) = p(x)$, and since all stuff is to be delivered, $\sum_x \gamma(x, y) = q(y)$. Now, for any transport plan $\gamma$ we get a total cost of $\sum_{x,y} \gamma(x, y)c(x, y)$, and we are interested in the optimum

$$\check{c}(p, q) = \inf_\gamma \Bigl\{ \sum_{xy} c(x, y)\gamma(x, y) \bigm| \gamma \text{ couples } p \text{ to } q \Bigr\}. \tag{36}$$

This is called the primal problem, to which there is also a dual problem. In economic language it concerns *pricing schemes*, that is, pairs of functions $\Phi : X \to \mathbb{R}$ and $\Psi : Y \to \mathbb{R}$ satisfying the inequality

$$\Phi(x) - \Psi(y) \leq c(x, y) \quad \text{for all} \;\; x \in X, \; y \in Y, \tag{37}$$

and demands to maximize

$$\hat{c}(p, q) = \sup_{\Phi,\Psi} \Bigl\{ \sum_x \Phi(x)p(x) - \sum_y \Psi(y)q(y) \bigm| (\Phi, \Psi) \text{ is a pricing scheme} \Bigr\}. \tag{38}$$

In Villani's example [21], think of a consortium of bakeries and cafés, that used to organize the transport themselves according to some plan $\gamma$. Now they are thinking of hiring a contractor, which offers to do the job, charging $\Phi(x)$ for every unit picked up from bakery $x$, and giving $\Psi(y)$ to café $y$ on delivery (these numbers can be negative). Their offer is that this will reduce overall costs, since their pricing scheme satisfies (37). Indeed, the overall charge to the consortium will be

$$\sum_x \Phi(x)p(x) - \sum_y \Psi(y)q(y) = \sum_{xy} \bigl(\Phi(x) - \Psi(y)\bigr)\gamma(x, y) \leq \sum_{xy} c(x, y)\gamma(x, y). \tag{39}$$

Taking the sup on the left hand side of this inequality (the company will try to maximize their profits by adjusting the pricing scheme $(\Phi, \Psi)$) and the inf on the right hand side (the transport plan $\gamma$ was already optimized), we get $\hat{c}(p,q) \leq \check{c}(p,q)$. It can be shown via the general duality theory of linear programming [20] that the duality gap closes in this case, i.e., we actually always have

$$\hat{c}(p,q) = \check{c}(p,q). \tag{40}$$

So the consortium will face the same transport costs in the end if the contractor chooses an optimal pricing scheme. (Note that both the infimum and the supremum in the definitions of $\check{c}$ and $\hat{c}$, respectively, are attained as $X$ and $Y$ are finite sets.)

What is especially interesting for us, however, is that the structure of the optimal solutions for both variational problems is very special, and both problems can be reduced to a combinatorial optimization over finitely many possibilities, which furthermore can be constructed independently of $p$ and $q$. Indeed, pricing schemes and transport plans are both related to certain subsets of $X \times Y$. We define $S(\gamma) \subseteq X \times Y$ as the *support* of $\gamma$, i.e., the set of pairs on which $\gamma(x,y) > 0$. For a pricing scheme $(\Phi, \Psi)$ we define the *equality set* $E(\Phi, \Psi)$ as the set of points $(x,y)$ for which equality holds in (37). Then equality holds in (39) if and only if $S(\gamma) \subset E(\Phi, \Psi)$. Note that for $\gamma$ to satisfy the marginal condition for given $p$ and $q$, its support $S(\gamma)$ cannot become too small (depending on $p$ and $q$). On the other hand, $E(\Phi, \Psi)$ cannot be too large, because the resulting system of equations for $\Phi(x)$ and $\Psi(y)$ would become overdetermined and inconsistent. The kind of set for which they meet is described in the following Definition.

**Definition 1.** *Let $X, Y$ be finite sets and $c : X \times Y \to \mathbb{R}$ a function. Then a subset $\Gamma \subset X \times Y$ is called* **cyclically $c$-monotone** *("ccm" for short), if for any sequence of distinct pairs $(x_1, y_1) \in \Gamma, \ldots, (x_n, y_n) \in \Gamma$, and any permutation $\pi$ of $\{1, \ldots, n\}$ the inequality*

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{\pi i}) \tag{41}$$

*holds. When $\Gamma$ is not properly contained in another cyclically $c$-monotone set, it is called* **maximally** *cyclically $c$-monotone ("mccm" for short).*

A basic example of a ccm set is the equality set $E(\Phi, \Psi)$ for any pricing scheme $(\Phi, \Psi)$. Indeed, for $(x_i, y_i) \in E(\Phi, \Psi)$ and any permutation $\pi$ we have

$$\sum_{i=1}^{n} c(x_i, y_i) = \sum_{i=1}^{n} \big(\Phi(x_i) - \Psi(y_i))\big) = \sum_{i=1}^{n} \big(\Phi(x_i) - \Psi(y_{\pi i})\big) \leq \sum_{i=1}^{n} c(x_i, y_{\pi i}) \tag{42}$$

The role of ccm sets in the variational problems (36) and (38) is summarized in the following proposition.

**Proposition 2.** *Let $X, Y, c, p, q$ be given as above. Then*

*(1) A coupling $\gamma$ minimizes* (36) *if and only if $S(\gamma)$ is ccm.*

*(2) The dual problem* (38) *has a maximizer $(\Phi, \Psi)$ for which $E(\Phi, \Psi)$ is mccm.*

*(3) If $\Gamma \subseteq X \times Y$ is mccm, there is a pricing scheme $(\Phi, \Psi)$ with $E(\Phi, \Psi) = \Gamma$, and $(\Phi, \Psi)$ is uniquely determined by $\Gamma$ up to the addition of the same constant to $\Phi$ and to $\Psi$.*

*Sketch of proof.*
(1) Suppose $(x_i, y_i) \in S(\gamma)$ $(i = 1, \ldots, n)$, and let $\pi$ be any permutation. Set $\delta = \min_i \gamma(x_i, y_i)$. Then we can modify $\gamma$ by subtracting $\delta$ from any $\gamma(x_i, y_i)$ and adding $\delta$ to $\gamma(x_i, y_{\pi i})$. This operation keeps $\gamma \geq 0$ and does not change the marginals. The target functional in the infimum (36) is changed by $\delta$ times the difference of the two sides of (41). For a minimizer $\gamma$ this change must be $\geq 0$, which gives inequality (41). For the converse we need a Lemma, whose proof will be sketched below.

**Lemma 2.** *For any ccm set $\Gamma$ there is some pricing scheme $(\Phi, \Psi)$ with $E(\Phi, \Psi) \supseteq \Gamma$.*

By applying this to $\Gamma = S(\gamma)$ we find that the duality gap closes for $\gamma$, i.e., equality holds in (39), and hence $\gamma$ is a minimizer.

(2) Every subset $\Gamma \subset X \times Y$ can be thought of as a bipartite graph with vertices $X \cup Y$ and an edge joining $x \in X$ and $y \in Y$ iff $(x, y) \in \Gamma$ (see Fig. 9). We call $\Gamma$ connected, if any two vertices are linked by a sequence of edges. Consider now the equality set $E(\Phi, \Psi)$ of some pricing scheme. We modify $(\Phi, \Psi)$ by picking some connected component, and setting $\Phi'(x) = \Phi(x) + a$ and $\Psi'(y) = \Psi(y) + a$ for all $x, y$ in that component. If $|a|$ is sufficiently small, $(\Phi', \Psi')$ will still satisfy all the inequalities (37), and $E(\Phi', \Psi') = E(\Phi, \Psi)$. The target functional in the optimization (38) depends linearly on $a$, so moving in the appropriate direction will increase, or at least not decrease it. We can continue until another one of the inequalities (37) becomes tight. At this point $E(\Phi', \Psi') \supsetneq E(\Phi, \Psi)$. This process can be continued until the equality set $E(\Phi, \Psi)$ is connected. Then $(\Phi, \Psi)$ is uniquely determined by $E(\Phi, \Psi)$ up to a common constant.

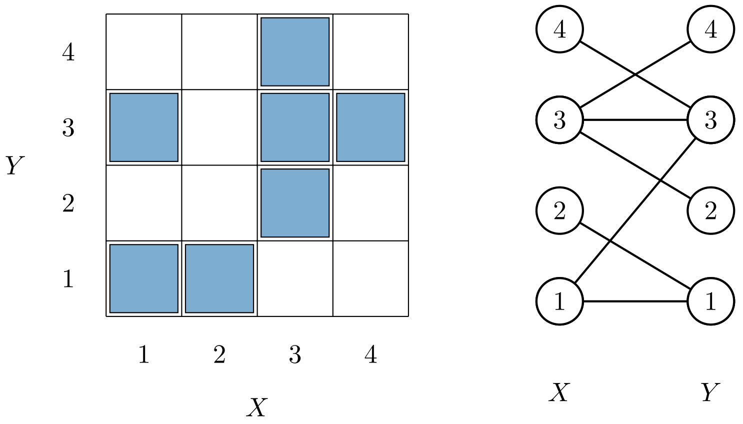


Figure 9: Representation of a subset $\Gamma \subset X \times Y$ (left) as a bipartite graph (right). The graph is a connected tree.

It remains to show that connected equality sets $E(\Phi, \Psi)$ are mccm. Suppose that $\Gamma \supseteq E(\Phi, \Psi)$ is ccm. Then by Lemma 2 we can find a pricing scheme $(\Phi', \Psi')$ with $E(\Phi', \Psi') \supseteq E(\Phi, \Psi)$. But using just the equalities in (37) coming from the connected $E(\Phi, \Psi)$, we already find that $\Phi' = \Phi + a$ and $\Psi' = \Psi + a$, so we must have $E(\Phi', \Psi') = E(\Phi, \Psi)$.

