



**HAL**  
open science

# The interplay between quantum contextuality and Wigner negativity

Pierre-Emmanuel Emeriau

► **To cite this version:**

Pierre-Emmanuel Emeriau. The interplay between quantum contextuality and Wigner negativity. Quantum Physics [quant-ph]. Sorbonne Université, 2021. English. NNT: 2021SORUS419. tel-03987476v2

**HAL Id: tel-03987476**

**<https://theses.hal.science/tel-03987476v2>**

Submitted on 14 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



SORBONNE UNIVERSITÉ - EDITE DE PARIS

*Laboratoire d'Informatique de Sorbonne Université (LIP6)*

---

---

# The interplay between quantum contextuality and Wigner negativity

---

---

BY PIERRE-EMMANUEL EMERIAU

PHD THESIS IN COMPUTER SCIENCE

UNDER THE SUPERVISION OF ELHAM KASHEFI AND SHANE MANSFIELD

*Keywords:* quantum information, contextuality, Wigner negativity, optimisation theory, measure theory, continuous-variables

PhD defence held publicly in November 2021 with the following thesis committee:

- ABRAMSKY Samson, Christopher Strachey Professor of Computing at Oxford University, United Kingdom.
- ACÍN Antonio, Group Leader ICREA Professor at ICFO, Spain.
- ARRIGHI Pablo, Professeur HDR, Université de Paris-Saclay, France. *Rapporteur.*
- DEL RIO Lídia, Senior Researcher at ETH Zurich, Switzerland.
- GALVÃO Enersto, Group leader at INL, Portugal. *Rapporteur.*
- KASHEFI Elham, Directrice de Recherche HDR au CNRS, Sorbonne Université, France and Professor at University of Edinburgh, United Kingdom. *PhD supervisor.*
- MANSFIELD Shane, Head of Quantum Algorithms, Quandela. France, *PhD supervisor.*
- PARIGI Valentina, Associate Professor at LKB, Sorbonne Université, Paris, France.

This work is licensed under the [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

# ABSTRACT

QUANTUM physics has revolutionised our way of conceiving nature and is now bringing about a new technological revolution. The use of quantum information in technology promises to supersede the so-called classical devices used nowadays. Understanding what features are inherently non-classical is crucial for reaching better-than-classical performance.

This thesis focuses on two nonclassical behaviours: quantum contextuality and Wigner negativity. The former is a notion superseding nonlocality that can be exhibited by quantum systems. To date, it has mostly been studied in discrete-variable scenarios, where observables take values in discrete and usually finite sets. In those scenarios, contextuality has been shown to be necessary and sufficient for advantages in some cases. On the other hand, negativity of the Wigner function is another unsettling non-classical feature of quantum states that originates from phase-space formulation in continuous-variable quantum optics. Continuous-variable scenarios offer promising candidates for implementing quantum computations and informatic protocols. Wigner negativity is known to be a necessary resource for quantum speedup with continuous variables. However contextuality has been little understood and studied in continuous-variable scenarios.

We first set out a robust framework for properly treating contextuality in continuous variables. We also quantify contextuality in such scenarios by using tools from infinite-dimensional optimisation theory. This is achieved by a converging hierarchy of finite-dimensional semidefinite programs that approximates the contextual fraction.

Building upon this, we show that Wigner negativity is equivalent to contextuality in continuous variables with respect to Pauli measurements thus establishing a continuous-variable analogue of a celebrated result by Howard *et al.* in discrete variables.

We then introduce experimentally-friendly witnesses for Wigner negativity of single mode and multimode quantum states, based on fidelities with Fock states. They possess a threshold expectation value indicating whether the measured state has a negative Wigner function. We phrase the problem of finding the threshold values as infinite-dimensional linear programs, and we derive two converging hierarchies of semidefinite programs to approximate the threshold values.

We further extend the range of previously known discrete-variable results linking contextuality and advantage into a new territory of information retrieval. We introduce a discrete-variable communication game—called the Torpedo Game—where perfect quantum strategies stem from negativity of the discrete Wigner function. Sequential contextuality is shown not only to be necessary and sufficient for quantum advantage, but also to quantify the degree of advantage for information retrieval tasks.

## DEDICATION AND ACKNOWLEDGEMENTS

These last three years have been incredibly rich, both scientifically and personally. I am deeply indebted to my supervisors Elham and Shane who, from the very beginning, have continuously shared their love of Quantum Information in so many various topics. My grateful thanks to Elham for her amazing support and her perfectly adjusted mixture of guidance and freedom throughout the projects. I cannot begin to express my thanks to Shane, who has been a wonderful supervisor, with incredible insight into quantum foundational related questions. I can safely say that I learnt more about Quantum Information than on beer related topics by his side (and I did learn a great deal on craft beers!). Ní fhéidfinn go leor buíochas a ghabhadh le Shane. Tá an t-eipeagraf ar do shon.

I would also like to extend my deepest gratitude to my coauthors Mark, Rui, Frédéric, Ulysse and Robert. Mark has been continuously present despite the distance from the end of the first year onward. I will be forever grateful to Ulysse for the numerous insightful discussions I had with him and his impressive dedication as he would rather proofread this dissertation than watch a movie or read a book during a 15 hour flight. Many thanks to Robert and his passion for  $C^*$ -algebras; his strong mathematical background allowed for fruitful conversations. Special thanks to Antoine who opened the door to the optimisation theory realm for me. I must also thank Pablo Arrighi and Anthony Leverrier for their support with the supervising committee. Thanks to Ricardo and Anna with whom I had interesting discussions on quantum game theory.

Many thanks to Elham, Damian, Eleni, Frédéric and Alex for the wonderful QI team they have built in Paris and again Elham for the amazing Edipar group in Paris and Edinburgh. This was an invaluable asset during my PhD as I could have intense discussions in so many topics with passionate and diverse people. These three years have been a delight because of the numbers of people I could interact with, both in research and outside the lab—mostly at Baker street, Bières Cultes or around the 'quais St Bernard' or during amazing retreats. For that I thank the Edipar group and its brilliant members Ellen, Léo, Luka, Niraj, Atul, Brian, Dominik, Harold, Daniel, Mina, Alex, Rawad, Mina, Mahshid, Theodoros, Jonas, Ieva, Yao, Armando. I also thank the members of the marvelous QI team Raja, Victor, Federico, Nathan, Clément, Luis, Matthieu, Anu, Andrea, Alisa, Adrien, Simon, Shouvik, Matteo, Francesco, Paul, Ivan. I also thank everyone at Quandela; it has been extremely stimulating to work on experimentally related projects.

Évidemment un immense merci à tous mes merveilleux amis qui m'ont accompagné en dehors du labo. À Rémi et Xavier : Bagneux on n'oublie pas. À Robin parce qu'il est bonnard. À Eliot mashallah. À François pour ce but vide que tu m'as offert et que j'ai mis à côté scellant notre amitié. À Lucie pour ses piggybacks endiablés. À Raja pour sa joie de vivre au quotidien qui réchauffe les cœurs. À Ulysse pour toutes ces nombreuses galettes épinards partagées et

---

également Léonie pour ces soirées jeux de société et escape games brillamment réussis. À mon ami de plus longue date Johan. À mon fillot Antoine pour tous ces moments riches depuis Ginette. Aux choletais (et alentours) Mathieu, Marie et Thomas qui m'ont soutenu depuis si longtemps déjà. À Jean-Victor et Raphaëlle pour ces nombreux partages sensibles. Et aussi à Aude, Coline, Tamara, Mathilde G., Mathilde S. et Alice.

Un grand merci au groupe de Ginette Côme, Clément, Victor, Paul, Anne-Claire, Gustave, Pierre H et Peter Battle et également aux coincheurs Léa, Sylvain, Paul Pi Su Sardu. Au PTRC pour m'avoir fait découvrir ce sport magnifique et ses membres incroyables: un véritable exhutoire pendant ces 3 ans. Merci à Damien et Leïla. Merci à Léa M., Romane et Roméo, Yaelin et nos parties d'échecs, Hortense. Merci aussi à Ev et Tom. À Etienne et l'atelier Mala pour ses nouilles veggies. Enfin merci à Matthieu et Virginie pour m'avoir aiguillé sur le bon chemin postbac.

À tous les enseignants incroyables que j'ai eu la chance de rencontrer sur mon parcours et qui ont su me passionner; en particulier Cyril Ciaudo, Alexianne Poutevigne, Hamid Seddiq, Bertrand Delahaye, Christophe Collet, Françoise Thebault, Jean Nougareyde, Bertrand Delamotte, Jérôme Beugnon et Patrick Ciarlet.

Un immense merci à Alexandre, Jennifer, Pacôme et Manolie pour leur soutien sur et en dehors des tatamis.

Merci à Caroline et Keat pour tout ce qu'ils ont déjà fait pour moi.

Merci à tout ma famille, grand-parents, oncles, tantes et cousins.

Évidemment une reconnaissance éternelle à mes parents Isabelle et Bruno et ma sœur Mathilde pour m'avoir accompagné jusqu'ici, pour toutes les valeurs qu'ils m'ont transmises et pour leur générosité abyssale. Un immense merci à Mathieu pour sa curiosité sans borne et son partage d'expériences et de connaissances. À Alice pour ses "areu" encourageants.

À Héloïse pour son soutien immuable et pour l'équilibre apporté dans ma vie tout au long de cette thèse. Mes mots ne pourraient se hisser à la hauteur de l'amour et l'affection qu'elle m'a portés. Merci d'être dans ma vie; on a encore tant de choses à découvrir tous les deux.

Sa sagesse résonnait jusque dans ses silences.  
Jamais ses réponses n'étaient précipitées.  
Jamais mes questions n'étaient sous-estimées.

---

*Le roi n'avait pas ri*  
Guillaume MEURICE

# TABLE OF CONTENTS

	Page
<b>Introduction</b>	<b>1</b>
Summary of results . . . . .	4
Publications . . . . .	7
Notations . . . . .	8
<b>1 Frameworks and formalism</b>	<b>11</b>
1.1 Quantum information preliminaries . . . . .	11
1.1.1 Basics of quantum information theory . . . . .	11
1.2 Phase-space formalism . . . . .	14
1.2.1 Phase-space for $d$ -dimensional quantum states . . . . .	15
1.2.2 The case of continuous-variable quantum information . . . . .	16
1.2.3 The Wigner function . . . . .	18
1.3 Measure theory in a nutshell . . . . .	20
1.4 Optimisation theory . . . . .	23
1.4.1 Linear programming . . . . .	25
1.4.2 Semidefinite programming . . . . .	27
1.4.3 The Lasserre–Parrilo hierarchy . . . . .	28
1.4.4 Quadratic optimisation problems . . . . .	33
1.5 Sheaf-theoretic framework for contextuality . . . . .	35
1.5.1 Measurement scenarios and empirical models . . . . .	35
1.5.2 The FAB theorem . . . . .	40
1.5.3 Quantifying contextuality . . . . .	42
<b>2 Continuous-variable nonlocality and contextuality</b>	<b>46</b>
2.1 Framework . . . . .	48
2.2 A continuous-variable FAB theorem . . . . .	49
2.3 Quantifying contextuality . . . . .	53
2.3.1 The continuous-variable contextual fraction . . . . .	54
2.3.2 Monotonicity under free operations including binning . . . . .	54

2.3.3	Assumptions on the outcome spaces . . . . .	55
2.3.4	Infinite-dimensional linear programming for computing the contextual fraction . . . . .	56
2.3.5	Zero duality gap . . . . .	59
2.3.6	The case of local compactness . . . . .	60
2.4	Continuous generalisation of Bell inequalities . . . . .	63
2.5	Approximating the contextual fraction with SDPs . . . . .	65
2.5.1	Hierarchy of semidefinite relaxations for computing $\text{NCF}(e)$ . . . . .	65
2.5.2	Convergence of the hierarchy of SDPs . . . . .	71
2.6	Discussion and open problems . . . . .	72
<b>3</b>	<b>Equivalence of Wigner negativity and contextuality for continuous-variable Pauli measurements</b>	<b>75</b>
3.1	Measurement scenario under consideration . . . . .	76
3.2	Admissible empirical models . . . . .	80
3.3	Equivalence between Wigner negativity and contextuality . . . . .	83
3.4	Discussion and open problems . . . . .	85
<b>4</b>	<b>Witnessing Wigner negativity</b>	<b>87</b>
4.1	Wigner negativity witnesses . . . . .	89
4.2	Procedure for witnessing Wigner negativity . . . . .	93
4.2.1	Procedure . . . . .	93
4.2.2	Examples . . . . .	94
4.2.3	Numerical implementation . . . . .	99
4.3	Infinite-dimensional optimisation . . . . .	100
4.3.1	Sketch of the proofs . . . . .	101
4.3.2	Function spaces . . . . .	102
4.3.3	Computing the threshold value as a linear program . . . . .	108
4.3.4	Hierarchies of semidefinite programs . . . . .	114
4.3.5	Convergence of the hierarchies of semidefinite programs . . . . .	125
4.4	Witnessing multimode Wigner negativity . . . . .	140
4.4.1	Multimode notations . . . . .	140
4.4.2	Multimode Wigner negativity witnesses . . . . .	141
4.4.3	Infinite-dimensional linear programs . . . . .	142
4.4.4	Convergence of the multimode hierarchies . . . . .	152
4.4.5	Multimode example . . . . .	156
4.5	Discussion and open problems . . . . .	158
<b>5</b>	<b>Quantum advantage in Information Retrieval</b>	<b>159</b>

5.1	Information retrieval tasks . . . . .	161
5.1.1	Random access codes . . . . .	161
5.1.2	General Information Retrieval Tasks . . . . .	162
5.1.3	The Torpedo Game . . . . .	162
5.2	The discrete Wigner function and information retrieval . . . . .	164
5.3	Optimal strategies for the Torpedo Game . . . . .	166
5.3.1	Optimal quantum and post-quantum strategies . . . . .	166
5.3.2	Optimal classical strategies . . . . .	168
5.3.3	Comparison of quantum and classical game values . . . . .	171
5.3.4	Dimensional witness . . . . .	171
5.4	Sequential contextuality . . . . .	172
5.4.1	Empirical behaviours and hidden-variable Models . . . . .	173
5.4.2	Sequential contextuality in information retrieval tasks . . . . .	175
5.4.3	Quantified Contextual Advantage in Information Retrieval Tasks . . . . .	175
5.5	Discussion and open problems . . . . .	177
	<b>Conclusion and outlook</b>	<b>180</b>
	<b>Bibliography</b>	<b>183</b>
	<b>Alphabetical Index</b>	<b>209</b>

# INTRODUCTION

QUANTUM mechanics has been one of the major scientific revolutions of the 20<sup>th</sup> century. While philosophically puzzling [EPR35, Wig95, Sch05, PBR12, Man16], it has profoundly challenged our way of perceiving reality, giving rise to intriguing notions such as *entanglement* or *superposition*. It has also allowed us to perform remarkably precise calculation of properties of physical systems—especially at the particle scale. The so-called second quantum revolution is often traced to Feynman’s idea of a “probabilistic simulator of a probabilistic nature” [Fey82] to solve difficult problems originating from quantum physics together with the first quantum algorithms outperforming known classical ones [DJ92, Sho94].

This gave rise to the field of quantum information built around the idea that information may be encoded in quantum degrees of freedom of physical systems. Quantum information processing seeks to provide an advantage over classical information processing in various fields such as computing, communication, cryptography or sensing. To outperform standard classical computers at certain tasks, one can attempt to harness effects at the microscopic level which have no counterpart in classical physics.

Studying the fundamentally non-classical features of quantum mechanics is critical for quantum information. If non-classical features can be identified at the very root of quantum information protocols, and their advantage founded on those, then their quantumness can be inherently guaranteed. There is a lesson that can be learnt from the dequantization of the quantum recommendation system [KP17] that was proposed by Ewin Tang [Tan19]. If one wants to prevent against dequantization, it is necessary to ensure that algorithms rely upon some purely quantum phenomenon.<sup>1</sup> This leads naturally to the following questions:

*Which features of quantum physics are inherently non-classical?*

*Among those, which are relevant to achieving a quantum-over-classical advantage in information processing tasks?*

*How non-classicality can be related in a quantified way to advantages?*

---

<sup>1</sup>Note however that first it is still intriguing that a quantum algorithm had to be found in order to derive a classical algorithm with an improved complexity compared to known classical algorithms; secondly purely quantum notions will protect against dequantization but there might still exist classical algorithms that can be as efficient; finally the quantum algorithm still retains some advantage though it was massively diminished by Tang.

One of the first influential works indicating how quantum theory may depart from classical analogues was that of Einstein, Podolsky and Rosen [EPR35] in 1935. They identified early on that if the quantum description of the world is seen as fundamental then entanglement poses a problem of “spooky action at a distance”. It was argued that the principle of locality, a key axiom of Einstein’s special relativity, was in conflict with the description of the world given by quantum mechanics. The conclusion of their paper was that quantum theory should be consistent with a deeper or more complete description of the physical world, in which such problems would disappear. This is known as *local realism*. *Realism* means the introduction of a hidden variable model with hidden states describing the actual physics behind the scene. Indeed probabilistic predictions of quantum mechanics may result from an incomplete knowledge of the true state of a system rather than being a fundamental feature. There are other examples for this in physical theories: for instance statistical mechanics—a probabilistic theory—admits a deeper, very complex description in terms of classical mechanics, which is purely deterministic. *Local realism* is a constraint on the hidden variables that requires that they cannot be updated from a space-like separated region. This impossibility was further studied and instantiated as an experimentally verifiable inequality by Bell [Bel64] in 1964 that quantum mechanics is predicted to violate. This violation was verified experimentally e.g. [ADR82, RKM<sup>+</sup>01, GVW<sup>+</sup>15, SMSC<sup>+</sup>15, HBD<sup>+</sup>15]. This is an example of a *no-go theorem*: quantum physics cannot be described by a local hidden variable model.