(3) is trivial from the proof of (2) that mccm sets are connected. □

*Proof sketch of Lemma 2.* Our proof will give some additional information on the set of all pricing schemes that satisfy $E(\Phi, \Psi) \supset \Gamma$ and $\Phi(x_0) = 0$ for some reference point $x_0 \in X$ to fix the otherwise arbitrary additive constant. Namely we will explicitly construct the largest element $(\Phi_+, \Psi_+)$ of this set and the smallest $(\Phi_-, \Psi_-)$, so that all other schemes $(\Phi, \Psi)$ satisfy

$$\Phi_-(x) \le \Phi(x) \le \Phi_+(x) \quad \text{and} \quad \Psi_-(y) \le \Psi(y) \le \Psi_+(y) \tag{43}$$

for all $x \in X$ and $y \in Y$. The idea is to optimize the sums of certain costs over paths in $X \cup Y$.

We define a $\Gamma$-adapted path as a sequence of vertices $z_1, \ldots, z_n \in X \cup Y$ such that the $z_i \in X \Rightarrow (z_i, z_{i+1}) \in \Gamma$, and $z_i \in Y \Rightarrow z_{i+1} \in X$. For such a path we define

$$c(z_1, \ldots, z_n) = \sum_{i=1}^{n-1} c(z_i, z_{i+1}), \tag{44}$$

with the convention $c(y,x) := -c(x,y)$ for $x \in X,\ y \in Y$. Then $\Gamma$ is ccm if and only if $c(z_1,\ldots,z_n,z_1) \leq 0$ for every $\Gamma$-adapted closed path. This is immediate for cyclic permutations, and follows for more general ones by cycle decomposition. The assertion of Lemma 2 is trivial if $\Gamma = \emptyset$, so we can pick a point $x_0 \in X$ for which some edge $(x_0, y) \in \Gamma$ exists. Then, for any $z \in X \cup Y$, we define, for $z \neq x_0$,

$$\chi_+(z) := -\sup c(x_0,\ldots,z) \quad \text{and} \quad \chi_-(z) := \sup c(z,\ldots,x_0), \tag{45}$$

where the suprema are over all $\Gamma$-adapted paths between the specified endpoints, we define $\chi_+(x_0) := \chi_-(x_0) := 0$, and empty suprema are defined as $-\infty$. Then $\chi_\pm$ are the maximal and minimal pricing schemes, when written as two functions $\Phi_\pm(x) = \chi_\pm(x)$ and $\Psi_\pm(y) = \chi_\pm(y)$ for $x \in X$ and $y \in Y$.

For proving these assertions, consider paths of the type $(x_0,\ldots,y,x)$. For this to be $\Gamma$-adapted, there is no constraint on the last link, so

$$-\chi_+(y) - c(x,y) \leq -\chi_+(x), \quad \text{and } \sup_y \bigl\{-\chi_+(y) - c(x,y)\bigr\} = \chi_+(x). \tag{46}$$

Here the inequality follows because the adapted paths $x_0 \to x$ going via $y$ as the last step are a subclass of all adapted paths and give a smaller supremum. The second statement follows, because for $x \neq x_0$ there has to be some last step from $Y$ to $x$. The inequality (46) also shows that $(\Phi_+, \Psi_+)$ is a pricing scheme. The same argument applied to the decomposition of paths $(x_0,\ldots,x,y)$ with $(x,y) \in \Gamma$ gives the inequality

$$-\chi_+(x) + c(x,y) \leq -\chi_+(y) \quad \text{for } (x,y) \in \Gamma. \tag{47}$$

Combined with inequality (46) we get that $(\Phi_+, \Psi_+)$ has equality set $E(\Phi_+, \Psi_+)$ at least $\Gamma$. The corresponding statements for $\chi_-$ follow by first considering paths $(y,x,\ldots,x_0)$ and then $(x,y\ldots,x_0)$ with $(x,y) \in \Gamma$.

Finally, in order to show the inequalities (43), let $(\Phi, \Psi)$ be a tight pricing scheme with $\Phi(x_0) = 0$ and $E(\Phi, \Psi) \supset \Gamma$. Consider first any $\Gamma$-adapted path $(x_0, y_0, x_1, \ldots, x_n, y)$. Then,

$$\begin{aligned} c(x_0,\ldots,x_n,y) &= \sum_{i=0}^{n-1} \bigl(\Phi(x_i) - \Psi(y_i) - c(x_{i+1}, y_i)\bigr) + \Phi(x_n) - \Psi(y) \\ &= \Phi(x_0) - \Psi(y) + \sum_{i=0}^{n-1} \bigl(\Phi(x_{i+1}) - \Psi(y_i) - c(x_{i+1}, y_i)\bigr) \\ &\leq \Phi(x_0) - \Psi(y) = -\Psi(y), \end{aligned} \tag{48}$$

because the sum is termwise non-positive due to the pricing scheme property. Hence by taking the supremum we get $\chi_+(y) \geq \Psi(y)$. The other inequalities follow with the same arguments applied to paths of the type $(x_0,\ldots,y_n,x)$, $(x,y_0,\ldots,x_0)$, and $(y,x_1,\ldots,x_0)$. □

Let us summarize the consequences of Proposition 2 for the computation of minimal costs (36). Given any cost function $c$, the first step is to enumerate the corresponding mccm sets, say $\Gamma_\alpha$, $\alpha \in \mathcal{S}$, for some *finite* label set $\mathcal{S}$, and to compute for each of these the pricing scheme $(\Phi_\alpha, \Psi_\alpha)$ (up to an overall additive constant, see Proposition 2). This step depends only on the chosen cost function $c$. Then, for any distributions $p, q$ we get

$$\hat{c}(p,q) = \check{c}(p,q) = \max_{\alpha \in \mathcal{S}} \sum_x \Phi_\alpha(x) p(x) - \sum_y \Psi_\alpha(y) q(y). \tag{49}$$

This is very fast to compute, so the preparatory work of determining the $(\Phi_\alpha, \Psi_\alpha)$ is well invested if many such expressions have to be computed. However, even more important for us that (49) simplifies the variational problem sufficiently so that we can combine it with the optimization over joint measurements (see Sect. 4.1). Of course, this leaves open the question of how to determine all mccm sets for a cost function. Some remarks about this will be collected in the next subsection.

## A.2 How to find all mccm sets

We will begin with a basic algorithm for the general finite setting, in which $X, Y$, and the cost function $c$ are arbitrary. Often the task can be greatly simplified if more structure is given. These simplifications will be described in the following sections.

The basic algorithm will be a growth process for ccm subsets $\Gamma \subseteq X \times Y$, which stops as soon as $\Gamma$ is connected (cf. the proof of Proposition 2(2)). After that, we can compute the unique pricing scheme $(\Phi, \Psi)$ with equality on $\Gamma$ by solving the system of linear equations with $(x, y) \in \Gamma$ from (37). This scheme may have additional equality pairs extending $\Gamma$ to an mccm set. Hence, the same $(\Phi, \Psi)$ and mccm sets may arise from another route of the growth process. Nevertheless, we can stop the growth when $\Gamma$ is connected, and eliminate doubles as a last step of the algorithm. The main part of the algorithm will thus aim at finding all *connected ccm trees*, where by definition a tree is a graph containing no cycles. We take each tree to be given by a list of edges $(x_1, y_1), \ldots (x_N, y_N)$, which we take to be written in lexicographic ordering, relative to some arbitrary numberings $X = \{1, \ldots, |X|\}$ and $Y = \{1, \ldots, |Y|\}$. Hence the first element in the list will be $(1, y)$, where $y$ is the first element connected to $1 \in X$.

At stage $k$ of the algorithm we will have a list of all possible initial sequences $(x_1, y_1), \ldots (x_k, y_k)$ of lexicographically ordered ccm trees. For each such sequence the possible next elements will be determined, and all the resulting edge-lists of length $k + 1$ form the next stage of the algorithm. Now suppose we have some list $(x_1, y_1), \ldots (x_k, y_k)$. What can the next pair $(x', y')$ be? There are two possibilities:

(a) $x' = x_k$ is unchanged. Then lexicographic ordering dictates that $y' > y_k$. Suppose that $y'$ is already connected to some $x < x_k$. Then adding the edge $(x_k, y')$ would imply that $y'$ could be reached in two different ways from the starting node ($x = 1$). Since we are looking only for trees, we must therefore restrict to only those $y' > y_k$ which are yet unconnected.

(b) $x$ is incremented. Since in the end all vertices $x$ must lie in one connected component, the next one has to be $x' = x_k + 1$. Since the graphs at any stage should be connected, $y'$ must be a previously connected $Y$-vertex.

With each new addition we also check the ccm property of the resulting graph. The best way to do this is to store with any graph the functions $\Phi, \Psi$ on the set of already connected nodes (starting from $\Phi(1) = 0$), and update them with any growth step. We then only have to verify inequality (37) for every new node paired with every old one. Since the equality set of any pricing scheme is ccm, this is sufficient. The algorithm will stop as soon as all nodes are included, i.e., after $|X| + |Y| - 1$ steps.

## A.3 The linearly ordered case

When we look at standard quantum observables, given by a Hermitian operator $A$, the outcomes are understood to be the eigenvalues of $A$, i.e., real numbers. Moreover, we typically look at cost functions which depend on the difference $(x - y)$ of two eigenvalues, i.e.,

$$c(x, y) = h\big(x - y\big). \tag{50}$$

For the Wasserstein distances one uses $h(t) = |t|^\alpha$ with $\alpha \geq 1$. The following Lemma allows, in addition, arbitrary convex, not necessarily even functions $h$.

**Lemma 3.** *Let $h : \mathbb{R} \to \mathbb{R}$ be convex, and $c$ be given by* (50)*. Then for $x_1 \leq x_2$ and $y_1 \leq y_2$ we have*

$$c(x_1, y_1) + c(x_2, y_2) \leq c(x_1, y_2) + c(x_2, y_1), \tag{51}$$

*with strict inequality if $h$ is strictly convex, $x_1 < x_2$ and $y_1 < y_2$.*

*Proof.* Since $x_2 - x_1 \geq 0$ and $y_2 - y_1 \geq 0$, there exists $\lambda \in [0,1]$ such that $(1-\lambda)(x_2-x_1) = \lambda(y_2-y_1)$. This implies $x_1 - y_1 = \lambda(x_1-y_2)+(1-\lambda)(x_2-y_1)$, so that convexity of $h$ gives $c(x_1,y_1) = h(x_1-y_1) \leq \lambda h(x_1 - y_2) + (1-\lambda)h(x_2 - y_1) = \lambda c(x_1,y_2) + (1-\lambda)c(x_2,y_1)$. The same choice of $\lambda$ also implies $x_2 - y_2 = (1-\lambda)(x_1 - y_2) + \lambda(x_2 - y_1)$, so that similarly $c(x_2,y_2) \leq (1-\lambda)c(x_1,y_2) + \lambda c(x_2,y_1)$. Adding up the two inequalities yields the desired result. If $x_1 < x_2$ and $y_1 < y_2$ are strict inequalities, then $\lambda \in (0,1)$, so that strict convexity of $h$ gives a strict overall inequality. □

As a consequence, if $\Gamma$ is a ccm set for the cost function $c$ and $(x_1,y_1) \in \Gamma$, then all $(x,y) \in \Gamma$ satisfy either $x \leq x_1$ and $y \leq y_1$ or $x \geq x_1$ and $y \geq y_1$. Loosely speaking, while in $\Gamma$, one can only move north-east or south-west, but never north-west or south-east.

This has immediate consequences for ccm sets: In each step in the lexicographically ordered list (see the algorithm in the previous subsection) one either has to increase $x$ by one or increase $y$ by one, going from $(1,1)$ to the maximum. This is a simple drive on the Manhattan grid, and is parameterized by the instructions on whether to go north or east in every step. Of the $|X|+|Y|-2$ necessary steps, $|X|-1$ have to go in the east direction, so altogether we will have at most

$$r = \binom{|X|+|Y|-2}{|X|-1} \tag{52}$$

mccm sets and pricing schemes. They are quickly enumerated without going through the full tree search described in the previous subsection.

## A.4 The metric case

Another case in which a little bit more can be said is the following [21, Case 5.4, p.56]:

**Lemma 4.** *Let $X = Y$, and consider a cost function $c(x,y)$ which is a metric on $X$. Then:*
*(1) Optimal pricing schemes satisfy $\Phi = \Psi$, and the Lipshitz condition $|\Phi(x) - \Phi(y)| \leq c(x,y)$.*
*(2) All mccm sets contain the diagonal.*

*Proof.* Any pricing schemes satisfies $\Phi(x) - \Psi(x) \leq c(x,x) = 0$, i.e., $\Phi(x) \leq \Psi(x)$. For an optimal scheme, and $y \in X$, we can find $x'$ such that $\Psi(y) = \Phi(x') - c(x',y)$. Hence

$$\Psi(y) - \Psi(x) \leq \big(\Phi(x') - c(x',y)\big) + \big(c(x',x) - \Phi(x')\big) \leq c(y,x). \tag{53}$$

By exchanging $x$ and $y$ we get $|\Psi(y) - \Psi(x)| \leq c(y,x)$. Moreover, given $x$, some $y$ will satisfy

$$\Phi(x) = \Psi(y) + c(x,y) \geq \Psi(x), \tag{54}$$

which combined with the previous first inequality gives $\Phi = \Psi$. In particular, every $(x,x)$ belongs to the equality set. □

One even more special case is that of the discrete metric, $c(x,y) = 1 - \delta_{xy}$. In this case it makes no sense to look at error exponents, because $c(x,y)^\alpha = c(x,y)$. Moreover, the Lipshitz condition $|\Phi(x) - \Phi(y)| \leq c(x,y)$ is vacuous for $x = y$, and otherwise only asserts that $\Phi(x) - \Phi(y) \leq 1$, which after adjustment of a constant just means that $|\Phi(x)| \leq 1/2$ for all $x$. Hence the transportation cost is just the $\ell^1$ norm up to a factor, i.e.,

$$\check{c}(p,q) = \frac{1}{2} \sup_{|\Phi| \leq 1} \sum_x (p(x) - q(x))\Phi(x) = \frac{1}{2} \sum_x |p(x) - q(x)|. \tag{55}$$

# Appendix to: State-independent uncertainty relations and entanglement detection in noisy systems

## Appendix

### Strict monotonicity of the gap

We will consider measurements $A$ and $B$ which could also be represented by two general POVMs. Furthermore, we will assume that we already have an initial outer approximation of the corresponding set $\mathcal{C}$ by a polyhedron $\mathcal{P}(\mathcal{R}_0)$, constructed from initial directions $\mathcal{R}_0$. Such a set of directions can be constructed by taking the face normals of a cube, as in Fig. 3.

Let $\mathbf{v}^*$ be a vertex of $\mathcal{E}(\mathcal{R})$ on which the minimum of $\mu$ is attained, i.e. $c_-(\mathcal{R}) = \mu(v^*)$ and take $\mathbf{r}' = \nabla\mu|_{\mathbf{v}^*}$ as new direction such as $\mathcal{R}' := \mathcal{R} \cup \mathbf{r}'$ as new set of directions, in every step. Then the bound $c_-(\mathcal{R})$ will either increase after a finite round of such steps or attain a global minimum on $\mathcal{C}$.

*Proof.* We will show that, by taking $\mathbf{r}'$ as above, the point $\mathbf{v}^*$ will be removed from the resulting polyhedron $\mathbb{P}(\mathcal{R}')$ whenever $\mathbf{v}^*$ is not in $\mathcal{C}$. From this statement we can conclude that: If $\mathbf{v}^*$ is removed, and there is no point in $\mathcal{C}$ which attains the value $c_-(\mathcal{R}')$, the new bound $c_-(\mathcal{R}')$ will fail to increase if and only if there is another extremal point in $\mathcal{E}(\mathcal{R}')$ that also attains the same minimal value of $\mu$. However, because $\mathcal{E}(\mathcal{R}')$ is a finite set, all those points will be removed from it after a finite round of steps. Hence, the bound $c_-$ will increase after a finite round of steps.

In the alternative case, when $\mathbf{v}^*$ is in $\mathcal{C}$, the upper bound $c_+$ will also be attained on $\mathbf{v}^*$, such that $c_+ = c_- = c$. Therefore, we already would have found the optimal bound.

Consider a fixed $\mathbf{v}^*$ and the level-set $M_{\mathbf{v}*} = \{\mathbf{x} \in \mathbb{R}^3 | \mu(\mathbf{x}) \geq \mu(\mathbf{v}^*)\}$. The set $M_{\mathbf{v}*}$ is convex and obviously contains $\mathcal{C}$ and $\mathcal{P}(\mathcal{R})$ as subsets. $\mu$ is a quadratic functional, so its gradient is well defined everywhere. Moreover, the direction $\mathbf{r}' = \nabla\mu|_{\mathbf{v}^*}$ is the normal direction of the tangent space of $M_{\mathbf{v}^*}$ at the point $\mathbf{v}^*$. Due to convexity, $M_{\mathbf{v}^*}$ is described by linear inequalities corresponding to its tangent spaces, which implies that

$$\mathbf{r}'.\mathbf{x} \geq \mathbf{r}'.\mathbf{v}^* \tag{14}$$

for all points $\mathbf{x}$ from $M_{\mathbf{v}^*}$, and so, for all $\mathbf{x}$ from $\mathcal{C}$ and $\mathcal{P}(\mathcal{R})$, as well. More precisely, we have $\mathbf{r}'.\mathbf{v}^* = \min\{\mathbf{r}'.\mathbf{x} | \mathbf{x} \in M_{\mathbf{v}^*}\}$, with a minimum that is attained uniquely on $\mathbf{v}^*$, because $\mu$ is strictly concave. If we now consider the new set of direction $\mathcal{R}'$ and its corresponding inequalities for constructing $\mathcal{P}(\mathcal{R}')$, see (5), we have

$$\mathbf{r}'.\mathbf{x} \geq h(\mathbf{r}') = \min\{\mathbf{r}'.\mathbf{x} | \mathbf{x} \in \mathcal{C}\} \geq \mathbf{r}'.\mathbf{v}^* \tag{15}$$

for all $\mathbf{x} \in \mathcal{P}(\mathcal{R}')$. Hereby, equality in the last part of (15) holds if and only if $\mathbf{v}^* \in \mathcal{C}$. In all other cases the functional $\mathbf{r}'.\mathbf{x}$ separates the set $\mathcal{P}(\mathcal{R}')$ from the point $\mathbf{v}^*$. □

### Precision per Step

A crucial property of any numerical method is its performance. For the method provided in this work we measure it by the numerical precision in comparison to the number of steps required. As a benchmarking we computed several random examples and illustrated three of them in Fig. 5. We observe that in the typical case, there are two different kinds of scaling behaviour. During the first part of steps, the precision increases slower than in the second. In the regime of the first steps the algorithm improves the outer approximation at very different points. However, once the outer polyherdron is fine enough, the algorithm generates vertices close to the actual optimum. If this optimum is unique, the algorithm will go to a regime where all improvements are made locally. In Fig. 5 this transition from global to local optimization is marked. After this point the improvement of precision per step, measured in decimal places of the gap $\epsilon$, scales linear.