*Local realism* was further generalised<sup>2</sup> as *measurement non-contextuality* [KS75]. Contextuality is a crucial non-classical behaviour that can be exhibited by quantum systems. The Heisenberg uncertainty principle identified that certain pairs of quantum observables are incompatible, e.g., position and momentum. In operational terms, observing one will disturb the outcome statistics of the other. Imprudent commentators will sometimes cite this as evidence that position and momentum cannot simultaneously be assigned definite values. However, this is not quite right and a more careful conclusion is that we simply cannot observe these values simultaneously. To make a stronger statement requires measurement contextuality. Roughly speaking, the meaning of words can provide an intuition for a contextual behaviour. The meaning of “I had the appendix removed” varies greatly whether I am addressing a physician or my supervisors for the present dissertation. Pre-assigning a meaning to this sentence is not possible irrespective of the context [AS14a, AS14b]. In more operational terms, contextuality is present whenever the behaviour of a system is inconsistent with the basic assumptions that (i) all of its observable properties may be assigned definite values at all times, and (ii) jointly measuring *compatible* observables does not disturb the global value assignments, or, in other words, these assignments are context-independent. Aside from its foundational importance, today contextuality

---

<sup>2</sup>Actually contextuality was originally considered to be similar but still a distinct non-classical feature rather than a generalisation of nonlocality. Fine [Fin82] proved that nonlocality is a special case of contextuality for the (2,2,2) Bell scenario and this was later generalised to any discrete-variable scenario by Abramsky and Brandenburger [AB11]. This is a consequence of the Fine–Abramsky–Brandenburger (FAB) theorem which we will present in Section 1.5.

is increasingly studied as an essential ingredient for enabling a range of quantum-over-classical advantages in informatic tasks, which include the onset of universal quantum computing in certain computational models [Rau13, HWVE14, ABM17, BVDB<sup>+</sup>17, ABdSZ17]. Moreover, contextuality has been tested experimentally e.g. [ZUZ<sup>+</sup>13, HLB<sup>+</sup>06, BKS<sup>+</sup>09, KZG<sup>+</sup>09].

The importance of seminal foundational results like the Bell [Bel64] and Bell–Kochen–Specker [Bel66, KS75] theorems is that they identify such non-intuitive behaviours and then rule out the possibility of finding *any* underlying model for them that would not suffer from the same issues. Incidentally, note that the EPR paradox was originally presented in terms of continuous variables, whereas the CHSH model (which is what was tested experimentally from Aspect onwards) addressed a discrete-variable analogue of it. Also frameworks for formalising contextuality have previously only focused on discrete-variable quantum information [AB11, CSW14, Spe05]. In Chapter 2, we provide the first robust framework formalising contextuality in continuous variables.

Quantum information with continuous variables [LB99]—where information is encoded in continuous degrees of freedom of quantum systems—is one of the promising directions for the future of quantum technologies [LB99]. From a theoretical point of view, quantum information with continuous variables is described via the formalism of infinite-dimensional Hilbert spaces. It offers different perspectives from discrete-variable quantum information. From a practical point-of-view, continuous-variable quantum systems are emerging as very promising candidates for implementing quantum informational and computational tasks [BVL05, WPGP<sup>+</sup>12, COR<sup>+</sup>13, BAV<sup>+</sup>21, Wal21, Cha21]. The main reason for this is that they offer unrivalled possibilities for deterministic generation of large-scale resource states over millions of modes [YYK<sup>+</sup>16, YUA<sup>+</sup>13] and also offer reliable and efficient detection methods, such as homodyne or heterodyne detection [Leo10, OE12]. Together they cover many of the basic operations required in the one-way or measurement-based model of quantum computing [RB01], for example. Typical implementations are in optical systems where the continuous variables correspond to the position-like and momentum-like quadratures of the quantised modes of an electromagnetic field. Indeed position and momentum, as mentioned previously in relation to the uncertainty principle, are the prototypical examples of continuous variables in quantum mechanics.

In order to handily manipulate continuous-variable states, mathematical tools initially inspired by quantum optics have been developed such as the phase-space formalism [Moy49] which has then been extended to discrete-variable systems [Gro06]. In this framework, quantum states are represented by a quasiprobability distribution over phase space, like the Wigner function [Wig32]. These representations provide a geometric intuition for quantum states [Lee91]. Quantum states are separated into two categories, Gaussian and non-Gaussian, depending on whether their Wigner function is a Gaussian function or not. The set of Gaussian states is well-understood [FOP05] but has limited power for computation. On the other hand characterising

the set of non-Gaussian states is an active research topic [AGPF18, TZ18, ZSS18, CMG20]. The latter are essential to a variety of quantum information processing tasks such as quantum state distillation [GC02, ESP02, Fiu02], quantum error-correction [NFC09], universal quantum computing [LB99, GS07] or quantum computational speedup [BSBN02, CFGM21]. Within those, an important subclass of non-Gaussian states are the states which display negativity in the Wigner function. These two classes of states coincide for pure states—namely, non-Gaussian pure states have a negative Wigner function. Indeed by Hudson’s theorem [Hud74, SC83], pure states with a positive Wigner function are necessarily Gaussian states. However, this is not the case for mixed states and the (convex) set of states with a positive Wigner function becomes much harder to characterise [MKC09, FMJ11]. In addition to its fundamental relevance as a non-classical property of physical systems [KŻ04], Wigner negativity is also essential for quantum computing as quantum computations described by positive Wigner functions can be simulated efficiently classically [ME12]. Wigner negativity is thus a necessary resource, though not sufficient [GACFF20], for quantum computational speedup. We explore intensively how to witness Wigner negativity, with reliable and accessible detection setup, in Chapter 4.

Knowing which characteristic lies at the source of better-than-classical performances can both allow for comparison of quantum systems in terms of their utility, and offer a heuristic for generating further examples of quantum-enhanced performance. This last point is highlighted in Chapter 5 where we derive an information retrieval game with a quantum-over-classical advantage. Since contextuality and Wigner negativity both seem to play a fundamental role as non-classical features enabling quantum-over-classical advantages in various tasks, another question that arises is:

*What is the precise relationship between Wigner negativity and contextuality?*

For discrete-variable systems of odd power-of-prime dimension, Howard *et al.* [HWVE14] showed that this negativity actually corresponds to contextuality with respect to Pauli measurements, thereby establishing the operational utility of contextuality for the gate-based model of quantum computation (particularly in a fault-tolerant setting). This equivalence was further established for odd dimensions [DOBV<sup>+</sup>17] and qubit systems [RBD<sup>+</sup>17, DAGBR15]. However this link has not previously been exhibited for continuous-variable systems and this is the subject of Chapter 3.

## Thesis structure and summary of results

**Chapter 1.** This introductory chapter presents the notions that will be used throughout this dissertation. It starts by giving an overview of quantum information in Section 1.1; then it gives the necessary tools for understanding the phase-space formulation of quantum mechanics in Section 1.2; it provides some background on measure theory in Section 1.3; it gives an overview

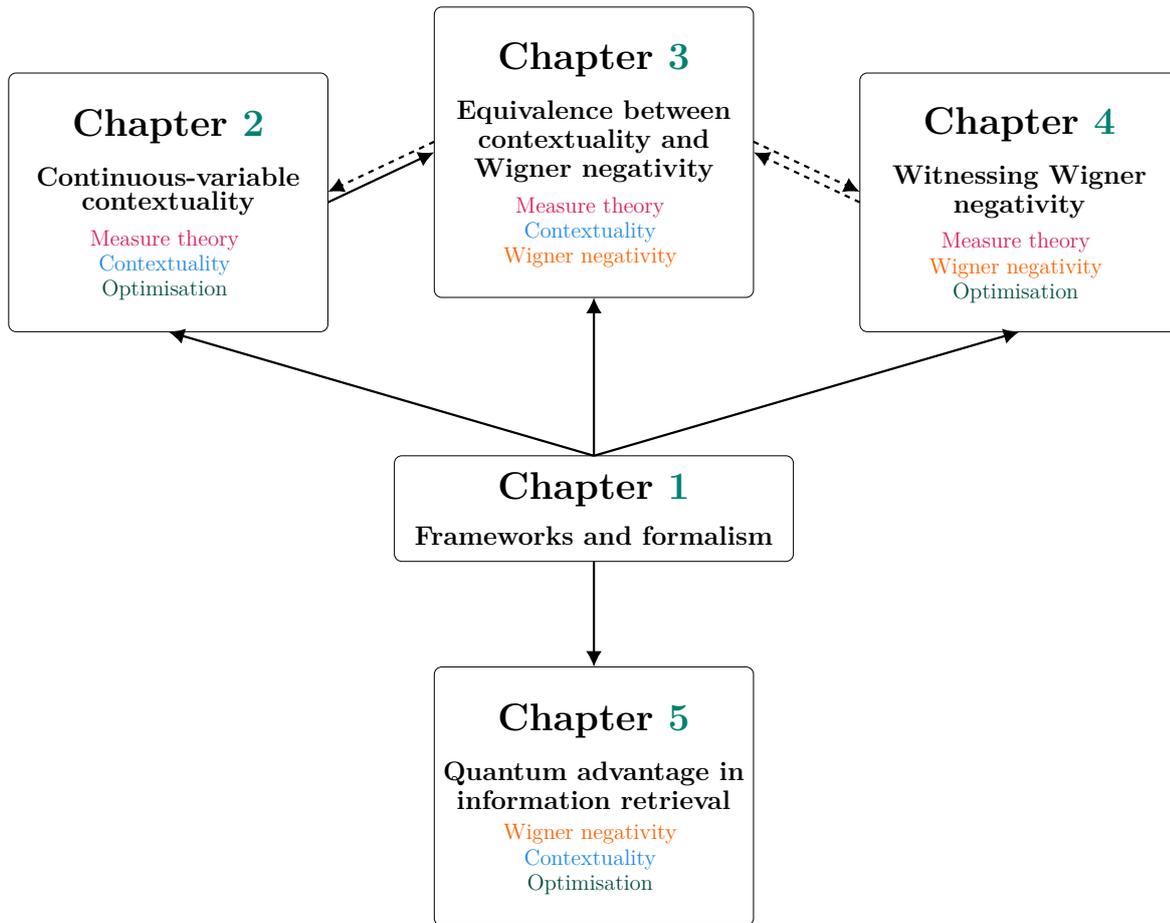


Figure 0.1: Dependencies between the chapters of this dissertation. The black arrows enlighten the direct dependencies between the chapters while the dashed arrows indicate a partial dependence. Key topics in the chapters are indicated below the title of the corresponding chapter and are colour coded. They are defined in Chapter 1. Note that Chapters 2, 3 and 4 use the continuous-variable (or infinite-dimensional) notions while Chapter 5 uses the discrete-variable ones. We did not emphasise the common denominator ‘Quantum Information’ as it is a trivial link throughout the dissertation.

of optimisation theory with a focus on infinite-dimensional linear programming in Section 1.4; and finally, it presents the sheaf theoretic framework for contextuality [AB11] in Section 1.5.

**Chapter 2.** This chapter investigates what happens for measurement contextuality as presented in [AB11]—known as the sheaf theoretic approach to contextuality—when one is dealing with continuous-variable systems. A robust framework for contextuality in continuous-variable scenarios is presented that follows along the lines of the discrete-variable framework introduced by Abramsky and Brandenburger [AB11]. Crucially the Fine–Abramsky–Brandenburger (FAB) theorem [Fin82, AB11] extends to continuous variables. We prove it in scenarios that comprise an uncountable number of measurement labels as it will be essential for the following chapter.

This establishes that noncontextuality of an empirical behaviour, originally characterised by the existence of a deterministic hidden-variable model [Bel64, KS75], can equivalently be characterised by the existence of a factorisable hidden-variable model, and that ultimately both of these are captured by the notion of extendability. An important consequence is that nonlocality may be viewed as a special case of contextuality in continuous-variable scenarios just as for discrete-variable scenarios. The contextual fraction, a quantifiable measure of contextuality that bears a precise relationship to Bell inequality violations and quantum advantages [ABM17], can also be defined in this setting using infinite-dimensional linear programming. It is shown to be a non-increasing monotone with respect to the free operations of a resource theory for contextuality [ABM17, ABKM19]. Crucially, these include the common operation of binning to discretise data which is usually employed in continuous-variable scenarios. A consequence is that any witness of contextuality on discretised empirical data also witnesses and gives a lower bound on genuine continuous-variable contextuality. While infinite linear programs are of theoretical importance and capture exactly the contextual fraction and Bell-like inequalities in which we are interested, they are not directly useful for actual numerical computations. To get around this limitation, we introduce a hierarchy of semidefinite programs [Las09] which are relaxations of the original problem and whose values converge monotonically to the contextual fraction. To use the results from [Las09], we establish that we can reduce the problem to the case of compact outcome spaces. In any case energy bounds of the experimental apparatus provide an argument for this but it is primordial for theoretical considerations where observables might be unbounded.

**Chapter 3.** Negativity of the Wigner function is a striking non-classical feature of quantum states that is related to contextuality for discrete-variable systems [HWVE14, DAGBR15, BVDB<sup>+</sup>17, DOBV<sup>+</sup>17]. It has been widely studied as a resource for quantum speed-up and advantage [Ga05, VFGE12, PWB15, CB18, SHP19]. In this chapter, we extend this result and make a connection between contextuality and Wigner negativity for continuous-variable systems. In particular, we show that contextuality is equivalent to Wigner negativity when we allow the measurement of all possible displacements on 2 or more continuous-variable systems.

**Chapter 4.** Beyond the fundamental relevance of Wigner negativity, it is also a necessary resource for quantum speedup with continuous variables. As quantum technologies emerge, the need to identify and characterise the resources which provide an advantage over existing classical technologies becomes more pressing. It is therefore desirable to detect Wigner negativity. In this chapter, we derive witnesses for Wigner negativity of single-mode and multimode quantum states, based on fidelities with Fock states, which can be reliably measured using standard detection setups. They possess a threshold expectation value indicating whether the measured state has a negative Wigner function. Moreover, the amount of violation provides an operational quantification of Wigner negativity. We phrase the problem of finding the threshold value for a given witness as an infinite-dimensional linear optimisation problem. By relaxing and

restricting the corresponding linear program, we derive two converging hierarchies of semidefinite programs, which provide numerical sequences of increasingly tighter upper and lower bounds for the sought threshold value. We further show that our witnesses form a complete family—each Wigner negative state is detected by at least one witness—thus providing a reliable method for experimentally witnessing Wigner negativity of quantum states from few measurements. From a foundational perspective, our findings provide insights on the set of positive Wigner functions which still lacks a proper characterisation.

**Chapter 5.** Here we focus on discrete-variable quantum information and on expanding the range of applications for which contextuality and negativity can be linked to and used to directly quantify quantum advantage. Random access codes have provided many examples of quantum advantage in communication, but concern only a specific communication game. In this chapter, we investigate a broad generalisation of those that are referred to as information retrieval tasks. We introduce and give a detailed analysis of an information retrieval task—the Torpedo Game—that is distinct from a random access code. We show that it admits a greater quantum advantage than the comparable random access code. Perfect quantum strategies involving prepare-and-measure protocols with experimentally accessible three-level systems [HXG<sup>+</sup>21] emerge via analysis in terms of the discrete Wigner function. The example is leveraged to an operational advantage in a pacifist version of the strategy game *Battleship*. We pinpoint a characteristic of quantum systems that enables quantum advantage in any bounded-memory information retrieval task. While preparation contextuality has previously been linked to advantages in random access coding we focus here on a different characteristic called sequential contextuality [MK18]. It is shown not only to be necessary and sufficient for quantum advantage, but also to quantify the degree of advantage for any information retrieval task.

## Publications

- The publication, pre-prints and work in progress on which this thesis is built are listed below.

[CEG21] “Witnessing Wigner Negativity”, U. Chabaud, P-E. Emeriau, F. Grosshans, **Quantum**, **5:471**, **June 2021**.

[BDE<sup>+</sup>19] “Continuous-variable nonlocality and contextuality”, R. S. Barbosa, T. Douce, P-E. Emeriau, E. Kashefi, S. Mansfield, **arxiv:1905.08267**.

[EHM20] “Quantum Advantage in Information Retrieval”, P-E. Emeriau, M. Howard, S. Mansfield, **arxiv:2007.15643**

[BCE21] “Equivalence between contextuality and Wigner negativity for generalised continuous-variable quadrature measurements”, R. Booth, U. Chabaud, P-E. Emeriau, in preparation.

## Notations

### Sets:

- $\mathbb{N}$  is the set of the natural numbers. For  $n \in \mathbb{N}$ ,  $\llbracket 0, n \rrbracket$  denotes the list of natural numbers between 0 and  $n$ .
- $\mathbb{R}$  is the set of the real numbers.
- $\mathbb{C}$  is the set of the complex numbers. An asterisk  $*$  adjacent to a complex number denotes that we take its complex conjugate.  $\mathbb{S}_1$  is the set of complex numbers of modulus 1.
- $\mathbb{Z}_d$  is the set of integers modulo  $d$  for  $d$  a positive integer.

We add an exponent  $*$  to these above sets when 0 is removed from them. For  $d \in \mathbb{N}^*$  and  $\mathbb{K} \in \{\mathbb{N}, \mathbb{R}, \mathbb{C}, \mathbb{Z}_d\}$ ,  $\mathbb{K}^d$  denotes the Cartesian product with itself  $d$  times. For  $\boldsymbol{\alpha} := (\alpha_i) \in \mathbb{K}^d$ ,  $\mathbb{K} = \mathbb{N}, \mathbb{C}$  or  $\mathbb{R}$ ,  $|\boldsymbol{\alpha}| = \sum_{i=1}^d \alpha_i$ .

The Kronecker delta is the function of two variables over nonnegative integers that assigns 1 if the variables are equal and 0 otherwise. It is denoted  $\delta_{ij}$ .  $\boldsymbol{\delta}_k \in \mathbb{K}^d$  designates the vector filled with 0 except for a 1 at the  $k^{\text{th}}$  coordinate.

- $\mathbb{N}_k^d := \left\{ \boldsymbol{\alpha} \in \mathbb{N}^d \mid |\boldsymbol{\alpha}| \leq k \right\}$  for  $k \in \mathbb{N}$ ,  $d \in \mathbb{N}^*$ .
- $\mathbb{R}[\boldsymbol{x}]$  is the ring of real polynomials in the variables  $\boldsymbol{x} \in \mathbb{R}^d$ .
- $\mathbb{R}[\boldsymbol{x}]_k \subset \mathbb{R}[\boldsymbol{x}]$  is the set of polynomials of total degree at most  $k \in \mathbb{N}$ .
- $\Sigma^2\mathbb{R}[\boldsymbol{x}] \subset \mathbb{R}[\boldsymbol{x}]$  is the set of sum-of-squares polynomials.
- $\Sigma^2\mathbb{R}[\boldsymbol{x}]_k \subset \Sigma^2\mathbb{R}[\boldsymbol{x}]$  is the set of sum-of-squares polynomials of total degree at most  $2k \in \mathbb{N}$ .