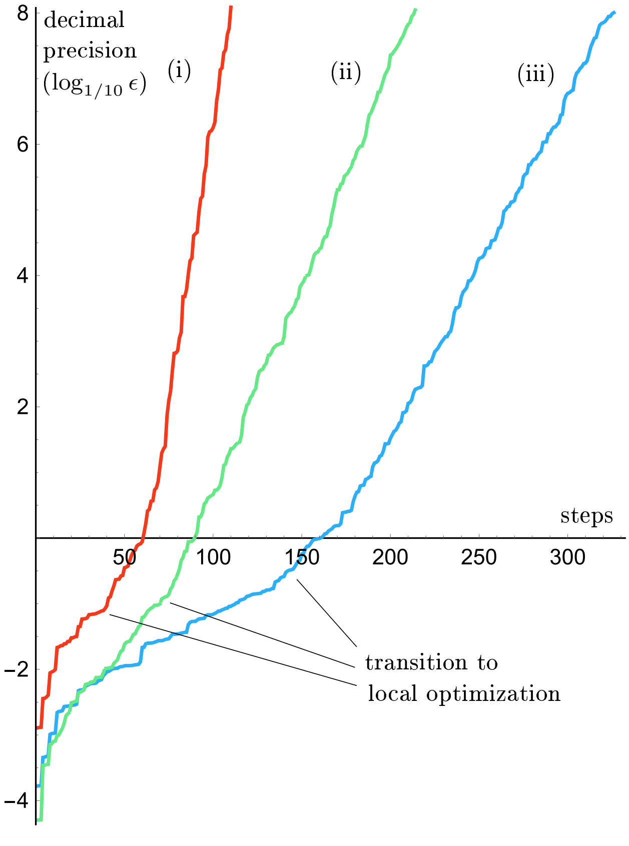


FIG. 5. Typical scaling behaviour of the gap $\epsilon$ in dependence of the steps made by the algorithm. Here depicted for three different randomly chosen pairs of operators, see Tab. I. In the above examples, the optimal uncertainty bound is attained at a unique point on $\mathcal{C}$, hence we observe a localization of our algorithm around this point. When this happens the scaling of the precision $\epsilon$, measured in decimal places, becomes linear, i.e. $\epsilon \approx 10^{-\lambda \#\text{steps}}$ .

| sample | method | size | steps |
|---|---|---|---|
| (i) | Haar random | $30 \times 30$ | 110 |
| (ii) | Haar random eigenvectors<br>uniform dist. spectrum | $200 \times 200$ | 214 |
| (iii) | Haar random | $200 \times 200$ | 326 |

TABLE I. Parameters of different random examples, see also Fig.-5, in order to benchmark the performance of the algorithm.

This behaviour agrees with the worst case example given in Fig. 6, where we considered two orthogonal angular momentum components, $L_z$ and $L_x$. Here rotations around the $y$-axis, in terms of spin components, impose a rotational degree of freedom on the set of all linear combinations of the operators $L_z$, $L_x$ and $L_x^2 + L_z^2 = s(s+1) - L_y^2$. This results in a region $\mathcal{C}$ that is rotational symmetric, as well. As the variance sum itself shares the same symmetry the optimum on $\mathcal{C}$ will be attained on a continuum of points.

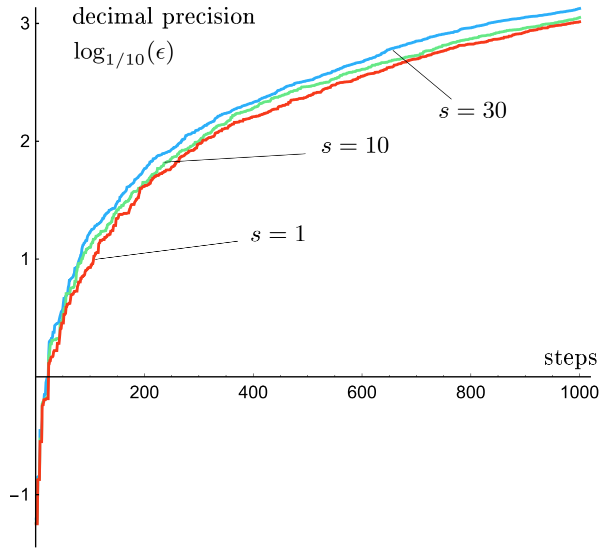


FIG. 6. Scaling behaviour of the precision $\epsilon$ for orthogonal angular momentum components. Depicted for different spins s: red=1, green =10, blue= 30. In these examples the optimal value of the uncertainty bound is not attained on a finite set of points, hence we have no localization of the algorithm. This benchmarks the worst case with respect to scaling. In the above example we rescaled the spectra of the operators to the unit interval. From the figure above can been seen that the algorithm shows the same scaling behaviour in all three cases. This illustrates that, the amount of steps the algorithm takes for reaching a certain target precision, is independent of the underlying Hilbert space dimension.

This is a *worst case* scenario, because our method has to improve the outer approximation on a continuum of points. Hence, no localization of the algorithm can be expected, and no transition in a linear scaling regime happens, for this compare Fig. 5 with Fig. 6. Note that, in this highly symetrical case there is no need to performe the algorithm on the whole set $\mathcal{C}$. If we take care of the ounderlying symmetry the problem reduces to a fast scaling problem on a two dimensioal subset of $\mathcal{C}$ again.

## Examples

### Non-orthogonal spin components:

We computed the minimal uncertainty for a measurement of two spin-s components that span an angle $\phi$. Without loss of generality we can assume one of the components to be given by $L_z$ and the other one to lie in the $L_z - L_x$ plane. So we can take

$$L_\phi = \cos\phi L_z + \sin\phi L_x. \tag{16}$$

The value of the uncertainty bound in dependence of the angle $\phi$ is shown in Fig. 7 and Tab. II.

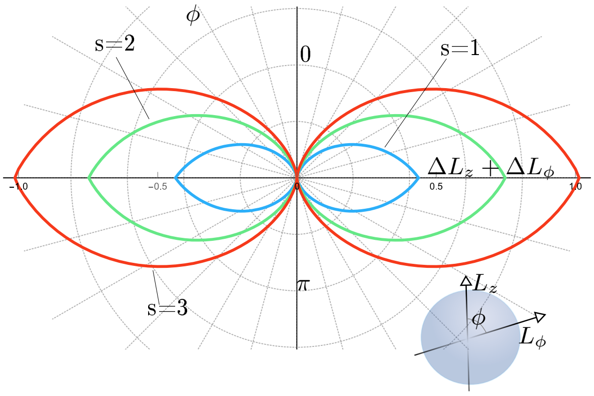


FIG. 7. Polar plot of the uncertainty bound between the non-orthogonal spin components $L_z$ and $L_\phi$ as a function of the angle $\phi$, depicted for spins $s = 1, 2, 3$

| spin/angle | 0 | $\frac{\pi}{8}$ | $\frac{\pi}{4}$ | $\frac{3\pi}{8}$ | $\frac{\pi}{2}$ |
|---|---|---|---|---|---|
| 1 | 0 | 0.0378 | 0.1431 | 0.2910 | 0.4365 |
| 2 | 0 | 0.0743 | 0.2754 | 0.5318 | 0.7478 |
| 3 | 0 | 0.1108 | 0.3984 | 0.7444 | 1.0131 |

TABLE II. Numerical values for the uncertainty of non-orthogonal spin components $L_z$ and $L_\phi$. Due to periodicity only angles from the interval $[0, \pi/2]$ are relevant.

### Entanglement detection with local noise

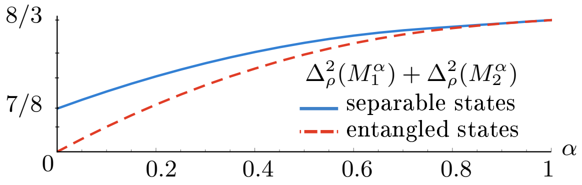


FIG. 8. Linear uncertainty bounds with equal weights in dependence of local noise, evaluated for measurements $M_1$ and $M_2$ on separable and entangled states. Any state that yields a variance sum below the blue solid line is entangled.

# Appendix to: Optimal uncertainty relations in a modified Heisenberg algebra

# APPENDIX

## APPENDIX A: PROOF OF THEOREM 1

Starting point of our examination is the Hilbert space $\mathcal{H} := \mathcal{L}^2(\mathbb{R}, \mathrm{d}x)$, witht the usual Borel-Lebesgue measure, together with the standard position operator $\mathbf{x}$ with dense domain $\mathcal{D}(\mathbf{x}) = \{\psi \in \mathcal{H} \,|\, \|\mathbf{x}\psi\|_2 < \infty\}$. Recall that $\mathbf{x}$ is self-adjoint on $\mathcal{D}(\mathbf{x})$ and has continuous spectrum $\sigma(\mathbf{x}) = \mathbb{R}$. The standard momentum operator $\mathbf{k}$ on $\mathcal{H}$ with its domain $\mathcal{D}(\mathbf{k})$ of weakly differentiable functions in $\mathcal{H}$ is related to $\mathbf{x}$ by the Fourier-transform on $\mathcal{H}$ and it is self-adjoint. They fulfill the standard Heisenberg commutation relation $[\mathbf{x}, \mathbf{k}]\psi = i\psi$ for $\psi$ in a dense domain of analyic vectors for both operators.