Below  $R$  is a subset of  $\mathbb{R}^d$  for some  $d \in \mathbb{N}^*$ . We implicitly consider real-valued functions.

- $C(R)$  is the space of continuous functions on  $R$ .
- $C_0(R)$  is the space of continuous functions on  $R$  that vanish at infinity in the case where  $R$  is not bounded.
- $C^\infty(R)$  is the space of smooth (infinitely-differentiable) functions over  $R$ .
- $L^1(R)$  is the space of integrable functions over  $R$ .
- $L^2(R)$  is the space of square-integrable functions over  $R$ .
- $L^2'(R)$  is the dual space  $L^2(R)$  over  $R$  which is isomorphic to  $L^2(R)$ .
- $\ell^2$  is the space of square-summable real sequences.
- $\mathcal{S}(R)$  is the space of Schwartz functions over  $R$  (the space of  $C^\infty$  functions that go to 0 at infinity faster than any inverse polynomial, as do their derivatives) [Sch47].
- $\mathcal{S}'(R)$  is the dual space of  $\mathcal{S}(R)$  which is the space of tempered distributions over  $R$  (i.e. distributions that have at most a polynomial growth at infinity).
- $\mathcal{S}(\mathbb{N})$  is the space of rapidly decreasing real sequences (i.e. that go to 0 at infinity faster than any inverse polynomial).

- $\mathcal{S}'(\mathbb{N})$  is the dual space of  $\mathcal{S}(\mathbb{N})$  which is the space of slowly increasing real sequences (i.e. that are upper bounded by a polynomial).

We add a subscript  $+$  when we take the nonnegative elements of all of the above sets (e.g.  $\mathbb{R}_+$  denotes the nonnegative reals,  $S_+(R)$  denotes the nonnegative Schwartz functions over  $R$ , etc).

- $\text{Mat}_m(\mathbb{R})$  is the space of  $m \times m$  real matrices for  $m \in \mathbb{N}^*$ .
- $\text{Sym}_m$  is the set of  $m \times m$  real symmetric matrices for  $m \in \mathbb{N}^*$ .

An exponent  $T$  on a matrix denotes its transpose while an exponent  $\dagger$  denotes its conjugate transpose. The symbol  $\succeq$  stands for semidefinite positiveness.

- $\mathbf{U} = \langle U, \mathcal{F}_U \rangle$  denotes a generic measurable space composed of a set  $U$  and a  $\sigma$ -algebra  $\mathcal{F}_U$  on  $U$ . We will use the convention of using bold font to refer to the measurable space and regular font to refer to the underlying set.
- $\mathbb{M}(\mathbf{U})$  is the set of measures on  $\mathbf{U}$ .
- $\mathbb{M}_{\pm}(\mathbf{U})$  is the set of finite-signed measures on  $\mathbf{U}$  (i.e. finite measures that are non necessarily nonnegative).
- $\mathbb{P}(\mathbf{U})$  is the set of probability measures on  $\mathbf{U}$ .
- $\mathcal{P}(X)$  denotes the powerset of a set  $X$ .
- $V \setminus U = V - U = \{x \in V | x \notin U\}$  denotes the difference of two sets  $U$  and  $V$  such that  $U \subseteq V$ .
- for  $U \subseteq V$ ,  $U^c$  denotes the complement of  $U$  in  $V$ .

For a measurable space  $\mathbf{U}$ ,  $E \in \mathcal{F}_U$ ,  $\mathbf{1}_E$  (resp.  $\mathbf{0}_E$ ) denotes the function that assigns 1 (resp. 0) to all elements of  $E$ .  $\mathbf{1}_E$  is the indicator function on  $E$ . We use  $\cong$  between two sets to mean they are isomorphic. For a vector space  $V$  over a field  $F$ ,  $V^*$  denotes the set of linear functions  $V \rightarrow F$ .<sup>3</sup>

For  $k \in \mathbb{N}$ , the  $k^{\text{th}}$  Laguerre function is  $\mathcal{L}_k(x): x \in \mathbb{R}_+ \rightarrow (-1)^k L_k(x) e^{-\frac{x}{2}}$  with  $L_k(x) := \sum_{l=0}^k \frac{(-1)^l}{l!} \binom{k}{l} x^l$  the  $k^{\text{th}}$  Laguerre polynomial.

## Quantum Information

$\mathcal{H}$  denotes a Hilbert space. When not explicit from the context,  $\mathcal{H}$  is a generic separable infinite-dimensional Hilbert space.  $\mathcal{D}(\mathcal{H})$  is the set of quantum states (i.e. positive semidefinite operators with unit trace) over  $\mathcal{H}$ .  $|\psi\rangle \in \mathcal{D}(\mathcal{H})$  represents a generic pure state of  $\mathcal{H}$  and  $\rho \in \mathcal{D}(\mathcal{H})$  a generic density operator. The identity operator is denoted  $\mathbb{1}$  ( $\mathbb{1}_n$  for  $n \in \mathbb{N}^*$  is the usual  $n \times n$  identity matrix) and the usual qubit Pauli operators are denoted  $X$ ,  $Y$  and  $Z$ .  $X$  and  $Z$  also designate the generalisation of their qubit counterpart to the qudit realm. It will be explicit from the context.

---

<sup>3</sup>This is the same notation than the one used to remove 0 from a set but it will be clear from context which notation we mean.

## Contextuality

From an operational perspective, an experimental setup can be formalised as a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  where  $\mathcal{X}$  is the set of measurement labels,  $\mathcal{M}$  the set of maximal contexts,  $\mathcal{O} = (\mathcal{O}_x)_{x \in \mathcal{X}}$  is the collection of outcome sets for each measurement. Empirical models will be usually denoted by  $e$  and  $\text{CF}(e)$  designates its contextual fraction while  $\text{NCF}(e) = 1 - \text{CF}(e)$  designates its noncontextual fraction.

## Multi-index notation

For  $M \in \mathbb{N}^*$ :

$$\begin{aligned}
 \mathbf{0} &= (0, \dots, 0) \in \mathbb{N}^M \\
 \mathbf{1} &= (1, \dots, 1) \in \mathbb{N}^M \\
 m\mathbf{1} &= (m, \dots, m) \\
 m\mathbf{k} &= (mk_1, \dots, mk_M) \\
 \pi_{\mathbf{k}} &= \prod_{i=1}^M (k_i + 1) \\
 s(m) &= \binom{M+m}{m} \\
 \hat{D}(\boldsymbol{\alpha}) &= \hat{D}(\alpha_1) \otimes \dots \otimes \hat{D}(\alpha_M) \\
 |\mathbf{k}\rangle &= |k_1\rangle \otimes \dots \otimes |k_M\rangle \\
 \langle \mathbf{k}| &= \langle k_1| \otimes \dots \otimes \langle k_M| \\
 \mathbf{k} \leq \mathbf{n} &\Leftrightarrow k_i \leq n_i \quad \forall i = 1, \dots, M \\
 L_{\mathbf{k}}(\boldsymbol{\alpha}) &= L_{k_1}(\alpha_1) \dots L_{k_M}(\alpha_M) \\
 \mathcal{L}_{\mathbf{k}}(\boldsymbol{\alpha}) &= \mathcal{L}_{k_1}(\alpha_1) \dots \mathcal{L}_{k_M}(\alpha_M) \\
 |\mathbf{k}| &= k_1 + \dots + k_M \\
 \boldsymbol{\alpha}^{\mathbf{k}} &= \alpha_1^{k_1} \dots \alpha_M^{k_M} \\
 \mathbf{k}! &= k_1! \dots k_M! \\
 \binom{\mathbf{n}}{\mathbf{k}} &= \binom{n_1}{k_1} \dots \binom{n_M}{k_M} \\
 \mathbf{k} + \mathbf{n} &= (k_1 + n_1, \dots, k_M + n_M) \\
 e^{\boldsymbol{\alpha}} &= e^{\alpha_1} \dots e^{\alpha_M}.
 \end{aligned} \tag{1}$$

# FRAMEWORKS AND FORMALISM

**I**N this chapter, we introduce the frameworks and formalism on which this dissertation is built. Section 1.1 introduces the basic notions of Quantum Information; Section 1.2 gives the necessary background for phase-space formalism and culminates with the expression of the Wigner function; Section 1.3 provides the necessary notions of measure theory; Section 1.4 provides the tools for (infinite-dimensional) optimisation theory; and Section 1.5 introduces the sheaf theoretic approach for contextuality.

## 1.1 Quantum information preliminaries

### 1.1.1 Basics of quantum information theory

We give a brief and pedestrian introduction to the basics of Quantum Information. We refer readers to [NC11] for a more extensive introduction.

Quantum information is the field that describes the behaviour of information when encoded in degrees of freedom of particles governed by the laws of quantum physics. The properties of a quantum system are represented by its quantum state. Formally, *pure* quantum states correspond to normalised state vectors in a separable Hilbert space  $\mathcal{H}$ . In quantum physics the Dirac notation is ubiquitous: a ket  $|\psi\rangle$  is a vector from  $\mathcal{H}$  which corresponds to a *pure* quantum state; a bra  $\langle\phi|$  is a linear form  $\phi: \mathcal{H} \rightarrow \mathbb{C}$ . The action of a linear form on a vector  $\langle\phi|(|\psi\rangle)$  gives a scalar denoted  $\langle\phi|\psi\rangle$ . It can be used to construct the projection onto the state  $|\psi\rangle$  as  $P_\psi: |\phi\rangle \in \mathcal{H} \mapsto \langle\psi|\phi\rangle |\psi\rangle \in \mathcal{H}$  or more concisely  $|\psi\rangle\langle\psi|$ .

From an operational point of view, a quantum experiment can be decomposed into three distinct steps: *preparation*, *evolution* and *measurement*. Note that the evolution could always be taken into account in the measurement or the preparation though it is often easier to consider

measurements or preparations in a fixed basis. Preparation might be used to encode classical information; evolution to manipulate quantum information; measurements to extract classical information again.

**Quantum states.** The simplest example is the two-dimensional Hilbert space  $\mathbb{C}^2$  where information is encoded in the so-called quantum bits or *qubits*. Given an orthonormal basis  $\{|0\rangle, |1\rangle\}$  of  $\mathbb{C}^2$ , a pure qubit state can be expressed as:

$$|\psi\rangle = a|0\rangle + b|1\rangle = \begin{pmatrix} a \\ b \end{pmatrix}, \quad (1.1)$$

where  $a, b \in \mathbb{C}$  with the normalisation condition that  $|a|^2 + |b|^2 = 1$ . Physically, it could for instance model the spin of an electron for which the two possible basis states are the spin pointing up and down; or the polarisation of a single photon where the two basis states are vertical and horizontal polarisation. Surprisingly, quantum physics allows for superposition, that is, it allows for complex linear combinations. That is why quantum theory is inherently probabilistic. Quantum randomness occurs when measuring a quantum system in superposition: in general, it is impossible to predict deterministically the outcome of such a measurement. This is not because of a lack of knowledge of the system but rather due to the probabilistic nature<sup>1</sup> of quantum theory. When measuring the state defined in Eq. (1.1) in the computational basis  $\{|0\rangle, |1\rangle\}$ , we will witness outcome 0 associated to  $|0\rangle$  with probability  $|a|^2$  and outcome 1 associated to  $|1\rangle$  with probability  $|b|^2$  according to the Born rule that we detail shortly after.

Quantum systems may also exhibit classical randomness. In that case, we will say that the quantum state is *mixed* rather than *pure*. Mathematically, it is represented by a semidefinite, Hermitian operator with unit trace acting on  $\mathcal{H}$  called a density matrix or density operator. We write  $\mathcal{D}(\mathcal{H})$  the set of density operators. Any density operator can be written as a convex combination of pure states. For instance, the mixed state  $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$  known as the maximally mixed state is classically probabilistic: it corresponds to the state obtained by flipping an unbiased coin; it is different from the pure superposition  $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$  whose density matrix is  $\frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|0\rangle\langle 1| + \frac{1}{2}|1\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$ .

To describe the state of a global system consisting of two independent subsystems  $|\psi_1\rangle$  and  $|\psi_2\rangle$  lying in respectively Hilbert spaces  $\mathcal{H}_1$  and  $\mathcal{H}_2$ , we will use a tensor product structure. We write the global state as  $|\psi_1\rangle \otimes |\psi_2\rangle = |\psi_1\psi_2\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ . However the two subsystems might not be independent. A pure quantum state of several particles which cannot be written as a tensor product of quantum states is said to be *entangled*. As an example, the state  $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$  is entangled while the state  $\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$  is separable. For ensembles, a mixed state is separable when it can be written as a mixture of separable pure states and entangled otherwise. Measurement on entangled states may give rise to correlations

---

<sup>1</sup>or the incompleteness...

unachievable with classical systems. For a composite system  $\rho_{AB} \in \mathcal{H}_A \otimes \mathcal{H}_B$ , the reduced state of the first subsystem is obtained by tracing out—that is taking the partial trace—over the second subsystem. It is described by  $\text{Tr}_{\mathcal{H}_B}(\rho)$ . Partial trace is the unique operator that describes part of a larger quantum system with the correct description for subsystems of the composite system [NC11].

The *fidelity*<sup>2</sup> between two arbitrary quantum states  $\rho$  and  $\sigma$  is defined as

$$F(\rho, \sigma) := \text{Tr}(\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}})^2. \quad (1.2)$$

When one of the states is pure, it reduces to  $F(\rho, \sigma) = \text{Tr}(\rho\sigma)$ . The *trace distance* between two quantum states  $\rho$  and  $\sigma$  is defined as

$$D(\rho, \sigma) := \frac{1}{2}\text{Tr}(\sqrt{(\rho - \sigma)^2}). \quad (1.3)$$

The trace distance can be related to the maximum probability of distinguishing between two quantum states. Fidelity and trace distance are related by  $1 - F \leq D \leq \sqrt{1 - F}$  [NC11].

**Evolution.** The simplest evolution of a closed quantum system is described by a unitary transformation with a unitary operator  $\hat{U}$  such that  $\hat{U}\hat{U}^\dagger = \mathbb{1}$ . It is generated by a Hermitian operator  $\hat{H}$ —often called Hamiltonian—such that  $\hat{U} = e^{i\hat{H}}$ . The unitary evolution of a pure state  $|\psi\rangle$  results in a pure state  $\hat{U}|\psi\rangle$  while for a mixed state  $\rho$ , it results in a mixed state  $\hat{U}\rho\hat{U}^\dagger$ . Important 2-dimensional examples are given by the Pauli X gate, the Pauli Z gate and the Pauli Y gate<sup>3</sup>:

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \quad (1.4)$$

Other important single qubit operations are the Hadamard gate  $H$  and the  $\frac{\pi}{8}$ -gate  $T = \sqrt{S}$ :

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\frac{\pi}{4}} \end{pmatrix}. \quad (1.5)$$

As an example, applying X on the state given by Eq. (1.1) yields:

$$X|\psi\rangle = b|0\rangle + a|1\rangle. \quad (1.6)$$

More general evolutions are given by quantum channels known as completely positive trace-preserving (CPTP) maps. Stinespring's dilation theorem guarantees that CPTP maps can ultimately be lifted to unitary operations [Sti55] on a higher dimensional Hilbert space. This can be seen as a purification of CPTP maps.

<sup>2</sup>Note that in [NC11], the fidelity is defined as the square root of Eq. (1.2). There is no fundamental difference between the two definitions.

<sup>3</sup>From a computational point of view, they correspond to the more intuitive errors that may occur: respectively *bit flip*, *phase flip* or both.

**Measurement.** Following the approach of von Neumann, a measurement is represented by a self-adjoint operator known as an *observable*  $\hat{O} = \hat{O}^\dagger$ . An observable has a family of eigenvectors that form an orthonormal basis for the Hilbert space. Measuring the observable yields an outcome that corresponds to one of the eigenvalues associated to the eigenvectors comprising the basis. The probability of obtaining each outcome follows from the Born rule. When the state of the system is a pure state  $|\psi\rangle$  it reads:

$$\Pr(\lambda) = \langle \psi | \Pi_\lambda | \psi \rangle, \quad (1.7)$$

where  $\lambda$  is an eigenvalue of  $\hat{O}$  and  $\Pi_\lambda$  is the projector onto the eigenspace<sup>4</sup> associated to  $\lambda$ . For qubits, Pauli operators are Hermitian and can thus be measured. As already mentioned, measuring  $Z$  on the state given by Eq. (1.1) yields two outcomes: 0 with probability  $\langle \psi | \Pi_0 | \psi \rangle = \langle \psi | 0 \rangle \langle 0 | \psi \rangle = |\langle 0 | \psi \rangle|^2 = |a|^2$ ; and 1 with probability  $\langle \psi | \Pi_1 | \psi \rangle = |b|^2$ . For a mixed state  $\rho$ , the Born rule reads:

$$\Pr(\lambda) = \text{Tr}(\Pi_\lambda \rho). \quad (1.8)$$

By linearity, the expectation value of  $\hat{O}$  can be expressed as:

$$\langle \hat{O} \rangle = \text{Tr}(\hat{O} \rho). \quad (1.9)$$

This procedure corresponds to a projection-valued measurement (PVM).

PVMs are generalised by positive operator-valued measurements (POVMs) in the same sense that mixed states generalise the notion of pure states. A POVM is a set of positive semidefinite matrices  $\{F_i\}_{i \in \mathcal{I}}$  that sum to the identity operator where  $\mathcal{I}$  denotes the set of outcomes. When the state of the system is  $\rho$ , the outcome  $i \in \mathcal{I}$  is obtained with probability:

$$\Pr(i) = \text{Tr}(F_i \rho). \quad (1.10)$$

A POVM reduces to a PVM when  $\{F_i\}_{i \in \mathcal{I}}$  are orthogonal projectors.

## 1.2 Phase-space formalism

We now turn our attention to the phase-space formulation of quantum physics, motivate its use and introduce the so-called Wigner function. We refer readers to a recent and comprehensive overview [RE21] for further details. This formulation is one of many possible alternative formulations of quantum physics, e.g. Schrödinger's wave formulation or Heisenberg, Born and Jordan's matrix mechanics both from 1925. The phase-space formulation is often favoured in quantum optics applications. The Wigner function was first introduced by Wigner in 1932 [Wig32] as the Fourier transform of the spatial autocorrelation wavefunction. It can be thought as the joint (quasi)probability distribution of position and momentum. The marginals of the

<sup>4</sup>Or more simply the eigenvector if the observable at hand is not degenerate.