Our aim is to examine the situation where there is another linear and self-adjoint operator $\mathbf{p}$ on $\mathcal{H}$, possibly unbounded and with domain $\mathcal{D}(\mathbf{p})$ such that there exists a sufficiently well-behaved function $f : \mathbb{R} \to \mathbb{R}$ such that

$$[\mathbf{x}, \mathbf{p}]\psi = if(\mathbf{p})\psi \tag{A1}$$

for some sensible choice of $\psi \in \mathcal{H}$.

More precisely we even expect the operator $\mathbf{p}$ having domain $\mathcal{D}(\mathbf{p})$ just on a sub-Hilbert space $\mathcal{P} \subset \mathcal{H}$ and such that it is self-adjoint only in $\mathcal{P}$ and the commutation relation (A1) only holds true for suitable vectors $\psi \in \mathcal{P}$. Even more, we want to show that under suitable assumptions on the function $f$ we can find a function $p : I \subseteq \mathbb{R} \to \mathbb{R}$ such that $\mathbf{p} = p(\mathbf{k})$ and $I$ is a possibly unbounded interval. The Hilbert space $\mathcal{P}$ can then be identified with $\mathcal{L}^2(I)$.

It is important here that we consider an embedding of $\mathcal{P}$ into the larger space $\mathcal{H} = \mathcal{L}^2(\mathbb{R})$ since this allows us later to interpret the operator $\mathbf{p}$ as a modified version of the momentum operator.

### 1. An example

We start with an example from which we extract the essential structure we use later in our assertion. For this consider, as above, the Hilbert space $\mathcal{H} = \mathcal{L}^2(\mathbb{R}, \mathrm{d}x)$, and standard position and momentum operators $\mathbf{x}$ and $\mathbf{k}$ satisfying the standard canonical commutation relations $\mathbf{x}\mathbf{k}\psi - \mathbf{k}\mathbf{x}\psi = i\psi$ on a common dense set $\mathcal{D}$ of analytic vectors $\psi$. Denote by $E_{\mathbf{k}}$ the spectral measure of $\mathbf{k}$. Note that both operators, $\mathbf{k}$ and $\mathbf{x}$, have simple spectrum $\sigma(\mathbf{k}) = \mathbb{R} = \sigma(\mathbf{x})$. That is, they are unitarily equivalent to a position operator on some $\mathcal{L}^2(\mathbb{R}, \mu)$, where $\mu$ is a suitable measure [64, 65]. We consider the representation, where $\mathbf{k}$ (the momentum operator) acts on $\mathcal{H}$ as the position operator, i.e. $\forall \psi \in \mathcal{D}(\mathbf{k}) : (\mathbf{k}\psi)(k) = k\psi(k)$, and where $\mathcal{D}(\mathbf{k})$ is the domain of $\mathbf{k}$.

Now consider the real-valued function $p$ on $\mathbb{R}$ given by $p(k) = \tan \circ \chi_I(k), k \in \mathbb{R}$, where $\chi_I$ is the characteristic function on the interval $I = [-\frac{\pi}{2}, \frac{\pi}{2}]$. From this we get a self-adjoint operator

$$\mathbf{p} := p(\mathbf{k}) = \int_{\mathbb{R}} p(\lambda) E_{\mathbf{k}}(\mathrm{d}\lambda).$$

with domain $\tilde{\mathcal{D}}(\mathbf{p}) = \{\psi \in \mathcal{H} \,|\, \int_{\mathbb{R}} |p(\lambda)|^2 \mathrm{d}\langle\psi, E_{\mathbf{k}}(\lambda)\psi\rangle < \infty\}$ and $\mathrm{d}\langle\psi, E_{\mathbf{k}}(\lambda)\psi\rangle$ is the unique measure (on the Borel-$\sigma$-algebra $\mathcal{B}(\mathbb{R})$) given by $\langle\psi, E_{\mathbf{k}}(\cdot)\psi\rangle$.

Let $\mathcal{P} := \mathcal{L}^2(I, \mathrm{d}x) \subset \mathcal{H}$ and denote by $P$ the projection $\mathcal{H} \to \mathcal{P}$. Then it is easy to check that $\mathbf{p} = p(P\mathbf{k}P)$, and, when regarded as an operator, $\mathbf{p}$ is self-adjoint on the domain $\mathcal{D}(\mathbf{p}) = \tilde{\mathcal{D}}(\mathbf{p}) \cap \mathcal{P}$. Note that the operator $P\mathbf{k}P$ is bounded, since the function $k \mapsto k$ is bounded on $I$.

Denote the operator $P\mathbf{k}P$, when regarded as operator on the Hilbert space $\mathcal{P}$, by $\mathbf{k}_P$. Then $\mathbf{k}_P$ has simple spectrum $\sigma(\mathbf{k}_P) = \mathbb{R}$, and the self-adjoint, unbounded operator $\mathbf{p}$, as operator on $\mathcal{P}$ has spectrum $\sigma(\mathbb{R})$. Furthermore, as such, $\mathbf{p} = p(\mathbf{k}_P)$, and since the function tan is bijective on $I$, $\sigma(\mathbf{p})$ is simple. Note that, when regarded as operator on $\mathcal{H}$ the spectrum is not simple anymore, since zero is then an infinitely degenerate eigenvalue of $\mathbf{p}$. In any case the spectral projections of $\mathbf{p}$ commute with the spectral projections of $\mathbf{k}$ and $\mathbf{k}_P$, i.e. the operators $\mathbf{k}_P$ and $\mathbf{p}$ strongly commute.

Since the tangent is analytic the dense subspace of analytic vectors $\mathcal{D}$ for $\mathbf{p}$ in $\mathcal{P}$ is contained in the space of analytic vectors for $\mathbf{x}$. For example, the smooth, compactly supported functions in $I$ that exponentially decay at the boundary of $I$ are analytic for $\mathbf{p}$, and also for $\mathbf{x}$. Moreover, for $\psi \in \mathcal{D}$ we have that $\mathbf{x}\mathbf{p}\psi - \mathbf{p}\mathbf{x}\psi = i(1 + \mathbf{p}^2)\psi$.

Now let $U : \mathbb{R} \times \mathcal{H} \to \mathcal{H}$ be the unitary 1-parameter group of translations generated by $\mathbf{x}$. Then for $\Omega \in \mathcal{B}(\mathcal{R})$ and $t \in \mathbb{R}$ we have that $U_t E_{\mathbf{k}}(\Omega) U_{-t} = E_{\mathbf{k}}(\Omega + t)$, where $\Omega + t = \{x + t \,|\, x \in \Omega\}$. If $g : \mathbb{R} \to \mathbb{R}$ is an analytic function we get $(U_t E_{\mathbf{k}}(\Omega) U_{-t}\psi)(k) = \chi_\Omega(g(k-t))\psi(k) = \chi_{\Omega_g}(k)\psi(k)$, where we set $\Omega_g := \{k \in \mathbb{R} \,|\, g(k-t) \in \Omega\} = g^{-1}(\Omega) + t$. Hence $U_t E_{\mathbf{p}}(\Omega) U_{-t} = E_{\mathbf{k}}(\Omega_p) = E_{\mathbf{p}}(\Omega') + Q$ with $\Omega' := p(\Omega_p) \cap [-\frac{\pi}{2}, \frac{\pi}{2}]$ and $Q = E_{\mathbf{k}}(\Omega_p \setminus [-\frac{\pi}{2}, \frac{\pi}{2}])$. This can also be used to see that for all $\epsilon \leq \frac{\pi}{2}$ there exists an $\epsilon' > 0$ such that $E_{\mathbf{k}}([-\epsilon, \epsilon]) = E_{\mathbf{p}}([-\epsilon', \epsilon'])$ by simply choosing $\epsilon' = \arctan \epsilon$.

### 2. The general case

**Assumptions.** *Given $\mathcal{H}$, $\mathbf{x}$, $\mathbf{k}$ as before. Let $f : \mathbb{R} \to \mathbb{R}$ be a function with the following properties:*

- *$f$ is smooth.*
- *$f(0) = 1$.*
- *$f$ is symmetric, i.e. $\forall p \in \mathbb{R} : f(-p) = f(p)$.*
- *$f$ is convex on $\mathbb{R}^+$.*

*Furthermore there exists a closed subspace $\mathcal{P} \subseteq \mathcal{H}$ with projection $P : \mathcal{H} \to \mathcal{P}$ such that $(\mathbf{p}, \mathcal{D}(\mathbf{p}))$ satisfies:*

- *$\mathcal{D}(\mathbf{p}) \subset \mathcal{P}$*

- $(\mathbf{p}, \mathcal{D}(\mathbf{p}))$ *is a self-adoint, linear operator on* $\mathcal{P}$
- *The spectrum* $\sigma(\mathbf{p})$ *is continuous, coincides with* $\mathbb{R}$, *and is simple. I.e. there exists a vector* $\psi \in \mathcal{P}$ *such that for any other vector* $\phi \in \mathcal{P}$ *there exists a function* $f \in \mathbb{L}^2(\sigma(\mathbf{p}), \mu)$ *such that* $\phi = \int_{\mathbb{R}} f(t) \mathrm{d}E_{\mathbf{p}}(t)\psi$ *and* $\mu$ *is the measure given by* $\mu(\Omega) = (E_{\mathbf{p}}(\Omega)\psi, \psi)$.
- *There exists a dense subspace* $\mathcal{D} \subset \mathcal{P}$ *such that:* $\forall \psi \in \mathcal{D} : \mathbf{x}\mathbf{p}\psi - \mathbf{p}\mathbf{x}\psi = if(\mathbf{p})\psi$.
- *For all* $\psi \in \mathcal{D}$ *and for all* $n \in \mathbb{N}$ *it holds that* $\mathbf{x}^n\psi \in \mathcal{D}$ *and* $\mathbf{p}^n\psi \in \mathcal{D}$. *In other words,* $\mathcal{D}$ *is a dense set of analytic vectors in* $\mathcal{P}$ *for both,* $\mathbf{x}$ *and* $\mathbf{p}$.