Wigner function reproduce the correct probability distribution of quadratures<sup>5</sup> i.e. integrating the Wigner function over position at a given momentum gives the correct probability associated with the chosen momentum and vice versa. Crucially, it is a quasiprobability distribution in phase-space: it sums to unity though one has to drop the nonnegativity requirement of probability distributions. This latter requirement is critical in quantum computing applications. We will see how Wigner functions emerge as the expectation value of a kernel<sup>6</sup> for finite-dimensional as well as infinite-dimensional systems. The phase-space formulation comes with a remarkable symplectic structure [ARL14].

### 1.2.1 Phase-space for $d$ -dimensional quantum states

It is possible to represent finite-dimensional quantum states as quasi-probability distributions over a phase space of discrete points. Fix  $d \in \mathbb{N}$  an odd power-of-prime. Wootters [GHW04, Woo87] introduced a method for constructing discrete Wigner functions (DWF) based on finite fields, wherein vectors from a complete set of mutually unbiased bases in  $\mathbb{C}^d$  are put in one-to-one correspondence with the lines of a finite affine plane of order  $d$ . Two orthonormal bases are said to be mutually unbiased [Sch60] if the square of the inner product between an element of the first basis and an element of the second basis equals the inverse of  $d$ . They are ‘unbiased’ in the following sense: if a system is prepared according to a basis state of one of the basis, all the outcomes of the measurement with respect to the other basis are predicted to occur with the same probability  $\frac{1}{d}$ .

Gross [Gro06] singled out one particularly symmetric definition of DWF that obeys the discrete version of Hudson’s Theorem [Hud74] that we recall after introducing the Wigner function. We introduce the necessary tools for this construction below.

We consider a  $d$ -dimensional Hilbert space  $\mathcal{H}$  equipped with the computational basis  $\{|0\rangle, |1\rangle, \dots, |d-1\rangle\}$ . Below all arithmetic will be modulo  $d$ . The phase-space of a single  $d$ -level system is  $V = \mathbb{Z}_d \times \mathbb{Z}_d$ —which can be thought of as an affine plane, or more precisely a toric  $d \times d$  grid. For  $m$  particles, it can be straightforwardly extended by associating a phase-space to each particle. We extend the definition of Pauli operators  $X$  (translation) and  $Z$  (boost) to  $\mathcal{H}$  as follows: for all  $k \in \mathbb{Z}_d$ ,

$$\begin{aligned} X|k\rangle &= |k+1\rangle \\ Z|k\rangle &= \omega^k |k\rangle, \end{aligned} \tag{1.11}$$

with  $\omega = \exp(i\frac{2\pi}{d})$ . Note that, unlike the qubit case, these are not Hermitian operators anymore. Strictly speaking they should not correspond to physical observational quantities. However their non-degenerate eigenstates still form a complete orthonormal basis of the Hilbert space; thus they can be used to define a von Neumann measurement. The qudit *Pauli group* is comprised

<sup>5</sup>Position  $\hat{q}$  and momentum  $\hat{p}$  are known as quadrature operators since they correspond to two ‘quadratures’ of the electromagnetic field. See Subsection 1.2.2.

<sup>6</sup>Kernel is rather versatile in mathematics. Here, by kernel, we mean *integral kernel*.

of *Weyl operators* (also known as generalised Pauli operators) which are products of powers of these operators, e.g.  $X^q Z^p$  for  $(q, p) \in V$ . The *Clifford group* is defined as containing the unitaries which map Weyl operators to other Weyl operators under conjugation. A unitary  $U$  stabilises a state  $|\psi\rangle$  if  $U|\psi\rangle = |\psi\rangle$ . A *stabiliser state* is the unique  $n$ -qudit state stabilised by a subgroup of size  $d^n$  of the Pauli group. Equivalently, stabiliser states may be understood as the image of computational basis states under the Clifford group.

We introduce the *displacement operator* at point  $(q, p) \in V$ :

$$\hat{D}(q, p) := \omega^{2^{-1}qp} X^q Z^p. \quad (1.12)$$

These are known as are Gross' Weyl operators [Gro06].  $2^{-1}$  is the multiplicative inverse of 2 in  $\mathbb{Z}_d$ , that is  $2^{-1} = (d+1)/2$ .

The *parity operator*—also known as the phase point operator at the origin—reads:

$$\hat{\Pi} := \sum_{k \in \mathbb{Z}_d} |-k\rangle\langle k|. \quad (1.13)$$

Parity transformation—or parity inversion—is essentially a flip in the sign of spatial coordinates. It is physically motivated by its continuous-variable analogue though we indeed have that  $\hat{\Pi}|k\rangle = |-k \bmod d\rangle$  for  $k \in \mathbb{Z}_d$ .

### 1.2.2 The case of continuous-variable quantum information

Before giving the expression of the Wigner function compatible with discrete and continuous systems we need to introduce some elements of quantum information with continuous variables. We refer readers to [ARL14, BVL05, Wal21] for extensive introductions to the material presented here.

In the continuous-variable (CV) formalism, information is encoded in a mode (or qumode) rather than in a finite dimensional system like a qubit or a qudit. The separable Hilbert space  $\mathcal{H}$  is infinite-dimensional and is equipped with a countable orthonormal basis  $\{|n\rangle\}_{n \in \mathbb{N}}$  known as the Fock basis, or photon number basis in an optical context. The phase-space associated to a single mode is now  $V = \mathbb{R}^2 \cong \mathbb{C}$  and it can be straightforwardly extended to several modes. A single-mode pure state can be expanded in the Fock basis as:

$$|\psi\rangle = \sum_{n \in \mathbb{N}} \psi_n |n\rangle \quad (1.14)$$

where for all  $n \in \mathbb{N}$ ,  $\psi_n \in \mathbb{C}$  such that  $\sum_n |\psi_n|^2 = 1$ . A general mixed state can be expressed as:

$$\rho = \sum_{k, l \in \mathbb{N}} \rho_{k, l} |k\rangle\langle l|, \quad (1.15)$$

where for all  $k, l \in \mathbb{N}$ ,  $\rho_{k, l} \in \mathbb{C}$ ,  $\rho_{k, l} = \rho_{l, k}^*$  and  $\sum_k \rho_{k, k} = 1$ .

Ladder operators are ubiquitous in CV quantum information. Canonical adjoint operators  $\hat{a}$  and  $\hat{a}^\dagger$ , known as the annihilation and creation operators, are defined by their action on the Fock basis as:

$$\begin{aligned}\hat{a}|n\rangle &= \sqrt{n}|n-1\rangle \quad \text{for } n \in \mathbb{N}^*, \\ \hat{a}|0\rangle &= 0, \\ \hat{a}^\dagger|n\rangle &= \sqrt{n+1}|n+1\rangle \quad \text{for } n \in \mathbb{N}.\end{aligned}\tag{1.16}$$

These operators obey the commutation relation:

$$[\hat{a}, \hat{a}^\dagger] = \mathbb{1}.\tag{1.17}$$

Note that they do not correspond naturally to physical observations since they are not Hermitian operators. However, from them, we can define well-known observables. Position and momentum operators—also known as the quadrature operators—are linked to these ladder operators as follows:

$$\begin{aligned}\hat{q} &= \frac{1}{\sqrt{2}}(\hat{a} + \hat{a}^\dagger), \\ \hat{p} &= \frac{1}{i\sqrt{2}}(\hat{a} - \hat{a}^\dagger).\end{aligned}\tag{1.18}$$

They obey the canonical commutation relation

$$[\hat{q}, \hat{p}] = i\mathbb{1},\tag{1.19}$$

which is one expression of the Heisenberg uncertainty principle [Hei85] i.e. that position and momentum cannot be simultaneously measured on the same state with arbitrary precision.

The eigenstates of  $\hat{a}$  are the coherent states  $\{|\alpha\rangle\}_{\alpha \in \mathbb{C}}$ :

$$|\alpha\rangle := e^{-\frac{1}{2}|\alpha|^2} \sum_{n \in \mathbb{N}} \frac{\alpha^n}{\sqrt{n!}} |n\rangle.\tag{1.20}$$

They define an overcomplete basis of  $\mathcal{H}$ . A central operator in phase-space formulation is the *displacement operator* by a value  $\alpha \in \mathbb{C}$  [WPGP<sup>+</sup>12] which reads

$$\hat{D}(\alpha) := e^{\alpha\hat{a}^\dagger - \alpha^*\hat{a}},\tag{1.21}$$

or similarly for  $\alpha = \frac{1}{\sqrt{2}}(q + ip)$  with  $(q, p) \in V$ :

$$\hat{D}(q, p) := e^{ip\hat{q} - iq\hat{p}} = e^{i\frac{qp}{2}} e^{-iq\hat{p}} e^{ip\hat{q}}.\tag{1.22}$$

Compared to the discrete-variable case in Eq. (1.12), this amounts to take the phase factor as  $\omega = e^i$  with the continuous-variable generalisation of the X gate as  $X(q) := e^{-iq\hat{p}}$  and of the Z gate as  $Z(p) := e^{ip\hat{q}}$ . This gives that  $\hat{D}(q, p) = e^{i\frac{qp}{2}} X(q)Z(p)$ .

The coherent state of amplitude  $\alpha \in \mathbb{C}$  can be obtained by displacing the vacuum state  $|0\rangle$  by an amount  $\alpha$ :

$$|\alpha\rangle = \hat{D}(\alpha)|0\rangle.\tag{1.23}$$

The coefficients of the displacement operator in Fock basis are given by [Wün98]:

$$\langle k|\hat{D}(\alpha)|l\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{p=0}^{\min k,l} \frac{\sqrt{k!l!}(-1)^{l-p}}{p!(k-p)!(l-p)!} \alpha^{k-p} \alpha^{*l-p}, \quad (1.24)$$

for all  $k, l \in \mathbb{N}$  and all  $\alpha \in \mathbb{C}$ .

Lastly, the *parity operator* is defined as:

$$\hat{\Pi} := (-1)^{\hat{a}^\dagger \hat{a}} = e^{i\pi \hat{a}^\dagger \hat{a}} = \sum_{n \in \mathbb{N}} (-1)^n |n\rangle \langle n|. \quad (1.25)$$

On a particle at position  $q \in \mathbb{R}$ , it acts as  $\hat{\Pi}\psi(q) = \psi(-q)$  where  $\psi(q) = \langle q|\psi\rangle$  and  $|q\rangle$  is an eigenstate of  $\hat{q}$ .

### 1.2.3 The Wigner function

We fix  $M \in \mathbb{N}^*$  the number of modes. The phase-space is  $V^M$  and it is endowed with its vector space structure and its canonical basis. As mentioned before, the *Wigner function* is a quasiprobability distribution [CG69] on  $V^M$  that can be expressed as the expectation value of a kernel—the displaced parity operator. For an operator  $\hat{O}$  acting on  $M$  discrete-variable or continuous-variable systems (see Table 1.1 for the comparison), the Wigner function at point  $\mathbf{r} = (q_1, \dots, q_M, p_1, \dots, p_M) \in V^M$  reads:

$$W_{\hat{O}}(\mathbf{r}) := \left(\frac{2}{c}\right)^M \text{Tr}(\hat{D}(\mathbf{r})\Pi^{\otimes M}\hat{D}(\mathbf{r})^\dagger\hat{O}) \quad (1.26)$$

where  $c = 2\pi$  for continuous-variable systems and  $c = 2d$  for a  $d$ -dimensional system and where  $\hat{D}(\mathbf{r}) = \hat{D}(q_1, p_1) \otimes \dots \otimes \hat{D}(q_M, p_M)$ .

	Discrete variables	Continuous variables
Phase-space $V^M$	$(\mathbb{Z}_d^2)^M$	$(\mathbb{R}^2)^M$
Phase factor $\omega$	$e^{i\frac{2\pi}{d}}$	$e^i$
Displacement at $(q, p) \in V^M$	$e^{\frac{2i\pi}{d}2^{-1}qp}\mathbf{X}^q\mathbf{Z}^p$	$e^{i\frac{qp}{2}}\mathbf{X}(q)\mathbf{Z}(p)$
Parity operator	$\sum_{k \in \mathbb{Z}_d}  -k\rangle \langle k $	$\sum_{k \in \mathbb{N}} (-1)^k  k\rangle \langle k $
Constant $c$	$2d$	$2\pi$

Table 1.1: Comparison between quantities comprising the Wigner function of  $m$  systems expressed in the discrete-variable setting of odd prime dimension  $d$  and in the continuous-variable setting.

A useful relation is the link between the Wigner function and the characteristic function that is derived using the symplectic structure of the phase-space  $V^M$ . The *characteristic function* of a density operator  $\rho$  is defined as:

$$\begin{aligned} \Phi_\rho : V^M &\longrightarrow \mathbb{R} \\ \mathbf{r} &\longmapsto \text{Tr}(\hat{D}(\mathbf{r})\rho) \end{aligned} \quad (1.27)$$

Due to the commutation relation between position and momentum operators there is a symplectic structure for  $V^M$ . We denote  $\Omega$  the symplectic form. For  $\mathbf{r}_1, \mathbf{r}_2 \in V^M$ ,

$$\Omega(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{r}_1 \cdot J\mathbf{r}_2 \quad \text{where} \quad J = \begin{pmatrix} 0 & \mathbb{1}_M \\ -\mathbb{1}_M & 0 \end{pmatrix} \quad (1.28)$$

We have that  $J^{-1} = J^T = -J$ . Then the Wigner function is the symplectic Fourier transform of the characteristic function:

$$W_\rho(\mathbf{r}) = \text{FT}[\Phi_\rho](-J\mathbf{r}) \quad (1.29)$$

where the Fourier transform of a  $L^1(V^M)$  function  $f$  is defined as (see table 1.1):

$$\text{FT}[f](\mathbf{r}) := \begin{cases} \sum_{\mathbf{p} \in V^M} (\omega^*)^{\mathbf{r} \cdot \mathbf{p}} f(\mathbf{p}) & \text{for discrete-variable systems,} \\ \int_{V^M} (\omega^*)^{\mathbf{r} \cdot \mathbf{p}} f(\mathbf{p}) \, d\mathbf{p} & \text{for continuous-variable systems,} \end{cases} \quad (1.30)$$

for  $\mathbf{r} \in V^M$ . The inverse Fourier transform is denoted  $\text{FT}^{-1}$  and it is defined as:

$$\text{FT}^{-1}[f](\mathbf{p}) := \begin{cases} \frac{1}{c^M} \sum_{\mathbf{r} \in V^M} \omega^{\mathbf{r} \cdot \mathbf{p}} f(\mathbf{r}) & \text{for discrete-variable systems,} \\ \frac{1}{c^M} \int_{V^M} \omega^{\mathbf{r} \cdot \mathbf{p}} f(\mathbf{r}) \, d\mathbf{r} & \text{for continuous-variable systems.} \end{cases} \quad (1.31)$$

Now we turn our attention to the celebrated *Hudson's theorem* [Hud74]. For continuous-variable systems, Hudson's theorem establishes that a pure state which has a nonnegative Wigner function is necessarily a Gaussian state (i.e. a state whose Wigner function is a Gaussian distribution). Its discrete-variable counterpart [Gro06] says that an odd-dimensional pure state is nonnegatively represented in the DWF if and only if it is a stabiliser state. Hudson's theorem has remarkable implications, providing large classes of quantum circuit with a local hidden variable model that enables efficient simulation [Ga05, CGaG<sup>+</sup>06, VFGE12, ME12].

The Wigner function is a quasiprobability distribution: a normalised distribution (i.e. its sum or integral over phase space equals 1) which can take negative values. Whenever it is everywhere nonnegative, it can be interpreted as a kind of probability distribution for the quadrature measurements. Crucially, Wigner negativity—the fact that the Wigner distribution can be negative in regions of phase space—is a necessary resource for quantum computational speedup. Informally, a result from Mari and Eisert [ME12] states that if the quantum computation (i) starts in a product state with a nonnegative Wigner function, (ii) is followed by quantum gates that can be represented by a nonnegative Wigner function (defined through the Choi matrix), (iii) and terminates with measurements associated with a nonnegative Wigner function then there exists an efficient classical algorithm that simulates the computation (assuming that local probability distributions may be sampled efficiently). This essentially amounts to sampling directly from the Wigner distribution which, in this case, is a proper probability distribution. This results applies in the DV and CV settings.

From a resource theoretic point of view, negativity of the Wigner function can only decrease under Gaussian operations [AGPF18], i.e., operations that map Gaussian states to Gaussian states. In particular, it is invariant under displacements. It is also a robust property, since two almost indistinguishable quantum states have similar Wigner functions. An operational measure of Wigner negativity for a quantum state  $\rho \in \mathcal{D}(\mathcal{H})$  is given by its distance to the set of states having a positive Wigner function [MKN<sup>+</sup>11]:

$$\eta_\rho = \inf_{\substack{\sigma \in \mathcal{D}(\mathcal{H}) \\ W_\sigma \geq 0}} D(\rho, \sigma), \quad (1.32)$$

where  $D$  denotes the trace distance (see Subsection 1.1.1). It quantifies the operational distinguishability between the state  $\rho$  and any state having a positive Wigner function [NC11].

To fix ideas, the Wigner functions of Fock states  $\{|k\rangle\}_k$  are given by [KŽ04]:

$$\forall \alpha \in \mathbb{C}, W_k(\alpha) = \frac{2}{\pi} \mathcal{L}_k(4|\alpha|^2), \quad (1.33)$$

where  $\mathcal{L}_k$  is the  $k^{\text{th}}$  Laguerre function [Sze59] defined as

$$\forall x \in \mathbb{R}_+, \mathcal{L}_k(x) := (-1)^k L_k(x) e^{\frac{x}{2}} \quad (1.34)$$

with  $L_k$  the  $k^{\text{th}}$  Laguerre polynomial defined as

$$L_k(x) := \sum_{l=0}^k \frac{(-1)^l}{l!} \binom{k}{l} x^l. \quad (1.35)$$

Except for the vacuum state  $|0\rangle$ , all other Fock states have a Wigner function with a negative part.

### 1.3 Measure theory in a nutshell

Introduced below are the necessary tools of measure theory that are used throughout this dissertation. We refer readers to [Bil08, Tao11] for a more thorough introduction to measure theory.

To avoid some pathological behaviours when dealing with probability distributions on a continuum, we first need to define  $\sigma$ -algebras which will give rise to a good notion of measurability.