*We say that the objects* $(f, \mathbf{p}, \mathcal{D}(\mathbf{p}), \mathcal{P})$ *are* admissible *if they satisfy all of the above assumptions.*

The existence of dense subsets of analytic vectors for $\mathbf{x}$, respectively $\mathbf{p}$ follows from the simplicity of their spectra [64, Theorem 69.3]. We want, however, that there exists a *common* set of analytic vectors for both operators, which is contained just in $\mathcal{P}$.

By [64, Theorem 69.2] (or [65, Proposition 5.18]) the assumption that $\mathbf{p}$ has simple spectrum implies that $\mathcal{P} \cong \mathcal{L}^2(\mathbb{R}, \mu)$, that is, vectors $\psi \in \mathcal{P}$ correspond to (equivalence classes of) functions $\psi \in \mathcal{L}^2(\mathbb{R}, \mu)$. Furthermore, by abuse of notation, $\forall \psi \in \mathcal{D}(\mathbf{p}) : (\mathbf{p}\psi)(p) = p\psi(p)$.

**Definition 1** ([35]). *Let* $A$ *and* $B$ *be unbounded self-adjoint operators on some Hilbert space* $\mathcal{H}$ *and* $E_A$ *and* $E_B$ *their projection-valued measures over* $\mathbb{R}$. *We say that* $A$ *and* $B$ strongly commute, *if for all measurable sets* $\Omega, \Omega' \subset \mathbb{R}$ *the according spectral projections commute, i.e.* $E_A(\Omega)E_B(\Omega') = E_B(\Omega')E_A(\Omega)$.

**Theorem 4.** *Given the standard position and momentum operators* $\mathbf{x}$ *and* $\mathbf{k}$ *on* $\mathcal{H} = \mathcal{L}^2(\mathbb{R})$ *and given* $(f, \mathbf{p}, \mathcal{D}(\mathbf{p}), \mathcal{P})$ *admissible. Denote by* $\mathcal{B}(\mathbb{R})$ *the Borel-*$\sigma$*-algebra on* $\mathbb{R}$ *and by* $E_{\mathbf{k}}, E_{\mathbf{p}}$ *the spectral measures for* $\mathbf{k}$ *and* $\mathbf{p}$, *respectively.*

*Consider the following conditions:*

- *There exists* $\epsilon_0$ *such that* $\forall\, 0 < \epsilon \leq \epsilon_0\, \exists \epsilon' > 0$ *and* $\exists \delta > 0$ *it holds:*
$$\|E_{\mathbf{k}}([-\epsilon, \epsilon]) - E_{\mathbf{p}}([-\epsilon', \epsilon'])\| < \delta \tag{A2}$$
*and* $\delta \in O(\epsilon^3)$.
- *Given the strongly-continuous 1-paramter group* $U_t$ *of translations generated by* $\mathbf{x}$ *we require that for any measurable set* $\Omega \in \mathcal{B}(\mathcal{R})$ *there exists a measureable set* $\Omega' \in \mathcal{B}(\mathbb{R})$ *and a projection* $Q \leq \mathbb{I} - P$, *where* $P$ *is the projection onto* $\mathcal{P}$, *such that* $U_t E_{\mathbf{p}}(\Omega)U_{-t} = E_{\mathbf{p}}(\Omega') + Q$.

*Then it follows that* $\mathbf{k}$, *when restricted to* $\mathcal{P}$, *has a self-adjoint extension* $\mathbf{k}_P$ *on* $\mathcal{P}$ *and that* $\mathbf{p}$ *and* $\mathbf{k}_P$ *strongly commute as operators on* $\mathcal{P}$.

By [65, Corollary 5.28] there then exists a dense linear subspace in $\mathcal{P}$ invariant under both operators $\mathbf{p}$ and the restricted $\mathbf{k}$, such that they commute on vectors from this domain.

*Proof.* The idea is to show that the spectral projections commute for any pair of finite Borel sets $\Omega, \Omega' \in \mathcal{B}(\mathcal{H})$. These sets can be partitioned into finitely many intervals of diameter $\epsilon$ and $\epsilon'$. The spectral projections of $\mathbf{k}$ for these intervals are simply translates of $[-\epsilon', \epsilon]$ by the unitary group $U_t$ generated by the position operator $\mathbf{x}$. The idea now is to show that we can approximate the projections $U_{-t}E_{\mathbf{p}}([-\epsilon, \epsilon])U_t$ by spectral projections $E_p(\tilde{\Omega})$ for some $\tilde{\Omega} \in \mathcal{B}(\mathbb{R})$ which depends on $\epsilon, \epsilon'$ and $t$. Note that since $\mathcal{B}(\mathbb{R})$ is generated by open and bounded intervals, and since spectral measures are countably additive the assertion holds for all measurable sets if its holds for bounded open intervals.

Let $\Omega, \Omega' \in \mathcal{B}(\mathbb{R})$ be bounded and open intervals. Choose any $\epsilon, \epsilon' > 0$ and let $\Omega^\alpha_\epsilon$ and $\Omega^\beta_\epsilon$ be a cover of $\Omega$ and $\Omega'$ by disjoint open intervals $\Omega^\alpha_\epsilon := (-\epsilon + \alpha, \epsilon + \alpha)$ and $\alpha, \beta$ are the corresponding indices for the shifts. By countable additivity it follows

$$[E_{\mathbf{k}}(\Omega), E_{\mathbf{p}}(\Omega')] = \sum_{\alpha,\beta} [E_{\mathbf{k}}(\Omega^\alpha_\epsilon), E_{\mathbf{p}}(\Omega^\beta_\epsilon)].$$

The unitary group $U_t$ generated by $\mathbf{x}$ acts as translations on $L^2(\mathbb{R})$ when we view $\mathbf{k}$ as multiplication operator. Hence we have for the spectral projections $E_{\mathbf{k}}(\Omega)$ for some measurable set $\Omega \in \mathcal{B}(\mathbb{R})$ that $U_t E_{\mathbf{k}}(\Omega) U_{-t} = E_{\mathbf{k}}(\Omega + t)$. I.e. for each index $\alpha$ we get $E_{\mathbf{k}}(\Omega^\alpha_\epsilon) = U_\alpha E_{\mathbf{k}}(\Omega_\epsilon) U_{-\alpha}$ where $\Omega_\epsilon := \Omega^0_\epsilon$.

Now choose $\epsilon < \epsilon_0$ and let $\epsilon'$ and $\delta$ such that $\|E_{\mathbf{k}}(\Omega_\epsilon) - E_{\mathbf{p}}(\Omega_{\epsilon'})\| < \delta$. Choose index sets $I, J$ such that $\Omega \subset \bigcup_{\alpha \in I} \Omega^\alpha_\epsilon =: \Omega^I_\epsilon$ and $\Omega' \subset \bigcup_{\beta \in J} \Omega^\beta_{\epsilon'} =: \Omega^J_{\epsilon'}$. It is enough to show that the commutators $[E_{\mathbf{k}}(\Omega^I_\epsilon), E_{\mathbf{p}}(\Omega^J_{\epsilon'})]$ are small since this implies that this is also true for the smaller projections $E_{\mathbf{k}}(\Omega)$ and $E_{\mathbf{p}}(\Omega')$.

By assumption we get for all such $\alpha$ and with the notation $\Omega_{\epsilon'} := \Omega^0_{\epsilon'}$

$$\|E_{\mathbf{k}}(\Omega^\alpha_\epsilon) - U_\alpha E_{\mathbf{p}}(\Omega_{\epsilon'}) U_{-\alpha}\| < \delta.$$

Furthermore, we have that for each $\alpha \in I$ there exists a measureable set $\Omega^\alpha_{\epsilon'}$ and a projections $Q_\alpha \leq \mathbb{I} - P$ such that

$$U_\alpha E_{\mathbf{p}}(\Omega_{\epsilon'}) U_{-\alpha} = E_{\mathbf{p}}(\Omega^\alpha_{\epsilon'}) + Q_\alpha.$$

Hence, we obtain the following estimate:

$$\begin{aligned} \|[E_{\mathbf{k}}(\Omega), E_{\mathbf{p}}(\Omega')]\| &\leq \sum_{\alpha,\beta} \|[E_{\mathbf{k}}(\Omega^\alpha_\epsilon), E_{\mathbf{p}}(\Omega^\beta_\epsilon)]\| \\ &\leq \sum_{\alpha,\beta} \big(\|[U_\alpha E_{\mathbf{p}}(\Omega_{\epsilon'}) U_{-\alpha}, E_{\mathbf{p}}(\Omega^\beta_\epsilon)\| + 2\delta\big) \\ &= \sum_{\alpha,\beta} \big[E_{\mathbf{p}}(\Omega^\alpha_{\epsilon'}) + Q_\alpha, E_{\mathbf{p}}(\Omega^\beta_\epsilon]\| + 2\delta\big) \\ &= \sum_{\alpha,\beta} 2\delta = 2|J||I|\delta \leq \frac{4|\Omega||\Omega'|}{\epsilon^2}\delta. \end{aligned}$$

Since this estimate only depends on the partitioning of $\Omega$ and $\Omega'$ into small intervals and since we can choose this partitioning arbitrarily small, it follows that

$$\forall\Omega,\Omega'\in\mathcal{B}(\mathbb{R}):\|[E_{\mathbf{k}}(\Omega),E_{\mathbf{p}}(\Omega')]\|=0.$$

In particular this implies that for all measurable sets $\Omega\in\mathcal{B}(\mathbb{R})$ the projections $E_{\mathbf{k}}(\Omega)$ commute with the projection $P:\mathcal{H}\to\mathcal{P}$. By [64, Theorem 75.1] this implies that $P$ is a function of $\mathbf{k}$, i.e. there exists a measurable function $\chi_P$ on $\mathbb{R}$ such that $P=\int_{\mathbb{R}}\chi_P(\lambda)E_{\mathbf{k}}(\mathrm{d}\lambda)$. Furthermore, $\chi_P$ is positive, $\chi_P^2=\chi_P$ and $\|\chi_P\|_\infty=1$, hence it is an indicator function of some measurable set $I\in\mathcal{B}(\mathbb{R})$, and the projection $P$ coincides with $E_{\mathbf{k}}(I)$. The restriction of $\mathbf{k}$ to $\mathcal{P}$ is therefore given by

$$\mathbf{k}_P=\int_{\mathbb{R}}\chi_P(\lambda)\lambda E_{\mathbf{k}}(\mathrm{d}\lambda).$$

q.e.d.