**Definition 1.1** ( $\sigma$ -algebras). *A  $\sigma$ -algebra on a set  $U$  is a family  $\mathcal{B}$  of subsets of  $U$  containing the empty set and closed under complementation and countable unions, that is:*

- (i)  $\emptyset \in \mathcal{B}$ .
- (ii) for all  $E \in \mathcal{B}$ ,  $E^c \in \mathcal{B}$ .
- (iii) for all  $E_1, E_2, \dots \in \mathcal{B}$ ,  $\cup_{i=1}^\infty E_i \in \mathcal{B}$ .

**Definition 1.2** (Measurable space). *A measurable space is a pair  $\mathbf{U} = \langle U, \mathcal{F}_U \rangle$  consisting of a set  $U$  and a  $\sigma$ -algebra  $\mathcal{F}_U$  on  $U$ .*

In some sense, the  $\sigma$ -algebra specifies the subsets of  $U$  that can be assigned a ‘size’, and which are therefore called the *measurable sets* of  $\mathbf{U}$ . We will use the convention of using boldface to refer to measurable spaces and regular font to refer to the underlying set. A trivial example of a  $\sigma$ -algebra over any set  $U$  is its powerset  $\mathcal{P}(U)$ , which gives the discrete measurable space  $\langle U, \mathcal{P}(U) \rangle$ , where every set is measurable. This is typically used when  $U$  is countable (finite or countably infinite), in which case this discrete  $\sigma$ -algebra is generated by the singletons. Another example, of central importance in measure theory, is the Borel  $\sigma$ -algebra  $\mathcal{B}_{\mathbb{R}}$  generated from the open sets of  $\mathbb{R}$ , whose elements are called the Borel sets. This gives the measurable space  $\langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$ . Working with Borel sets avoids the problems that would arise if we naively attempted to measure or assign probabilities to points in the continuum. More generally, any topological space gives rise to a Borel measurable space in this fashion. The product of measurable spaces can be defined as follows:

**Definition 1.3** (Finite product). *The product of measurable spaces  $\mathbf{U}_1 = \langle U_1, \mathcal{F}_1 \rangle$  and  $\mathbf{U}_2 = \langle U_2, \mathcal{F}_2 \rangle$  is the measurable space*

$$\mathbf{U}_1 \times \mathbf{U}_2 = \langle U_1 \times U_2, \mathcal{F}_1 \otimes \mathcal{F}_2 \rangle ,$$

where  $U_1 \times U_2$  is the Cartesian product of underlying sets, and the so-called tensor-product  $\sigma$ -algebra  $\mathcal{F}_1 \otimes \mathcal{F}_2$  is the  $\sigma$ -algebra on  $U_1 \times U_2$  which is generated by the subsets of the form  $E_1 \times E_2$  with  $E_1 \in \mathcal{F}_1$  and  $E_2 \in \mathcal{F}_2$ .

The product above is given by the *box topology*. In this dissertation, we will also need to deal with measurable spaces formed by taking the product of an uncountably infinite family of measurable spaces. In that case the box topology is no longer the most natural choice as enlightened by Tychonoff’s theorem [Tyc30] and we will use the *product topology*.

**Definition 1.4** (Infinite product). *Fix a possibly uncountably infinite index set  $I$ . The product of measurable spaces  $(\mathbf{U}_i = \langle U_i, \mathcal{F}_i \rangle)_{i \in I}$  is the measurable space:*

$$\prod_{i \in I} \mathbf{U}_i = \left\langle \prod_{i \in I} U_i, \prod_{i \in I} \mathcal{F}_i \right\rangle = \langle U_I, \mathcal{F}_I \rangle ,$$

where  $U_I = \prod_{i \in I} U_i$  is the Cartesian product of the underlying sets, and the  $\sigma$ -algebra  $\mathcal{F}_I = \prod_{i \in I} \mathcal{F}_i$  is obtained via the product topology i.e. it is generated by subsets of  $\prod_{i \in I} U_i$  of the form  $\prod_{i \in I} B_i$  where for all  $i \in I$ ,  $B_i \subseteq U_i$  and  $B_i \subsetneq U_i$  only for a finite number of  $i \in I$ .

Remarkably, the product topology is the smallest topology that makes the projection maps  $\pi_k: \prod_{i \in I} U_i \rightarrow U_k$  continuous. This definition is more general as it reduces straightforwardly to the case of a finite product. We can now formally define measurable functions and measures on those spaces.

**Definition 1.5** (Measurable function). *A measurable function between measurable spaces  $U = \langle U, \mathcal{F}_U \rangle$  and  $V = \langle V, \mathcal{F}_V \rangle$  is a function  $f: U \rightarrow V$  whose preimage preserves measurable sets, i.e. such that, for any  $E \in \mathcal{F}_V$ ,  $f^{-1}(E) \in \mathcal{F}_U$ .*

This is similar to the definition of a continuous function between topological spaces. Measurable functions compose as expected.

**Definition 1.6** (Measure). *A measure on a measurable space  $U = \langle U, \mathcal{F}_U \rangle$  is a function  $\mu: \mathcal{F}_U \rightarrow \overline{\mathbb{R}}$  from the  $\sigma$ -algebra to the extended real numbers  $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$  satisfying:*

- (i) (nonnegativity)  $\mu(E) \geq 0$  for all  $E \in \mathcal{F}_U$ ;
- (ii) (null empty set)  $\mu(\emptyset) = 0$ ;
- (iii) ( $\sigma$ -additivity) for any countable family  $(E_i)_{i=1}^{\infty}$  of pairwise disjoint measurable sets,  $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$ .

In particular, a measure  $\mu$  on  $U = \langle U, \mathcal{F}_U \rangle$  allows one to integrate well-behaved measurable functions  $f: U \rightarrow \mathbb{R}$  (where  $\mathbb{R}$  is equipped with its Borel  $\sigma$ -algebra  $\mathcal{B}_{\mathbb{R}}$ ) to obtain a real value, denoted

$$\int_U f \, d\mu \quad \text{or} \quad \int_{x \in U} f(x) \, d\mu(x). \tag{1.36}$$

A simple example of a measurable function is the *indicator function* of a measurable set  $E \in \mathcal{F}_U$ :

$$\mathbf{1}_E(x) := \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$$

For any measure  $\mu$  on  $U$ , its integral yields the ‘size’ of  $E$ :

$$\int_U \mathbf{1}_E \, d\mu = \mu(E). \tag{1.37}$$

A measure  $\mu$  on a measurable space  $U$  is said to be *finite* if  $\mu(U) < \infty$  and it is a *probability measure* if it is of unit mass i.e.  $\mu(U) = 1$ . We will denote by  $\mathbb{M}(U)$  and  $\mathbb{P}(U)$ , respectively, the sets of measures and probability measures on the measurable space  $U$ . Note that neither  $\mathbb{M}(U)$  nor  $\mathbb{P}(U)$  form a vector space. One way to recover this structure is to consider *finite-signed* measures. In comparison to the definition of a measure, one drops the nonnegativity requirement (i) in Definition 1.6, but insists that the values be finite. We denote this real vector space  $\mathbb{M}_{\pm}(U)$ . Equipped with total variation distance<sup>7</sup> it becomes a normed vector space.

A measurable function  $f$  between measurable spaces  $U$  and  $V$  carries any measure  $\mu$  on  $U$  to a measure  $f_*\mu$  on  $V$ . This is known as a *push-forward* operation. This push-forward measure is given by  $f_*\mu(E) = \mu(f^{-1}(E))$  for any set  $E$  measurable in  $V$ . An important use of

<sup>7</sup>For  $\mu, \nu \in \mathbb{M}_{\pm}(U)$ , the total variation distance is defined as  $\|\mu - \nu\| = |\mu - \nu|(U) = \sup_{A \in \mathcal{F}_U} |\mu(A) - \nu(A)|$ . When the set is countable, the total variation distance is related to the  $L^1$  norm by the identity  $\|\mu - \nu\| = \frac{1}{2} \sum_{x \in U} |\mu(x) - \nu(x)|$ .

push-forward measures is that for any integrable function  $g$  between measurable spaces  $\mathbf{V}$  and  $\langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$ , one can write the following change-of-variables formula

$$\int_U g \circ f \, d\mu = \int_{\mathbf{V}} g \, df_*\mu . \quad (1.38)$$

Importantly, the push-forward operation preserves the total measure, hence it takes  $\mathbb{P}(U)$  to  $\mathbb{P}(V)$ . A case that will be of particular interest to us is the push-forward of a measure  $\mu$  on a product space  $U_1 \times U_2$  along a projection  $\pi_i: U_1 \times U_2 \rightarrow U_i$ . This yields the *marginal measure*  $\mu|_{U_i} = \pi_{i*}\mu$ , where for any  $E \in \mathcal{F}_{U_1}$  measurable,  $\mu|_{U_1}(E) = \mu(\pi_1^{-1}(E)) = \mu(E \times U_2)$ . In the opposite direction, given a measure  $\mu_1$  on  $U_1$  and a measure  $\mu_2$  on  $U_2$ , a *product measure*  $\mu_1 \times \mu_2$  is a measure on the product measurable space  $U_1 \times U_2$  satisfying  $(\mu_1 \times \mu_2)(E_1 \times E_2) = \mu_1(E_1)\mu_2(E_2)$  for all  $E_1 \in \mathcal{F}_1$  and  $E_2 \in \mathcal{F}_2$ . For probability measures, there is a uniquely determined product measure.

The last ingredient we need from measure theory is called a *Markov kernel*.

**Definition 1.7** (Markov kernel). *A Markov kernel (or probability kernel) between measurable spaces  $U = \langle U, \mathcal{F}_U \rangle$  and  $V = \langle V, \mathcal{F}_V \rangle$  is a function  $k: U \times \mathcal{F}_V \rightarrow [0, 1]$  (the space  $[0, 1]$  is assumed to be equipped with its Borel  $\sigma$ -algebra) such that:*

- (i) *for all  $E \in \mathcal{F}_V$ ,  $k(-, E): U \rightarrow [0, 1]$  is a measurable function;*
- (ii) *for all  $x \in U$ ,  $k(x, -): \mathcal{F}_V \rightarrow [0, 1]$  is a probability measure.*

Markov kernels generalise the discrete notion of stochastic matrices.

## 1.4 Optimisation theory

Below we give a brief overview of the optimisation problems that arise in this dissertation. We begin by recalling some basic definitions [BBV04]. Let  $E$  be a real topological vector space.

**Definition 1.8** (Convex set). *A set  $F \subset E$  is said to be convex if for any  $x, y \in F$  and  $\lambda \in [0, 1]$ , we have that  $\lambda x + (1 - \lambda)y \in F$ .*

**Definition 1.9** (Convex function). *Let  $F$  be a convex subset of  $E$ .  $f: F \rightarrow \mathbb{R}$  is said to be convex if  $f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y)$  for any  $x, y \in F$  and  $\lambda \in [0, 1]$ .*

If  $-f$  is convex then we say that  $f$  is a *concave* function.

**Definition 1.10** (Cone). *A set  $C \subset E$  is a cone if for all  $x \in C$  and  $\lambda \in \mathbb{R}^+$ ,  $\lambda x \in C$ .*

If  $C$  is a convex subset of  $E$  and a cone, we say that  $C$  is a *convex cone*.

**Definition 1.11** (Proper cone). *A cone  $C \subset E$  is a proper cone if:*

- *$C$  is convex.*

- $C$  is closed (i.e. the limit of vectors with respect to the topology on  $E$  in  $C$  belong to  $C$ ).
- $C$  is solid (i.e. it has nonempty interior).
- $C$  is pointed (i.e. if  $x \in C$  and  $-x \in C$  then  $x = 0_E$ ).

A proper cone  $C \subset E$  induces a partial ordering on  $E$ :  $x \geq_C y \iff x - y \in C$  for  $x, y \in E$ . Thus  $x =_C y$  if  $x - y \in C$  and  $y - x \in C$ . In the rest of the manuscript, we will only deal with proper cones that we will refer to as cones.

A *convex optimisation problem* is an optimisation problem in which the objective function and the feasible set are convex. A detailed introduction to the general problem as well as algorithmic techniques for solving it can be found in e.g. [BTN01, BBV04, HUL13]. A convex optimisation problem can be generally expressed as:

$$\inf_{x \in F} f(x). \tag{1.39}$$

$F \subset E$  is called the *feasible set*. It is a convex subset specified by constraints: for instance linear inequalities or semidefinite inequalities.  $f: E \rightarrow \mathbb{R}$  is a convex real-valued function called the *objective function*. Note that the maximisation problem of optimising a concave function  $f$  can be expressed in this form with the convex function  $-f$ . Usually,  $E$  can be embedded into  $\mathbb{R}^n$  for some  $n \in \mathbb{N}^*$ . However, we will also consider other cases e.g. the space of real valued continuous functions. Such problems will yield *infinite-dimensional* convex optimisation problems. Convex programs have the remarkable property that a local optimum is always a global optimum. This fact is greatly exploited by solvers such as point interior methods.

A *feasible plan* or *feasible solution* is an element  $x \in E$  which belongs to  $F$  i.e. an element that satisfies all the constraints. We say that a convex optimisation program (P) is *consistent* if it possesses a feasible solution. A *strictly feasible solution* strictly satisfies all the constraints. If the infimum in Eq. (1.39) is attained, an *optimal plan* or *optimal solution* is a feasible plan that reaches this value. The *value* of a program (P), denoted  $\text{val}(\text{P})$ , is the value returned by the optimisation problem. It is the value reached by an optimal solution if there exists an optimal plan. It can be infinite if the program diverges.

In particular, we will be interested in a subfield of convex optimisation known as *conic programming*:

$$\inf_{Ax =_C b} f(x), \tag{1.40}$$

where  $A$  is a linear operator acting on  $E$ ,  $b \in E$  and  $C$  is a convex cone in  $E$ . Conic programming thus consists of the optimisation of a convex function over the intersection of an affine subspace  $\{y \in E | y = Ax - b\}$  and a convex cone  $C$ . This can be extended to constraints of the form  $Ax \geq_C b$ . We will mainly focus on linear objective functions.

Importantly conic optimisation reduces to *linear programming* (LP) when the convex cone is a nonnegative orthant; and to *semidefinite programming* (SDP) when it is the convex cone of positive semidefinite matrices. Remarkably, these two classes come with a duality theory: from a

*primal program* expressed as a minimisation problem, one can derive a *dual program* expressed as a maximisation problem. *Weak duality* always holds meaning that the value of the minimisation problem is lowerbounded by the value of the maximisation problem. Furthermore when the values of the programs agree, we have *strong duality* between the programs. Equivalently, we say there is *no duality gap*. The latter will only hold under specific conditions that we will detail later. There exist robust and efficient classical algorithms for solving linear programs and semidefinite programs [BTN01, BBV04], in particular via interior point methods.

Finally we will also encounter more exotic optimisation problems where we will have to deal with quadratic terms. Terms like  $x^T Q x$  for some vector  $x$  and some symmetric matrix  $Q$  will only be convex if  $Q$  is positive semidefinite. In that case, we will deal with quadratic SDPs. However we will often encounter optimisation problems where the objective function is no longer convex. The resulting program will not fit in the convex optimisation category. In particular we loose the very desirable property that a local optimum is automatically a global optimum resulting in a numerical resolution which is often much harder.

### 1.4.1 Linear programming

We present below the generic form of an infinite-dimensional linear program and main results regarding strong duality. Extensive introductions can be found in [Bar02, DT06] and we will follow the form given in [Bar02, IV-(6.1)] for such programs.

Let  $E_1, F_1$  and  $E_2, F_2$  be pairs of dual topological vector spaces equipped with the weak topology<sup>8</sup> induced by the dualities  $\langle -, - \rangle_{1,2}$ . They are defined as  $\langle -, - \rangle_1: E_1 \times F_1 \rightarrow \mathbb{R}$  and  $\langle -, - \rangle_2: E_2 \times F_2 \rightarrow \mathbb{R}$ .

We fix convex cones  $K_1 \subseteq E_1$  and  $K_2 \subseteq E_2$ . Convex dual cones  $K_1^*$  and  $K_2^*$  are defined as:

$$K_i^* = \{f \in F_i \text{ such that } \forall e \in K_i, \langle e, f \rangle \geq 0\} .$$

Furthermore, we fix the linear transformation  $A: E_1 \rightarrow E_2$ . We also fix its dual transformation  $A^*: F_2 \rightarrow F_1$  such that  $\forall e \in E_1, \forall f \in F_2, \langle A(e), f \rangle_2 = \langle e, A^*(f) \rangle_1$ . Finally, we fix  $c \in F_1$  and  $b \in E_2$ . A canonical example of a (potentially infinite-dimensional) primal<sup>9</sup> linear program can be expressed as:

$$(LP) \quad \left\{ \begin{array}{l} \text{Find } e_1 \in E_1 \\ \text{minimising } \langle e_1, c \rangle_1 \\ \text{subject to:} \\ \quad A(e_1) \geq_{K_2} b \\ \quad e_1 \geq_{K_1} 0 . \end{array} \right.$$

<sup>8</sup>The weak topology on  $E_1$  induced by  $F_1$  and the duality  $\langle -, - \rangle_1$  is the weakest (or coarsest) topology on  $E_1$  that makes all the maps  $\langle -, f_1 \rangle_1: E_1 \rightarrow \mathbb{R}$  continuous as  $f_1$  ranges over  $F_1$ . And similarly for  $E_2, F_1$  and  $F_2$ .

<sup>9</sup>Note that the naming ‘primal’ and ‘dual’ is purely arbitrary and could be exchanged.