**Proposition 1.** *Given the standard position and momentum operators* $\mathbf{x}$ *and* $\mathbf{k}$ *on* $\mathcal{H}=\mathcal{L}^2(\mathbb{R})$ *and given* $(f,\mathbf{p},\mathcal{D}(\mathbf{p}),\mathcal{P})$ *admissible, satisfying the assumptions in Theorem 4. Then there exists an interval* $I\subseteq\mathbb{R}$ *such that* $\mathcal{P}=\mathcal{L}^2(I)$*, and a function* $p:I\to\mathbb{R}$ *such that* $\mathbf{p}=p(\mathbf{k})$ *on* $\mathcal{P}$*. The interval is determined by the function* $f$ *in the following sense:*

- *If the function* $g(p)=\int_0^p\frac{1}{f(s)}\mathrm{d}s$ *is bounded then* $I=[-k_{max},k_{max}]$ *and* $k_{max}$ *is given by*

$$k_{max}=\int_0^\infty\frac{1}{f(p)}\mathrm{d}p.$$

- *If* $g$ *is unbounded then* $I=\mathbb{R}$*.*

*Proof.* By the previous theorem the operators $\mathbf{k}_P$ and $\mathbf{p}$ strongly commute on $\mathcal{P}$. Let $I\subset\mathbb{R}$ the measureable set with $E_{\mathbf{k}}(I)=P$ and $P:\mathcal{H}\to\mathcal{P}$ the projection onto $\mathcal{P}$. By assumption the spectrum of $\mathbf{p}$ is simple and therefore the spectral projections $E_{\mathbf{k}}(\Omega)$ with $\Omega\subset I$ are bounded functions of $\mathbf{p}$ [35, Theorem VII.5]. Since for any such $\Omega$ $E_{\mathbf{k}}(\Omega)$ is a projection there exists a measureable set $\Theta\subset\mathbb{R}$ such that $E_{\mathbf{k}}(\Omega)=E_{\mathbf{p}}(\Theta)$. Therefore there exists an almost everywhere finite, measureable, real-valued function $g$ on $\mathbb{R}$ such that

$$P\mathbf{k}P=\int_{\mathbb{R}}g(\lambda)E_{\mathbf{p}}(\mathrm{d}\lambda),$$

i.e. $P\mathbf{k}P=g(\mathbf{p})$. Now let $\psi\in\mathcal{D}$. Then, by assumption

$$[\mathbf{x},\mathbf{p}]\psi=if(\mathbf{p})\psi.$$

Then

$$[\mathbf{x},g(\mathbf{p})]\psi=ig'(\mathbf{p})f(\mathbf{p})\psi.$$

But, since $[\mathbf{x},\mathbf{k}]\psi=i\psi$ we must have that $g'=\frac{1}{f}$, therefore

$$g(p)=\int_0^p\frac{1}{f(s)}\mathrm{d}s.$$

Since the spectrum $\sigma(g(\mathbf{p}))$ of $g(\mathbf{p})$ is the essential range of the function $g$ we see that the spectrum of $P\mathbf{k}P$ is the interval $I=[-k_{max},k_{max}]$, if $g$ is bounded, and $\mathbb{R}$ if $g$ is unbounded. Hence the subspace $\mathcal{P}$ is isomorphic to $\mathcal{L}^2(I)$.

Conversely, since $\mathbf{k}$ has simple spectrum, there exists a function $h$ such that $\mathbf{p}=h(\mathbf{k})$, and this function is necessarily unbounded. q.e.d.

## APPENDIX B: PROOF OF THEOREM 2

We consider observables $\mathbf{x}$ and $\mathbf{p}$ that obey a modified commutation relation as described is Section II. Since we are interested in uncertainty relations that are optimal for all states, we need to consider mixed states as well. In order to focus on pure states one would need to first show that for all mixed states there exists a pure state that is "more optimal", a notion that we will make precise in the following.

We denote by $\Omega$ the set of mixed states $\rho$, given by the density operators from $\mathcal{B}(\mathcal{P})$. When considering variances, the corresponding uncertainty region is given by

$$\mathcal{U}=\left\{(\Delta_\rho\mathbf{x},\Delta_\rho\mathbf{p})\in\mathbb{R}_+^2\middle|\rho\in\Omega\right\}. \tag{B1}$$

We can give a precise definition of the trade-off curve by introducing a partial ordering relation "<" ("≤"), which is given by saying that $v<w$ ($v\leq w$), for $v,w\in\mathbb{R}^2$, if every component of $v$ is smaller (small or equal) than the corresponding component of $w$. The trade-off curve $\Gamma(\mathcal{U})$, i.e. the desired optimal and state-independent uncertainty relation, is then given by all tuples from $\mathcal{U}$ that are minimal in $\mathcal{U}$ with respect to the above ordering "<", i.e.

$$\Gamma(\mathcal{U})=\left\{v\in\mathcal{U}\middle|\nexists w\in\mathcal{U}:w<v\right\}. \tag{B2}$$

Unfortunately, no general efficient method for computing this trade-off curve is known. However, in Theorem 5, we circumvent this circumstance by first providing a lower bound on $\Gamma(\mathcal{U})$ and then showing that this bound can be attained. For the first step we need the notion of a lower convex hull $\mathcal{U}_{lc}$: This is obtained by first filling up $\mathcal{U}$ with all points that are more uncertain than, at least, some point from $\mathcal{U}$, and then taking the convex hull of this set, i.e.

$$\mathcal{U}_{lc}=\mathrm{Conv}\left(\left\{v\in\mathbb{R}^2\middle|\exists w\in\mathcal{U}:w\leq v\right\}\right). \tag{B3}$$

Note that this does not add any additional extremal points other than those already contained in the convex hull of $\mathcal{U}$ (which are therefore already contained in $\mathcal{U}$).

**Theorem 5.** *Let* $\mathbf{x}$*,* $\mathbf{p}$*,* $\Omega$*,* $\mathcal{U}$ *and* $\mathcal{U}_{lc}$ *be given as described above and let* $\mathcal{U}_{00}$ *denote the set of attainable tuples of second moments, i.e.*

$$\mathcal{U}_{00}=\left\{(\mathrm{tr}(\rho\mathbf{x}^2),\mathrm{tr}(\rho\mathbf{p}^2))|\rho\in\Omega\right\}\ . \tag{B4}$$

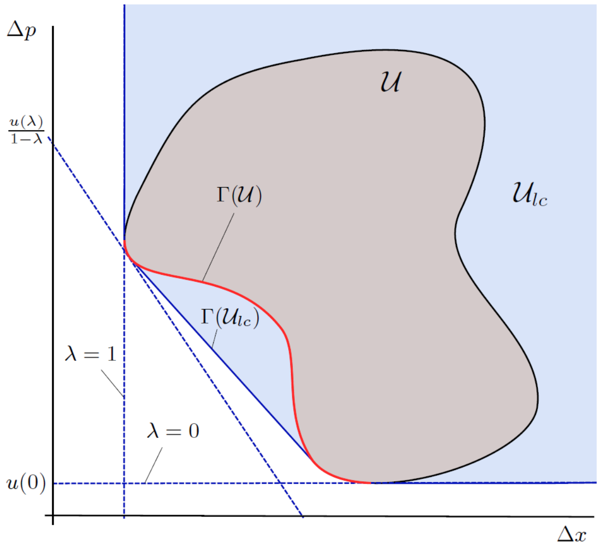


Fig. 10: The lower convex hull $\mathcal{U}_{lc}$ (blue region) of some set $\mathcal{U}$ (grey region) is obtained by considering hyperplanes (dashed blue lines) for $\lambda \in (0, 1)$. As in (20) the trade-off curve $\Gamma(\mathcal{U}_{lc})$ (solid blue line) of the convex hull can be parametrised by values $u(\lambda)$. The trade-off curve $\Gamma(\mathcal{U}_{lc})$ is never "worse" than the trade-off curve $\Gamma(\mathcal{U})$ (solid red line); both curves coincide if $\mathcal{U}$ is convex itself.