Its dual problem can be expressed as:

$$(D-LP) \quad \left\{ \begin{array}{l} \text{Find } f_2 \in F_2 \\ \text{maximising } \langle b, f_2 \rangle_2 \\ \text{subject to:} \\ \quad A^*(f_2) \leq_{K_1^*} c \\ \quad f_2 \geq_{K_2} 0 . \end{array} \right.$$

**Example.** To fix ideas, let us consider an example that will be of particular importance in this manuscript. For some Borel measurable space  $\mathbf{U}$  of a compact Hausdorff set  $U$ ,  $E_1$  and  $E_2$  are the set of finite signed measures  $\mathbb{M}_\pm(\mathbf{U})$  and  $F_1$  and  $F_2$  are the set of continuous real valued functions on  $U$ ,  $C(U)$ . Note that  $\mathbb{M}_\pm(\mathbf{U})$  is indeed a concrete realisation of the topological dual space of  $C(U)$  via the Riesz–Markov–Kakutani representation theorem [Kak41]. The duality is then naturally given by:

$$\forall \mu \in \mathbb{M}_\pm(\mathbf{U}), \forall f \in C(U), \langle \mu, f \rangle = \int_U f \, d\mu. \quad (1.41)$$

$K_1$  and  $K_2$  are the convex cone of positive finite measure.  $K_1^*$  and  $K_2^*$  are then the set of positive real-valued functions on  $\mathbb{R}$ . We further fix  $A$  and  $A^*$  to be the identity map.  $b \in E_2$  is a fixed finite measure and  $c \in F_1$  a continuous function on  $U$ . Program (LP) then seeks the infimum of the real number  $\langle e_1, c \rangle_1$  over all finite-signed measures  $e_1$  under the constraint that  $e_1$  is a nonnegative measure and that it is lower bounded by the finite measure  $b$ , meaning that  $e_1 - b$  is itself a nonnegative finite measure. Likewise, program (D-LP) seeks the supremum of the real number  $\langle b, f_2 \rangle_2$  over all real-valued continuous functions  $f_2$  under the constraints that  $f_2$  is a nonnegative function and that it is upper bounded by the continuous function  $c$ , meaning that  $c - f_2$  is itself a nonnegative continuous function.

**Weak duality.** It always holds that  $\text{val}(\text{LP}) \geq \text{val}(\text{D-LP})$ .

**Strong duality.** This only holds under specific conditions. It is often a very desirable property. For instance, the computation of the contextual fraction [ABM17] (see Section 1.5.3) can be phrased as a linear program. Its dual program provides a method for calculating the maximally violated normalised Bell inequality which only holds because strong duality applies in this setting.

When the problem is *finite-dimensional*, a sufficient condition for strong duality is that the program has a finite value and its feasible region has an interior point. The latter is known as Slater’s condition. More formally, it requires the existence of a strictly feasible plan i.e. a point that satisfies the constraints with strict inequality. For instance, for problem (LP), it amounts to the existence of a positive  $e_1 \in E_1$  such that  $A(e_1) > b$ .

When the problem is *infinite-dimensional*, this becomes slightly more cumbersome. The following theorem, expressed with notation from (LP), provides a sufficient condition:

**Theorem 1.1** (Th. (7.2) [Bar02]). *Suppose that the set  $\{(A(x), \langle x, c \rangle_1) : x \in K_1\}$  is closed in  $E_2 \oplus \mathbb{R}$  and that there exists a primal feasible plan i.e. that the primal program is consistent. Then there is no duality gap i.e.  $\text{val}(\text{LP}) = \text{val}(\text{D-LP})$ . If  $\text{val}(\text{LP})$  is finite then there exists a primal optimal plan.*

### 1.4.2 Semidefinite programming

Semidefinite programs are a generalisation of linear programs where one optimises in the cone of semidefinite matrices rather than a positive orthant. Here we give the canonical form of finite-dimensional semidefinite programs [VB96]. Let  $M, N \in \mathbb{N}$ ,  $\mathbf{b} = (b_1, \dots, b_M) \in \mathbb{R}^M$ ,  $C \in \text{Sym}_N$  (the set of  $N \times N$  symmetric matrices), and  $B^{(i)} \in \text{Sym}_N$  for all  $i \in \llbracket 1, M \rrbracket$ .

$$(\text{SDP}) \quad \left\{ \begin{array}{l} \text{Find } X \in \text{Sym}_N \\ \text{minimising } \text{Tr}(C^T X) \\ \text{subject to:} \\ \quad \forall i \in 1, \dots, M, \quad \text{Tr}(B^{(i)} X) = b_i \\ \quad X \succeq 0 . \end{array} \right.$$

A matrix  $X \in \text{Mat}_M(\mathbb{R})$  is positive semidefinite, denoted  $X \succeq 0$ , if and only if for all  $\mathbf{u} \in \mathbb{R}^M$ ,  $\mathbf{u}^T X \mathbf{u} \geq 0$ . Its dual semidefinite program reads:

$$(\text{D-SDP}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} \in \mathbb{R}^M \\ \text{maximising } \mathbf{b}^T \mathbf{y} \\ \text{subject to:} \\ \quad \sum_{i=1}^M y_i B^{(i)} \preceq C . \end{array} \right.$$

Compared to the very general programs given in the previous subsection, the programs given above seem more rigid because we only account for finite-dimensional semidefinite programs in this dissertation—a similar simpler form could be given for finite-dimensional LP. Infinite-dimensional linear programs can either be *relaxed* (for instance when some constraints are relaxed or even suppressed i.e. when the optimisation is performed on a larger search space) or *restricted* (for instance when constraints are added). It turns out that with a clever choice of restrictions and relaxations, infinite-dimensional linear programs may sometimes be approximated by converging upper and lower hierarchies of finite dimensional semidefinite programs (which can actually be ran on a computer). In the case of optimising over measures, it is sometimes possible to rephrase the problem in terms of the infinite sequence of moments of the measure

by requiring that the so-called moment matrix is positive semidefinite. We then truncate the sequence of moment to obtain a finite dimensional semidefinite program [Las09, Las01]. This is known as the *Lasserre hierarchy* (or Lasserre relaxation) that we introduce shortly after. Dual to this vision, it is also sometimes possible to find sum-of-squares representation for continuous real-valued nonnegative functions. Then by truncating at a certain degree of the sum-of-squares representation, it can be rephrased as a semidefinite program [Par03, Las09].

**Weak duality.** It always holds that  $\text{val}(\text{SDP}) \geq \text{val}(\text{D-SDP})$ .

**Strong duality.** If one of the programs has a finite value and a feasible region with an interior point (a point that strictly satisfies all the constraints) then we have strong duality i.e.  $\text{val}(\text{SDP}) = \text{val}(\text{D-SDP})$ .

**Numerical implementation.** There exist nowadays numerous software packages to solve SDPs. A great open-source resource is the library Picos [SS12] which allows one to write high-level SDPs in Python. It can be interfaced with robust SDP solvers such as Mosek [ApS19] or SDPA [FKNY02]. Note that SDPA features a highly accurate multiple-precision arithmetic SDP solver called SDPA-GMP [Nak10]. While much slower, this significantly improves accuracy if numerical issues are critical.

### 1.4.3 The Lasserre–Parrilo hierarchy

Below we introduce the Lasserre–Parrilo hierarchy for relaxing infinite-dimensional linear programs known as Generalised Moment Problems [Las01, Las09, Par03].

We start by giving insightful results: Subsection 1.4.3.1 provides results concerning the representation of positive polynomials while Subsection 1.4.3.2 provides results to understand when a sequence can be represented by a Borel measure.

#### Notation, terminology

We work in  $\mathbb{R}^d$  for  $d \in \mathbb{N}^*$ . We fix  $\mathbf{K}$  to be a Borel measurable subspace of  $\mathbb{R}^d$ . We use the multi-index notation in Eq. (1), which we recall briefly. Let  $\mathbb{R}[\mathbf{x}]$  denote the ring of real polynomials in the variables  $\mathbf{x} \in \mathbb{R}^d$ , and let  $\mathbb{R}[\mathbf{x}]_k \subset \mathbb{R}[\mathbf{x}]$  contain those polynomials of total degree at most  $k$ . The latter forms a vector space of dimension  $s(k) := \binom{d+k}{k}$ , with a canonical basis consisting of monomials  $\mathbf{x}^\alpha := x_1^{\alpha_1} \cdots x_d^{\alpha_d}$ <sup>10</sup> indexed by the set  $\mathbb{N}_k^d := \{\alpha \in \mathbb{N}^d \mid |\alpha| \leq k\}$  where  $|\alpha| := \sum_{i=1}^d \alpha_i$ . For  $k \in \mathbb{N}$ ,  $\mathbf{x} \in \mathbb{R}^d$ , we define  $\mathbf{v}_k(\mathbf{x}) := (\mathbf{x}^\alpha)_{|\alpha| \leq k} = (1, x_1, \dots, x_n, x_1^2, x_1 x_2, \dots, x_n^k)^T$  the vector of monomials of total degree less or equal than  $k$ .

Any  $p \in \mathbb{R}[\mathbf{x}]_k$  is associated with a vector of coefficients  $\mathbf{p} := (p_\alpha) \in \mathbb{R}^{s(k)}$  via expansion in the canonical basis as  $p(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}_k^d} p_\alpha \mathbf{x}^\alpha$ .

<sup>10</sup>We also extend this notation to noncommutative variables in Subsection 1.4.4.

### Moment problem in probability

Given a finite set of indices  $\Gamma$ , a set of reals  $\{\gamma_j : j \in \Gamma\}$  and polynomials  $h_j : K \rightarrow \mathbb{R}$ ,  $j \in \Gamma$  (integrable with respect to every measure  $\mu \in \mathbb{M}_{\pm}(\mathbf{K})$ ) the corresponding *Global Moment Problem* (GMP) can be expressed as:

$$(\text{GMP}) \quad \left\{ \begin{array}{l} \text{Find } \mu \in \mathbb{M}_{\pm}(\mathbf{K}) \\ \text{maximising } \mu(\mathbf{K}) \\ \text{subject to:} \\ \forall j \in \Gamma, \quad \int_K h_j \, d\mu \leq \gamma_j . \end{array} \right.$$

Its dual program can be expressed as:

$$(\text{D-GMP}) \quad \left\{ \begin{array}{l} \text{Find } \boldsymbol{\lambda} \in \mathbb{R}^{\Gamma} \\ \text{minimising } \sum_{j \in \Gamma} \gamma_j \lambda_j \\ \text{subject to:} \\ \forall \mathbf{x} \in K, \quad \sum_{j \in \Gamma} \lambda_j h_j(\mathbf{x}) - \mathbf{x} \geq 0 \\ \forall j \in \Gamma, \quad \lambda_j \geq 0 . \end{array} \right.$$

#### 1.4.3.1 Positive polynomials and sum-of-squares

Here we present the link between positive polynomials and sum-of-squares representation so that we can derive a converging hierarchy of restriction problems for program (D-GMP).

**Definition 1.12** (sum-of-squares polynomial). *A polynomial  $p \in \mathbb{R}[\mathbf{x}]$  is a sum-of-squares (SOS) polynomial if there exists a finite family of polynomials  $(q_i)_{i \in I}$  such that  $p = \sum_{i \in I} q_i^2$ .*

SOS polynomials are widely used in convex optimisation. We will denote by  $\Sigma^2\mathbb{R}[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$  the set of (multivariate) SOS polynomials, and  $\Sigma^2\mathbb{R}[\mathbf{x}]_k \subset \Sigma^2\mathbb{R}[\mathbf{x}]$  the set of SOS polynomials of degree at most  $2k$ . The following proposition hints towards the reason why it is desirable to be able to look for a sum-of-squares decomposition: it can be cast as a semidefinite optimisation problem.

**Proposition 1.1** (Prop. 2.1, [Las09]). *A polynomial  $p \in \mathbb{R}[\mathbf{x}]_{2k}$  has a sum-of-squares decomposition if and only if there exists a real symmetric positive semidefinite matrix  $Q \in \text{Sym}_{s(k)}$  such that  $\forall \mathbf{x} \in \mathbb{R}^d$ ,  $p(\mathbf{x}) = \mathbf{v}_k(\mathbf{x})^T Q \mathbf{v}_k(\mathbf{x})$ .*

Then we will be looking at conditions under which a nonnegative polynomial can be expressed as a sum-of-squares polynomial. This is in essence the question raised by Hilbert in his 17<sup>th</sup> conjecture [Hil88].

**Definition 1.13** (quadratic module). For a family  $q = (q_j)_{j \in \{1, \dots, m\}}$  of polynomials, the set:

$$Q(q) := \left\{ \sum_{j=0}^m \sigma_j q_j \mid (\sigma_j)_{j \in \{0, \dots, m\}} \subset \Sigma^2 \mathbb{R}[\mathbf{x}] \right\} \quad (1.42)$$

is a convex cone in  $\mathbb{R}[\mathbf{x}]$  called the quadratic module generated by the family  $q$  with, for convenience,  $q_0 = 1$  added. For  $k \in \mathbb{N}$ , we define  $Q_k(q)$  to be the quadratic module  $Q(q)$  where we further impose that  $(\sigma_j)_{j \in \{0, \dots, m\}} \subset \Sigma^2 \mathbb{R}[\mathbf{x}]_k$  i.e. we limit the degree of SOS polynomials.

**Assumptions 1.1.** Let  $K \subset \mathbb{R}^d$ . We make the following three assumptions on  $K$ .

- (i) Suppose  $K$  is a basic semi-algebraic set i.e. there exists a family of polynomials  $g = (g_j)_{j \in \{1, \dots, m\}} \in \mathbb{R}[\mathbf{x}]^m$  of degrees  $\deg(g_j)$  respectively such that:

$$K := \left\{ \mathbf{x} \in \mathbb{R}^d \mid \forall j = 1, \dots, m, g_j(\mathbf{x}) \geq 0 \right\}. \quad (1.43)$$

- (ii) Further suppose that  $K$  is compact.

- (iii) Finally suppose that there exists  $u \in Q(q)$  such that the level set  $\left\{ \mathbf{x} \in \mathbb{R}^d \mid u(\mathbf{x}) \geq 0 \right\}$  is compact.

In the family of polynomials  $(g_j)_{j \in \{1, \dots, m\}}$  we add  $g_0 = 1$  for convenience.

The following theorem is the key result that we will exploit for deriving the hierarchy of SDP restrictions for the dual program (GMP).

**Theorem 1.2** (Putinar's Positivellensatz [Put93]). Let  $K \subset \mathbb{R}^d$  satisfy Assumptions 1.1. If  $p \in \mathbb{R}[\mathbf{x}]$  is strictly positive on  $K$  then  $p \in Q(g)$ , that is

$$p = \sum_{j=0}^m \sigma_j g_j \quad (1.44)$$

for some sum-of-squares polynomials  $\sigma_j \in \Sigma^2 \mathbb{R}[\mathbf{x}]$  for  $j = 0, 1, \dots, m$ .

A proof can also be found in [Lau09].

Using the result above and Assumption 1.1, one can derive a hierarchy of SDPs [Las09] which provide a converging sequence of optimal values towards the value of program (D-GMP):

$$(D-GMP^k) \quad \left\{ \begin{array}{l} \text{Find } \boldsymbol{\lambda} = (\lambda_j)_{j \in \Gamma} \in \mathbb{R}^\Gamma \text{ and } \forall j = 0, \dots, m, f_j \in \Sigma^2 \mathbb{R}[\mathbf{x}]_{k - \lceil \frac{\deg(g_j)}{2} \rceil} \\ \text{minimising } y_0 \\ \text{subject to:} \\ \sum_{j \in \Gamma} \lambda_j h_j - \mathbf{1}_K = \sum_{j=0}^m f_j g_j \\ \forall j \in \Gamma, \quad \lambda_j \geq 0. \end{array} \right.$$

### 1.4.3.2 Moment sequences and moment matrices

In this subsection, we want to understand why the program (P-CF<sup>CV</sup>) can be relaxed so that a converging hierarchy of SDPs can be derived. The program (P-CF<sup>CV</sup>) is essentially a maximisation problem on finite-signed Borel measures with additional constraints such as the fact that these are proper measures (i.e. they are nonnegative). We will represent a measure by its moment sequence and find conditions for which this moment sequence has a (unique) representing Borel measure.

**Definition 1.14** (Riesz functional  $L_{\mathbf{y}}$ ). *Given a sequence  $\mathbf{y} = (y_{\alpha})_{\alpha \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$ , we define the linear functional  $L_{\mathbf{y}}: \mathbb{R}[\mathbf{x}] \rightarrow \mathbb{R}$  by*

$$L_{\mathbf{y}}(p) := \sum_{\alpha \in \mathbb{N}^d} p_{\alpha} y_{\alpha}. \quad (1.45)$$

**Definition 1.15** (Moment sequence). *Given a measure  $\mu \in \mathbb{M}(\mathbf{K})$ , its moment sequence  $\mathbf{y} = (y_{\alpha})_{\alpha \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$  is given by*

$$y_{\alpha} := \int_{\mathbf{K}} \mathbf{x}^{\alpha} d\mu(\mathbf{x}). \quad (1.46)$$

We say that  $\mathbf{y}$  has a unique representing measure  $\mu$  when there exists a unique  $\mu$  such that Eq. (1.46) holds. If  $\mu$  is unique then we say it is determinate (i.e. determined by its moments).

The linear functional  $L_{\mathbf{y}}$  then gives integration of polynomials with respect to  $\mu$  i.e. for any  $p \in \mathbb{R}[\mathbf{x}]$ :

$$\begin{aligned} L_{\mathbf{y}}(p) &= \sum_{\alpha \in \mathbb{N}^d} p_{\alpha} y_{\alpha} = \sum_{\alpha \in \mathbb{N}^d} p_{\alpha} \int_{\mathbf{K}} \mathbf{x}^{\alpha} d\mu(\mathbf{x}) = \int_{\mathbf{K}} \sum_{\alpha \in \mathbb{N}^d} p_{\alpha} \mathbf{x}^{\alpha} d\mu(\mathbf{x}) \\ &= \int_{\mathbf{K}} p(\mathbf{x}) d\mu(\mathbf{x}) \\ &= \int_{\mathbf{K}} p d\mu, \end{aligned} \quad (1.47)$$

where we reversed summation and integration because the sum is finite since  $p$  is a polynomial.

The following theorem is often used in optimisation theory over measures as it provides a necessary and sufficient condition for a sequence to have a representing measure.

**Theorem 1.3** (Riesz-Haviland [Hav36]). *Let  $\mathbf{y} = (y_{\alpha})_{\alpha \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$  and suppose that  $K \subseteq \mathbb{R}^d$  is closed. Then  $\mathbf{y}$  has a representing (nonnegative) measure i.e. there exists  $\mu$  a measure on  $K$  such that:*

$$\forall \alpha \in \mathbb{N}^d, \int_K \mathbf{x}^{\alpha} d\mu = y_{\alpha}$$

*if and only if  $L_{\mathbf{y}}(p) \geq 0$  for all polynomials  $p \in \mathbb{R}[\mathbf{x}]$  nonnegative on  $K$ .*

We recall that for  $k \in \mathbb{N}$ ,  $s(k) = \binom{d+k}{k}$ .

**Definition 1.16** (Moment matrix). For each  $k \in \mathbb{N}$ , the moment matrix of order  $k$   $M_k(\mathbf{y}) \in \text{Sym}_{s(k)}$  of a truncated sequence  $(y_\alpha)_{\alpha \in \mathbb{N}_{2k}^d}$  is the  $s(k) \times s(k)$  symmetric matrix with rows and columns indexed by  $\mathbb{N}_k^d$  (i.e. by the canonical basis for  $\mathbb{R}[\mathbf{x}]_k$ ) defined as follows: for any  $\alpha, \beta \in \mathbb{N}_k^d$ ,

$$(M_k(\mathbf{y}))_{\alpha, \beta} := L_{\mathbf{y}}(\mathbf{x}^{\alpha+\beta}) = y_{\alpha+\beta} . \quad (1.48)$$

**Definition 1.17** (Localising matrix). Given a polynomial  $p \in \mathbb{R}[\mathbf{x}]$ , the localising matrix  $M_k(p\mathbf{y}) \in \text{Mat}_{s(k)}(\mathbb{R})$  of a moment sequence  $(y_\alpha)_{\alpha \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$  is defined by: for all  $\alpha, \beta \in \mathbb{N}_k^d$ ,

$$(M_k(p\mathbf{y}))_{\alpha, \beta} := L_{\mathbf{y}}(p(\mathbf{x})\mathbf{x}^{\alpha+\beta}) = \sum_{\gamma \in \mathbb{N}^d} p_\gamma y_{\alpha+\beta+\gamma} . \quad (1.49)$$

The localising matrix reduces to the moment matrix for  $p = 1$ . For well-defined moment sequences, i.e. sequences that have a representing finite Borel measure, moment matrices and localising matrices are positive semidefinite, which provides insight on the reason why problem (P-CF<sup>CV</sup>) can be relaxed to a problem with positive semidefiniteness constraints.

**Proposition 1.2.** Let  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$  be a sequence of moments for some finite Borel measure  $\mu$  on  $\mathbf{K}$ . Then for all  $k \in \mathbb{N}$ ,  $M_k(\mathbf{y}) \succeq 0$ . If  $\mu$  has support contained in the set  $\{\mathbf{x} \in K \mid g(\mathbf{x}) \geq 0\}$  for some polynomial  $g \in \mathbb{R}[\mathbf{x}]$  then, for all  $k \in \mathbb{N}$ ,  $M_k(g\mathbf{y}) \succeq 0$ .

*Proof.* Let  $\mathbf{y} = (y_\alpha)$  be the moment sequence of a given Borel measure  $\mu$  on  $\mathbf{K}$ . Fix  $k \in \mathbb{N}$ . For any vector  $\mathbf{v} \in \mathbb{R}^{s(k)}$  (noting that  $\mathbf{v}$  is canonically associated with a polynomial  $v \in \mathbb{R}[\mathbf{x}]_k$  in the basis  $(\mathbf{x}^\alpha)$ ):

$$\mathbf{v}^T M_k(\mathbf{y}) \mathbf{v} = \sum_{\alpha, \beta \in \mathbb{N}_k^d} v_\alpha y_{\alpha+\beta} v_\beta \quad (1.50)$$

$$= \sum_{\alpha, \beta \in \mathbb{N}_k^d} v_\alpha v_\beta \int_K \mathbf{x}^{\alpha+\beta} d\mu \quad (1.51)$$

$$= \int_K \left( \sum_{\alpha \in \mathbb{N}_k^d} v_\alpha \mathbf{x}^\alpha \right)^2 d\mu \quad (1.52)$$

$$= \int_K v^2(\mathbf{x}) d\mu \geq 0 . \quad (1.53)$$

Thus  $M_k(\mathbf{y}) \succeq 0$ .

Similarly we can prove that the localising matrix  $M_k(g\mathbf{y})$  is positive semidefinite when  $g$  is a nonnegative polynomial on the support of  $\mu$ . Indeed for all  $\mathbf{v} \in \mathbb{R}^{s(k)}$ :

$$\mathbf{v}^T M_k(g\mathbf{y}) \mathbf{v} = \int_K v^2(\mathbf{x}) g(\mathbf{x}) d\mu \geq 0 , \quad (1.54)$$

which concludes the proof.  $\square$

The following theorem, which is the dual facet of Theorem 1.2, is the key result for deriving the hierarchy of SDP relaxations for the primal problem (P-CF<sup>CV</sup>). It provides a necessary and sufficient condition for a sequence to have a representing measure.

**Theorem 1.4** (Th. 3.8 [Las09]). *Let  $\mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}^d} \in \mathbb{R}^{\mathbb{N}^d}$ . Let  $K \subset \mathbb{R}^d$  satisfy Assumptions 1.1. Then  $\mathbf{y}$  has a finite Borel representing measure with support contained in  $K$  if and only if:*

$$M_k(\mathbf{y}) \succeq 0, \quad \forall k \in \mathbb{N}, \quad (1.55)$$

$$M_k(g_j \mathbf{y}) \succeq 0, \quad \forall j = 1, \dots, m, \quad \forall k \in \mathbb{N}. \quad (1.56)$$

Using the above result and Assumption 1.1, one can derive a hierarchy of SDPs [Las09] which provide a converging sequence of optimal values towards the value of program (GMP):

$$(\text{GMP}^k) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}_{2k}^d} \in \mathbb{R}^{s(2k)} \\ \text{maximising } y_0 \\ \text{subject to:} \\ \quad \forall j \in \Gamma, \quad L_{\mathbf{y}}(h_j) \leq \gamma_j \\ \quad M_k(\mathbf{y}) \succeq 0 \\ \quad \forall i = 1, \dots, m, \quad M_{k - \lceil \frac{\deg(g_i)}{2} \rceil}(g_i \mathbf{y}) \succeq 0. \end{array} \right.$$

We refer readers to [Las09] for the proof of convergence of the hierarchies given by programs (GMP<sup>k</sup>) and (D-GMP<sup>k</sup>).

#### 1.4.4 Quadratic optimisation problems

Quadratic problems play an important role in Quantum Information since jointly optimising an expectation value over quantum states  $\rho$ 's and PVMs  $\{E_i\}_i$  amounts to optimising quadratic terms like  $\text{Tr}(\rho E_i)$  under the constraints that these operators are positive semidefinite and that the PVMs sum to the identity. However these terms are not convex so the resulting optimisation problem is challenging to solve (it is NP-hard in general).

Suppose we have a maximisation problem for a nonconvex bilinear quadratic problem. One obvious strategy to obtain a lower bound is to decompose the optimisation into two SDP sub problems: one when the PVMs are fixed; one when the quantum state is fixed. The solution of one sub program can then be used to fix the quantities in the other program. This is known as a see-saw iteration algorithm and they usually provide a tight lower bound though convergence is never guaranteed.

To obtain an upper bound on the value of the nonconvex problem, one can implement the Navascués–Pironio–Acín (NPA) hierarchy [NPA08] and in particular the finite-dimensional

variation known as the Navascués-Vértesi (NV) hierarchy [NFAV15] which will optimise from outside the set of quantum correlations to provide an upper bound on the sought value. Below we briefly introduce noncommutative polynomial optimisation (NPO) and give the NV hierarchy associated. We focus on this variant as the dimension of the Hilbert space is bounded and this will be required for our problems.

Consider the set  $S$  of all  $d$ -tuples of Hermitian operators  $\mathbf{X} = (X_1, \dots, X_d)$  satisfying  $\mathcal{R} = \{q_i(\mathbf{X}) \succeq 0 : i = 1, \dots, m\}$  where  $(q_i)_{i=1, \dots, m}$  are Hermitian polynomials (i.e.  $q_i(\mathbf{X})$  is Hermitian for  $\mathbf{X}$  a  $d$ -tuple of Hermitian operators). Given  $p$  a Hermitian polynomial and  $D \in \mathbb{N}^*$ , a generic NPO problem can be expressed as:

$$\text{(NPO)} \quad \left\{ \begin{array}{l} \text{Find a Hilbert space } \mathcal{H}, \psi \in \mathcal{D}(\mathcal{H}) \\ \text{and } \mathbf{X} \text{ a } d\text{-tuple of Hermitian operators on } \mathcal{H} \\ \text{maximising } \langle \psi | p(\mathbf{X}) | \psi \rangle \\ \text{subject to:} \\ \quad \dim(\mathcal{H}) \leq D \\ \quad \forall i = 1, \dots, m, \quad q_i(\mathbf{X}) \succeq 0 . \end{array} \right.$$

A sequence  $\mathbf{y} = (y_\alpha) \in \mathbb{R}^{s(2k)}$  admits a valid quantum representation if there exists  $\mathbf{X} \subset \mathcal{B}(\mathcal{H})$  satisfying  $\mathcal{R}$  with  $\dim(\mathcal{H}) \leq D$  and a pure state  $|\psi\rangle \in \mathcal{H}$  such that  $y_\alpha = \langle \psi | \mathbf{X}^\alpha | \psi \rangle$ . The moment vectors  $\mathbf{y}$  that admit a quantum representation with a dimensional constraint satisfy a number of additional linear restrictions which depend on  $D$ . For instance for  $D = 1$ , there are additional constraints that ensure the variables commute. Denote  $S_D^k$  the span of the set of feasible sequences  $\mathbf{y} = (y_\alpha) \in \mathbb{R}^{s(2k)}$ . The NV hierarchy is generally expressed as:

$$\text{(SDP-NPO}^k) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} = (y_\alpha)_{\alpha \in \mathbb{N}_{2k}^d} \in \mathbb{R}^{s(k)} \\ \text{maximising } L_{\mathbf{y}}(p) \\ \text{subject to:} \\ \quad \mathbf{y} \in S_D^k \\ \quad y_{\mathbf{0}} = 1 \\ \quad M_k(\mathbf{y}) \succeq 0 \\ \quad \forall i = 1, \dots, m, \quad M_k(q_i \mathbf{y}) \succeq 0 . \end{array} \right.$$

The key to implement this program is to characterise the subspace  $S_D^k$  and [NFAV15] provide methods to do so. It can be achieved by a randomised method where  $\mathbf{X}$  and  $|\psi\rangle$  are randomly generated to build a basis for  $S_D^k$ . Another method is deterministic and involve symbolic computations to retrieve a basis for  $S_D^k$ .

## 1.5 Sheaf-theoretic framework for contextuality

In this section, we aim at introducing the traditional notion of contextuality as first considered by Bell [Bel64] and Kochen and Specker [KS75]. We will define it following the later, more general approach developed in the seminal paper of Abramsky and Brandenburger often referred to as the sheaf-theoretic approach to measurement contextuality [AB11]—though in the present exposition, we will not dive deeply into the sheaf structure itself. The main ingredients are *empirical models* which can be thought of as providing formal descriptions of tables of data specifying the probabilities of joint outcomes for compatible measurements. These empirical models need an underlying abstract description of an experiment which is given by a *measurement scenario*. A crucial result is the Fine–Abramsky–Brandenburger (FAB) theorem which pinpoints a unified structure describing nonlocality and contextuality. Moreover, contextuality can be quantified using the contextual fraction [ABM17] which can be computed by a linear program.

### 1.5.1 Measurement scenarios and empirical models

#### Measurement scenarios

An abstract description of a particular experimental setup is formalised as a *measurement scenario*.

**Definition 1.18** (Measurement scenario). *A measurement scenario is a triple  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  where:*

- $\mathcal{X}$  is a finite set of measurement labels.
- $\mathcal{M}$  is a covering family of  $\mathcal{X}$  i.e. it is a set of subsets of  $\mathcal{X}$  such that  $\bigcup_{C \in \mathcal{M}} C = \mathcal{X}$ . We require that each element  $C \in \mathcal{M}$  is a maximal measurement context in the sense that  $\forall C, C' \in \mathcal{M}$  such that  $C \subseteq C'$  then  $C = C'$ . It represents a maximal set of compatible measurements. No element of this family is a proper subset of another.
- $\mathcal{O} = (\mathcal{O}_x)_{x \in \mathcal{X}}$  is a finite set of outcomes for each measurement. For some set of measurements  $U \subseteq \mathcal{X}$ , the joint outcome set is given by the product of the respective outcome spaces:  $\mathcal{O}_U = \prod_{x \in U} \mathcal{O}_x$ .

A very convenient way to picture contextuality is via bundle diagrams as introduced in [ABK<sup>+</sup>15]. The elements of  $\mathcal{X}$  are represented by the vertices of the base of the bundle diagram (see green vertices in figure 1.1). The elements of  $\mathcal{M}$  are represented by edges (or hyperedges in general) between those vertices (see blue lines). The outcomes (i.e. the elements of  $\mathcal{O}$ ) are displayed in fibres above each measurement (see purple vertices). The connection to bundles is very carefully set up in [ABK<sup>+</sup>15].

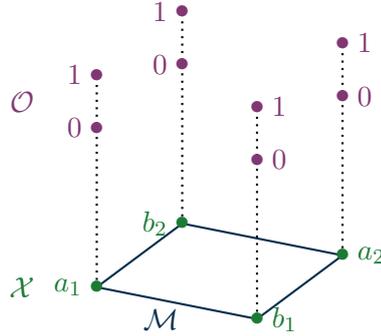


Figure 1.1: The example of the (2, 2, 2) Bell measurement scenario represented in the bundle diagram format.

**Example.** Let us express the standard (2,2,2) Bell scenario (2 parties, 2 observables each, 2 possible outcomes). It is represented as:

$$\mathcal{X} = \{a_1, a_2, b_1, b_2\}, \quad \mathcal{M} = \{\{a_1, b_1\}, \{b_1, a_2\}, \{a_2, b_2\}, \{b_2, a_1\}\}, \quad \forall x \in \mathcal{X}, \mathcal{O}_x = \{0, 1\}$$

### The language of sheaves

Different conventions have been used for defining the set of *sections*. We will adopt the one from [BKM21] and call a *section* an element of  $\mathcal{O}_U$  for  $U \in \mathcal{P}(\mathcal{X})$  a set of measurement labels. Sections from  $\mathcal{O}_U$  are joint outcomes for all measurements in  $U$ . A *global section* is a global assignment i.e. it is an element of  $\mathcal{O}_{\mathcal{X}}$ . The sheaf  $\mathcal{E}^{11}$  maps  $U \in \mathcal{P}(\mathcal{X})$  to  $\mathcal{E}(U) = \mathcal{O}_U$ . It is called the *event sheaf* as it assigns, to any set of measurements, information about the outcome events that could result from jointly performing them. Note that as well as applying the map to valid contexts  $U \in \mathcal{M}$  we will see that it can also be of interest to consider hypothetical outcome spaces for sets of measurements that do not necessarily form valid contexts, in particular  $\mathcal{E}(X) = \mathcal{O}_X$ , the joint outcome space for *all* measurements. Sheaves are widely used in modern mathematics. See [MM92] for a comprehensive reference. They might roughly be thought of as providing a means of assigning information to the open sets of a topological space. This is performed in such a way that information can be restricted to smaller open sets and consistent information on family of open sets can be uniquely ‘glued’ on their union. We are concerned with discrete topological spaces whose points represent measurements, and the information that we are interested in assigning has to do with outcome spaces for these measurements and probability measures on these outcome spaces. Sheaves can be defined concisely in category-theoretic terms as contravariant functors (presheaves) satisfying an additional gluing condition, though in what follows we will also give a more concrete description in terms of restriction maps.

Sheaves come with a notion of *restriction*. Restriction arises in the following way: whenever  $U, V \in \mathcal{P}(\mathcal{X})$  with  $U \subseteq V$  we have an obvious restriction map  $\rho_U^V: \mathcal{E}(V) \rightarrow \mathcal{E}(U)$  which simply

<sup>11</sup>It is a functor  $\mathcal{E}: \mathcal{P}(\mathcal{X}) \rightarrow \text{Set}$

projects from the product outcome space for  $V$  to that for  $U$ . Note that  $\rho_U^U$  is the identity map for any  $U \in \mathcal{P}(\mathcal{X})$  and that if  $U \subseteq V \subseteq W$  in  $\mathcal{P}(\mathcal{X})$  then  $\rho_U^V \circ \rho_V^W = \rho_U^W$ . Already this is enough to show that  $\mathcal{E}$  is a *presheaf*. For an inclusion  $U \subseteq V$  and a section  $o \in \mathcal{E}(V) = \mathcal{O}_V$ , it is often more convenient to use the notation  $o|_U$  to denote  $\rho_U^V(o) \in \mathcal{E}(U) = \mathcal{O}_U$ , the restriction of  $o$  to  $\mathcal{O}_U$  which selects outcomes in  $o$  that corresponds to measurement labels in  $U$ .

Additionally, the unique gluing property holds for  $\mathcal{E}$ . Suppose that  $\mathcal{N} \subseteq \mathcal{P}(X)$  and we have an  $\mathcal{N}$ -indexed family of sections  $(o_U \in \mathcal{O}_U)_{U \in \mathcal{N}}$  that is compatible in the sense that its elements agree on overlaps, i.e. that for all  $U, V \in \mathcal{N}$ ,  $o_U|_{U \cap V} = o_V|_{U \cap V}$ . Then these sections can always be ‘glued’ together in a unique fashion to obtain a section  $o_N$  over  $N := \cup \mathcal{N}$  such that  $o_N|_U = o_U$  for all  $U \in \mathcal{N}$ . This makes  $\mathcal{E}$  a *sheaf*. For a concise view on the reason why contextuality might be approached by topological arguments see [Man20].

### Empirical model.

An empirical model is an object that allows one to specify the probabilities of observed joint outcomes on a given measurement scenario.

**Definition 1.19** (Empirical model). *An empirical model  $e$  (also called empirical behaviour) on the measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  is a family  $e = (e_C)_{C \in \mathcal{M}}$ , where  $e_C$  is a probability distribution on the joint outcome space  $\mathcal{O}_C$  for each maximal context  $C$ . These probability distributions further satisfy the compatibility conditions:*

$$\forall C, C' \in \mathcal{M}, \quad e_C|_{C \cap C'} = e_{C'}|_{C \cap C'}.$$

The notation  $e_C|_U$  for  $C \in \mathcal{M}$  and  $U \subset C$  is shorthand for the standard marginalisation of a probability distribution: for  $t \in \mathcal{O}_U$ ,  $e_C|_U(t) := \sum_{s \in \mathcal{O}_C, s|_U = t} e_C(s)$ . The compatibility condition, which requires that the marginals of these distributions agree on overlapping contexts, is sometimes referred to as the *no-disturbance* condition or as the *generalised no-signalling* condition. A special case is no-signalling which occurs when measurements are performed at space-like separated regions. Compatibility then ensures that the choice of which measurement to perform at one location does not affect the empirical prediction at another location. The no-disturbance principle is satisfied by all empirical models that arise from quantum predictions. Note that due to the structure inherited from probability distributions, empirical models are closed under convex combinations.