*The notation indicates that $\mathcal{U}_{00}$ corresponds to the set obtained by restricting $\mathcal{U}$ to states with expectations $\langle \mathbf{x} \rangle = \langle \mathbf{p} \rangle = 0$. Then*

$$\Gamma(\mathcal{U}_{lc}) = \Gamma(\mathcal{U}) = \Gamma(\mathcal{U}_{00}) \ . \tag{B5}$$

*Proof.* For any two sets $\mathcal{V}_1$ and $\mathcal{V}_2$ we introduce a partial ordering relation "$\preceq$" on the corresponding curves $\Gamma(\mathcal{V}_1)$ and $\Gamma(\mathcal{V}_2)$, respectively, by saying that $\Gamma(\mathcal{V}_1) \preceq \Gamma(\mathcal{V}_2)$ if and only if for all points $v_2 \in \Gamma(\mathcal{V}_2)$ there is a point $v_1 \in \Gamma(\mathcal{V}_1)$ such that $v_1 \leq v_2$.

By construction we know that $\mathcal{U}_{lc}$ is convex and $\Gamma(\mathcal{U}_{lc}) \preceq \Gamma(\mathcal{U})$. Note that $\mathcal{U}_{00}$ is also convex since it is obtained as the range of an affine map of the convex set $\Omega$. Additionally, points in $\Gamma(\mathcal{U}_{00})$ are always in $\mathcal{U}$ and hence

$$\Gamma(\mathcal{U}_{lc}) \preceq \Gamma(\mathcal{U}) \preceq \Gamma(\mathcal{U}_{00}) \ . \tag{B6}$$

In the following we will prove that in fact $\Gamma(\mathcal{U}_{lc}) = \Gamma(\mathcal{U}_{00})$, which immediately implies the desired statement (B5).

Since $\mathcal{U}_{lc}$ is convex, it is fully characterised by the intersection of its supporting halfspaces (see for example [33]), i.e. there is a function $u(\lambda)$ such that

$$\mathcal{U}_{lc} = \left\{ (v_1, v_2) \in \mathbb{R}^2_+ | \forall \lambda \in \mathbb{R} : \lambda v_1 + (1-\lambda) v_2 \geq u(\lambda) \right\} \tag{B7}$$

The boundary of $\mathcal{U}_{lc}$ is the set of points which have an intersection with a supporting hyperplane, i.e. $\lambda v_1 + (1-\lambda) v_2 = u(\lambda)$. Moreover, $\Gamma(\mathcal{U}_{lc})$ consists of points on the boundary, hence that attain equality for $\lambda \in (0, 1)$. Conversely, $u(\lambda)$ can be obtained by minimizing $\lambda v_1 + (1-\lambda) v_2$ over all points in $\mathcal{U}_{lc}$, for a fixed $\lambda$. However, for $\lambda$ in $[0, 1]$, this minimization of a linear functional will attain its minimum also on extremal points of $\mathcal{U}_{lc}$, which are contained in the boundary of $\mathcal{U}$. We therefore can write

$$u(\lambda) = \inf_{\rho \in \Omega} \lambda \Delta_\rho \mathbf{x} + (1-\lambda) \Delta_\rho \mathbf{p} \ . \tag{B8}$$

At this stage we can reformulate the variance of an observable $A$ as

$$\Delta_\rho A = \min_{a \in \mathbb{R}} \langle (A - a)^2 \rangle_\rho \ , \tag{B9}$$

such that the r.h.s of (B8) turns into

$$\min_{\alpha, \eta \in \mathbb{R}} \inf_{\rho \in \Omega} \lambda \langle (\mathbf{x} - \eta)^2 \rangle_\rho + (1-\lambda) \langle (\mathbf{p} - \alpha)^2 \rangle_\rho \ . \tag{B10}$$

The expectation of $\mathbf{x}$ can be shifted by multiplying with $\exp(i\eta\mathbf{k})$. As $\mathbf{p}$ commutes with $\mathbf{k}$, (see section II) this procedure will not affect the variance $\mathbf{\Delta p}$, such that we can always set $\eta = 0$ in the following. If we now represent $\mathbf{p}$ as a function of the coordinate $k$ and $\mathbf{x}$ as $i\partial_k$, the minimization over $\rho$ in (B10) corresponds to finding the ground state energy, $E_\alpha$, of the Schrödinger operator

$$H_\alpha := -\lambda \partial_k^2 + V_\alpha(k) \tag{B11}$$

with potential $V_\alpha(k) = (1-\lambda)(p(k) - \alpha)^2$, i.e.

$$u(\lambda) = \min_\alpha E_\alpha . \tag{B12}$$

As $V_\alpha(k)$ is positive and thus bounded from below for every $\alpha$, there is a unique function $\check{V}_\alpha(k)$ which gives the best convex approximation to $V_\alpha(k)$ from below, i.e the super graph of $\check{V}_\alpha(k)$ is the convex hull of the super graph of $V_\alpha(k)$. If needed, $\check{V}_\alpha(k)$ can be obtained by Legendre transforming $V_\alpha(k)$ twice. Now, for all states $\rho$, we have $\langle V_\alpha(k) \rangle_\rho \geq \langle \check{V}_\alpha(k) \rangle_\rho$ and hence we can lower bound $E_\alpha$ by $\check{E}_\alpha$, which is the ground state energy of the Schrödinger operator $-\lambda \partial_k^2 + \check{V}_\alpha(k)$.

Note that $\check{V}_\alpha(k)$ is a convex function in $\alpha$, because $V_\alpha(k)$ is convex in $\alpha$. We can therefore employ Corollary 13.6 from [66] to show that $\check{E}_\alpha$ is a convex function of $\alpha$, too.

Moreover, $\check{V}_\alpha$ inherits the symmetry $\check{V}_\alpha(-k) = \check{V}_{-\alpha}(k)$ from $V_\alpha(k)$, which can be implemented by a unitary automorphism on $\Omega$. This shows the symmetry $\check{E}_\alpha = \check{E}_{-\alpha}$, which directly implies that $\check{E}_\alpha$ becomes minimal for $\alpha = 0$. But here we have that $\check{V}_0(k) = V_0(k)$, because $V_0(k)$ is already a convex function, which yields

$$\check{E}_0 = E_0 = u(\lambda). \tag{B13}$$

We thus know that, for $\lambda \in (0, 1)$, the extremal points of $\mathcal{U}_{lc}$ (which are in $\Gamma(\mathcal{U}_{lc})$) have zero expectation in $\mathbf{x}$ and $\mathbf{p}$, i.e. they lie within $\mathcal{U}_{00}$. Since $\Gamma(\mathcal{U}_{lc}) \preceq \Gamma(\mathcal{U}_{00})$ we even know that on these points the boundaries $\Gamma(\mathcal{U}_{lc})$ and $\Gamma(\mathcal{U}_{00})$ coincide. But as $\mathcal{U}_{lc}$ and $\mathcal{U}_{00}$ are both convex, this implies that $\Gamma(\mathcal{U}_{lc}) = \Gamma(\mathcal{U}_{00})$. q.e.d.

## APPENDIX C: A STATE-INDEPENDENT BUT NOT SO OPTIMAL BOUND

Assume a modification $f(p)$ of the Heisenberg algebra with Taylor expansion $f(p) = 1 + \sum_{n=1}^{\infty} a_n p^{2n}$ and $a_n \geq 0$ for all $n \in \mathbb{N}$. This implies that $f$ is convex, monotonously increasing for $p \geq 0$, smooth and symmetric around the origin with $f(0) = 1$, hence fulfilling our assumptions (i-iii) in the main text. If we set

$$g(p) := f(-\sqrt{|p|}) \ , \tag{C1}$$

we can see that

$$g(p^2) = f(-\sqrt{|p^2|}) = f(-|p|) = f(|p|) = f(p) \ . \tag{C2}$$

Now, restricted to $p > 0$, $g(p)$ arises as a concatenation of convex functions, thus $g$ is also convex in this parameter range. Inserting (C2) into the Robertson Kennard relation (16), and using Jensen's inequality together with the convexity of $g$, we get

$$\mathbf{\Delta x \Delta p} \geq \frac{1}{4} |\langle f(\mathbf{p}) \rangle|^2 \geq \frac{1}{4} |g(\langle \mathbf{p}^2 \rangle)|^2 \ .$$

Now we can substitute $\langle p^2 \rangle = \mathbf{\Delta p} + \langle \mathbf{p} \rangle^2$ and use the properties of $g$ as well as the convexity of the absolute square, to arrive after a simple calculation at

$$\begin{aligned} \mathbf{\Delta x \Delta p} &\geq \frac{1}{4} |g(\mathbf{\Delta p} + \langle \mathbf{p} \rangle^2)|^2 \\ &\geq \frac{1}{4} |g(\mathbf{\Delta p})|^2 + \frac{1}{4} |g(\langle \mathbf{p} \rangle^2)|^2 - 1 \ . \end{aligned}$$

Here the state-dependent term $\frac{1}{4}|g(\langle \mathbf{p} \rangle^2)|^2$ is greater than or equal to one, such that we can conclude the state-independent bound

$$\mathbf{\Delta x \Delta p} \geq \frac{1}{4} g(\mathbf{\Delta p})^2 \ . \tag{C3}$$

However this bound is not optimal for general modifications. Examples for this can be seen in Fig. 5.

# Curriculum Vitae

| | |
|---|---|
| Full Name: | René Schwonnek |
| Date of Birth: | 18.05.1987 |
| Place of Birth: | Langenhagen |
| Citizenship: | German |

## Academic career

1999-2006 Abitur at the Goetheschule Hannover

2006-2013 Studies in Physics at the Leibniz University Hannover

2010 Bachelorthesis at the Institute for Theoretical Physics
*Störungstheorie und asymptotische Reihen*

2013 Masterthesis at the Institute for Theoretical Physics
*Dynamische Modellierung von Quantenrepeater-Prozessen*

2013-2018 Research scientist at the Institute for Theoretical Physics of the Leibniz University Hannover