**Example.** We express an empirical model for the Clauser–Horne–Shimony–Holt (CHSH) experiment [CHSH69] on the (2,2,2) Bell scenario as the table 1.2. Note that

$$e^{\text{CHSH}} = \frac{\eta}{4} e^{\mathbb{1}} + (1 - \eta) e^{\text{PR box}} \quad (1.57)$$

where  $\eta = 1 - \frac{\sqrt{2}}{2}$ .  $e^{\text{CHSH}}$  is an empirical model corresponding to the CHSH experiment,  $e^{\mathbb{1}}$  the one corresponding to a maximally mixed state and  $e^{\text{PR box}}$  an empirical model for a PR box.

A	B	00	01	10	11
$a_1$	$b_1$	$\eta_1$	$\eta_2$	$\eta_2$	$\eta_1$
$a_1$	$b_2$	$\eta_1$	$\eta_2$	$\eta_2$	$\eta_1$
$a_2$	$b_1$	$\eta_1$	$\eta_2$	$\eta_2$	$\eta_1$
$a_2$	$b_2$	$\eta_2$	$\eta_1$	$\eta_1$	$\eta_2$

Table 1.2: Empirical model on the  $(2, 2, 2)$  Bell scenario specifying the probabilities of the joint outcomes for the CHSH model with  $\eta_1 = \frac{2+\sqrt{2}}{8}$  and  $\eta_2 = \frac{2-\sqrt{2}}{8}$ . Here  $a_i$  and  $b_i$  for  $i = 1, 2$  represent quantum observables for CHSH with eigenvalues relabelled as 0 and 1. Joint probabilities are obtained by the Born rule.

For instance the first row indicates that for the context  $C = \{a_1, b_1\}$ , the empirical model assigns probability  $\frac{1}{2}$  to the local sections  $(0, 0)$  and  $(1, 1)$  in context  $\{a_1, b_1\}$ . To understand the compatibility condition, let us focus on the first and second rows of the table. Fix  $C = \{a_1, b_1\}$  and  $C' = \{a_1, b_2\}$ . Then  $U = C \cap C' = \{a_1\}$ . Consider any  $t \in \mathcal{O}_U$ : for instance the measurement associated to  $a_1$  gives the outcome 0. For the first row, the only local sections  $s \in \mathcal{O}_C$  such that  $s|_U = t$  are  $(0, 0)$  and  $(0, 1)$  which, marginalised on the second outcome corresponding to the measurement of  $a_2$ , gives the outcome 0 for  $a_1$ . Summing the probabilities associated to these sections:  $e_C|_U(0) = e_C((0, 0)) + e_C((0, 1)) = \eta_1 + \eta_2 = \frac{1}{2}$ . With the same procedure for the second row:  $e_{C'}|_U(0) = e_{C'}((0, 0)) + e_{C'}((0, 1)) = \eta_2 + \eta_1 = \frac{1}{2}$  as required.

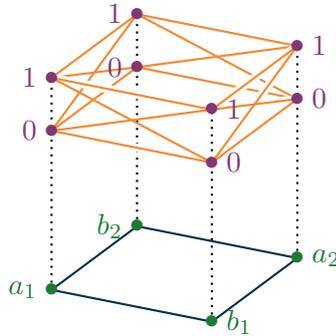


Figure 1.2: Bundle diagram for the CHSH model presented in Table 1.2 on the  $(2, 2, 2)$  Bell scenario.

We can now complete the bundle diagram 1.1 with the joint outcomes specified by the empirical model  $e$  (see orange edges in figure 1.2). Note that bundle diagrams are often used to witness possibilistic contextuality (i.e. logical or strong contextuality that we will define shortly) so that it is not always necessary to represent the probabilities associated to joint outcomes. Rather we only represent an edge between outcomes whenever the associated probability is non zero.<sup>12</sup>

<sup>12</sup>Note, however, that for the CHSH model it would be necessary to take account of the precise probabilities in order to witness contextuality.

## Extendability and Contextuality

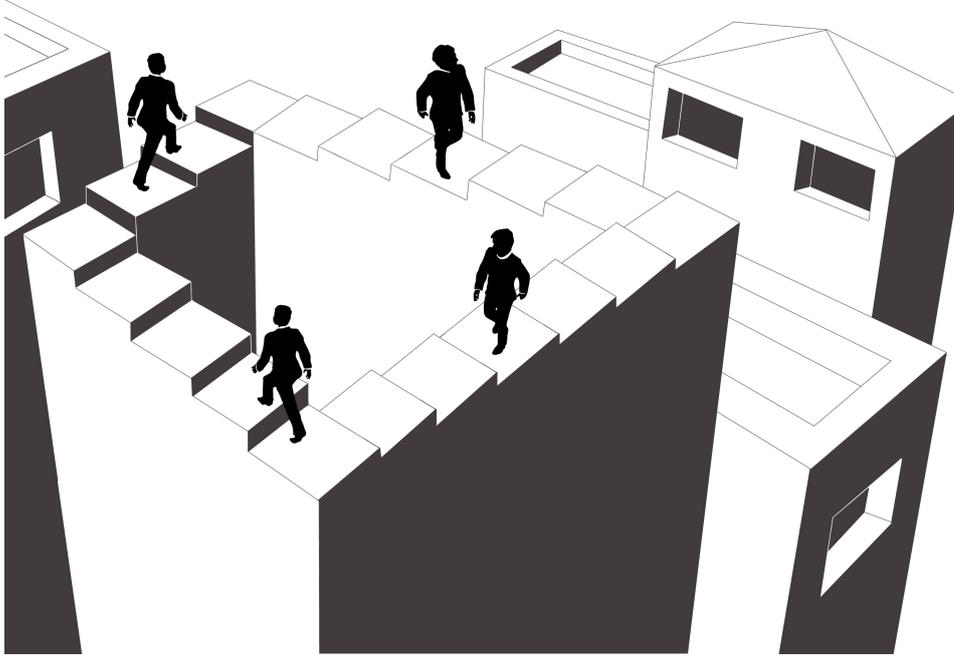


Figure 1.3: A drawing inspired by M.C. Escher's lithograph *Klimmen en dalen*.

We are now able to state the notion of measurement contextuality (or rather noncontextuality) which is a characteristic of the empirical behaviors. Informally noncontextuality can be thought of as an extendability property: local sections can be glued together consistently so that the empirical data can be described from global sections only. Local predictions can then be obtained from marginalising a probability distribution on global sections.

**Definition 1.20** (Noncontextuality or extendability). *An empirical model  $e$  on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  is said to be noncontextual (or extendable) if there exists a global probability distribution  $d$  on global assignments  $\mathcal{O}_{\mathcal{X}}$  such that  $\forall C \in \mathcal{M}, e_C = d|_C$ .*

When such a global distribution  $d$  cannot be found then we say the empirical model  $e$  is *contextual*. Equivalently, it is contextual when compatible local sections cannot be glued into global sections. Contextuality thus naturally arises as the tension between local and global consistency [AB11, BKM21]. Such tension might give rise to apparent paradoxes. This is beautifully illustrated by M.C. Escher's lithograph *Klimmen en dalen* (ascending and descending). A drawing inspired by this is given in Figure 1.3. Locally looking at the staircase, every character and its adjacent neighbours are going up the stairs consistently. However, taking a global look at the staircase gives an impossible figure in 3 dimensional Euclidean geometry. Notions of partial extendability have also been considered in [MB14, Sim17].

A stronger version of contextuality is referred to as *logical contextuality*. It arises when there exists a local section in the empirical model that cannot be obtained from a global section that is consistent with the support of the model. This can easily be pictured on a bundle diagram. On figure 1.4 which illustrates the Hardy paradox [Har93], the yellow path shows a local section ( $\{1, 1\}$  in context  $a_1, b_1$ ) that cannot be extended consistently. Finally, *strong contextuality*—a stronger form of logical contextuality—arises when no local sections can be extended to a global one consistently i.e. when there is no global section  $g \in \mathcal{O}_X$  such that  $\forall C \in \mathcal{M}, e_C(g|_C) > 0$ . The Popescu-Rohrlich (PR) box<sup>13</sup> [PR94] is such an example (see figure 1.5 where no path can be extended consistently).

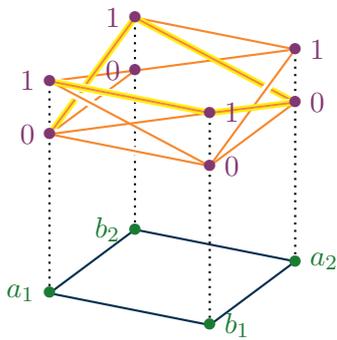


Figure 1.4: Logical contextual empirical model on the  $(2, 2, 2)$  Bell scenario (Hardy model).

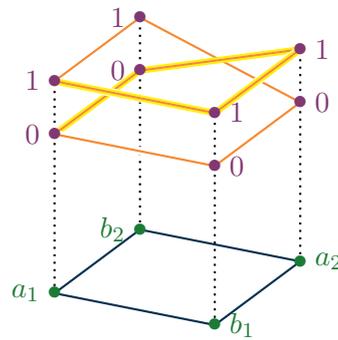


Figure 1.5: Strongly contextual empirical model on the  $(2, 2, 2)$  Bell scenario (PR box).

### 1.5.2 The FAB theorem

Previously, we characterised noncontextuality of an empirical model by the extendability property. Global sections are sufficient to capture noncontextual empirical behaviours via deterministic global states that assign predefined outcomes to all measurements. This is precisely the model referred to in the Kochen-Specker theorem [KS75]. On the other hand, Bell's theorem—focused on a multi-party experiment in which the parties may be spacelike separated—identifies another classical feature: factorisability rather than determinism. Fine unified these two notions in the case of the  $(2, 2, 2)$  Bell scenario [Fin82]. Later, Abramsky and Brandenburger [AB11] showed that this existential equivalence holds for any measurement scenario with observables with a discrete spectrum. It establishes an unambiguous, unified treatment of locality and noncontextuality, which is captured in a canonical way by the notion of extendability.

We begin by introducing the notion of hidden-variable models (HVM). Note that HVMs are often referred to as ontological models [Spe05] nowadays. The latter has become widely

<sup>13</sup>Interestingly the example of these local distributions that cannot be explained by a global distribution was already introduced in [Vor62]

used in quantum foundations in recent years. It indicates that the hidden variable—or the ontic state—is supposed to provide an underlying description of the physical world at perhaps a more fundamental level than the empirical-level description via the quantum state for example. The idea behind the introduction of HVMs is that there exists some space  $\Lambda$  of hidden variables predetermining the empirical behaviour. The motivation is that hidden variables could *explain away* some of the more non-intuitive aspects of the empirical predictions of quantum mechanics, which would then arise as resulting from an incomplete knowledge of the true state of a system rather than being a fundamental feature. There is some precedent for this in physical theories: for instance, statistical mechanics—a probabilistic theory—admits a deeper, albeit usually very complex, description in terms of classical mechanics, which is purely deterministic. It is desirable to further impose constraints on hidden-variable models which will restrict the set of achievable empirical behaviours and require that it behaves *classically* in some sense. In the case of Bell locality, we require that the hidden-variable model must be local i.e. factorisable (in a sense made precise below). However, hidden variables may not be directly accessible themselves so we allow that we only have probabilistic information about which hidden variable pertains in the form of a probability distribution  $p$  on  $\Lambda$ . The empirical behaviour should then be obtained as an average over the hidden-variable behaviours.

**Definition 1.21** (Hidden-variable model). *A hidden-variable model on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  consists of the triple  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  where:*

- $\Lambda$  is the finite space of hidden variables;
- $p$  is a probability distribution on  $\Lambda$ ;
- for each  $\lambda \in \Lambda$ ,  $h^\lambda$  is empirical model i.e.  $h^\lambda = (h_C^\lambda)_{C \in \mathcal{M}}$  is a family where, for each context  $C \in \mathcal{M}$ ,  $h_C^\lambda$  is a probability distribution on the joint outcome space  $\mathcal{O}_C$ . For each  $\lambda \in \Lambda$ ,  $h_\lambda$  satisfies the following compatibility condition:

$$\forall C, C' \in \mathcal{M}, \quad h_C^\lambda|_{C \cap C'} = h_{C'}^\lambda|_{C \cap C'}.$$

Crucially, our definition of hidden-variable model assumes  $\lambda$ -independence [Dic98] and parameter-independence [Jar84, Shi86].  $\lambda$ -independence corresponds to the requirement that the probability distribution  $p$  is independent of the measurement context. It is crucial for otherwise every behaviour (including contextual ones) could be modelled trivially by HVMs. Parameter-independence corresponds to the compatibility condition, which also ensures that the corresponding hidden-variable behaviour satisfies no-disturbance [BK13].

A hidden-variable model  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  gives rise to an empirical behaviour as follows:

$$\forall C \in \mathcal{M}, \forall s \in \mathcal{O}_C, \quad e_C(s) = \sum_{\lambda \in \Lambda} p(\lambda) h_C^\lambda(s). \quad (1.58)$$

This is a valid empirical behaviour since it is a convex sum of empirical behaviours  $(h_\lambda)_{\lambda \in \Lambda}$ .

As mentioned, to ensure that the hidden-variable model behaves *classically* we need to impose further constraints. This motivates the notions of deterministic and of factorisable hidden-variable models.

**Definition 1.22** (Deterministic HVM). *A hidden-variable model  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  is said to be deterministic if for every  $\lambda \in \Lambda$  and for every maximal context  $C \in \mathcal{M}$  the probability distribution  $h_C^\lambda$  is a Kronecker delta i.e. there exists an assignment  $o \in \mathcal{O}_C$  such that  $h_C^\lambda = \delta_o$ .*

Note that global sections  $\mathcal{O}_\mathcal{X}$  can be seen as a canonical form of deterministic hidden variables, which assign a definite outcome to each measurement, independent of the measurement context. Each global section  $g \in \mathcal{O}_\mathcal{X}$  gives rise to a Kronecker delta distribution  $\delta_g$ . This induces a distribution on each context that can be obtained by marginalisation; this is again a Kronecker delta at the restriction of the global section  $g$  to the given context:  $\delta_g|_C = \delta_{g|_C}$ . Because the set of global assignments is finite, we can always restrict the hidden-variable space to be of finite dimension as an infinite-dimensional space will not capture more behaviours. This fact will become clear from the statement of the FAB theorem.

**Definition 1.23** (Factorisable HVM). *A hidden-variable model  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  is factorisable if the probability assigned to a joint outcome factors as the product of the probabilities assigned to the individual measurements i.e. for every  $\lambda \in \Lambda$  and for every maximal context  $C \in \mathcal{M}$ ,  $h_C^\lambda$  factorises as a product probability distribution. That is, for  $\lambda \in \Lambda$ ,  $C \in \mathcal{M}$  and for every  $s \in \mathcal{O}_C$ :*

$$h_C^\lambda(s) = \prod_{m \in C} h_C^\lambda|_{\{m\}}(s|_{\{m\}}) = \prod_{m \in C} h_m^\lambda(s|_{\{m\}}).$$

Due to the assumption of parameter-independence, we can unambiguously write  $h_m^\lambda$  for  $h_C^\lambda|_{\{m\}}$  as the marginalisation is independent of the context  $C$ .

We can now state the FAB theorem.

**Theorem 1.5** (FAB theorem [AB11]). *Let  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  a measurement scenario with finite number of measurement labels and finite outcome sets. Let  $e$  be an empirical behaviour on  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ . Then the following propositions are equivalent:*

- (1)  *$e$  is extendable.*
- (2)  *$e$  admits a realisation by a deterministic hidden-variable model.*
- (3)  *$e$  admits a realisation by a factorisable hidden-variable model.*

### 1.5.3 Quantifying contextuality

A more refined question than asking whether a given empirical behaviour is contextual or not is to ask:

*To what extent is a given empirical behaviour contextual?*

Indeed some fraction of the empirical model may admit a noncontextual explanation. This allows for a quantitative statement in terms of a measure known as the *contextual fraction*, originally introduced in the seminal work [AB11] and further developed in [ABM17]. Instead of asking for a proper probability distribution on global assignments that allows the retrieval of the empirical behaviour  $e$  by marginalisation at each context, the idea is to ask for a subprobability distribution  $\mathbf{b}$  (a distribution that sums to less than 1) on global sections  $\mathcal{O}_{\mathcal{X}}$  explaining some fraction of the empirical behaviour i.e. we require that  $\forall C \in \mathcal{M}, \mathbf{b}|_C \leq e_C$ .

We can create an empirical model by taking a convex sum of empirical behaviours i.e. for two empirical models  $e$  and  $e'$  on the same measurement scenario and for  $\lambda \in [0, 1]$ ,  $\lambda e + (1 - \lambda)e'$  is again a well-defined empirical model. An equivalent way of explaining some fraction of the empirical behaviour with a noncontextual behaviour is to ask for a convex decomposition of the form:

$$e = \lambda e^{NC} + (1 - \lambda)e', \quad (1.59)$$

where  $\lambda \in [0, 1]$ ,  $e^{NC}$  is a noncontextual empirical model and  $e'$  is some other (no-signalling) empirical model. The maximum weight  $\lambda$  on the noncontextual part or, equivalently, the maximum mass (i.e. the quantity  $\mathbf{1} \cdot \mathbf{b} = \sum_{i=1}^{|\mathcal{O}_{\mathcal{X}}|} b_i$ ) of a subprobability distribution is called *the noncontextual fraction* of  $e$  and denoted  $\text{NCF}(e) \in [0, 1]$ . It generalises the local fraction [EPR92, BKP06, AGA<sup>+</sup>12]. The contextual fraction  $\text{CF}(e)$  is defined as:  $\text{CF}(e) := 1 - \text{NCF}(e)$ . It was originally introduced in [AB11] where it was proven that a model with contextual fraction of 1 is necessarily strongly contextual. A convex decomposition as in Eq. (1.59) with maximal weight on the noncontextual part necessarily results in the following:

$$e = \text{NCF}(e)e^{NC} + \text{CF}(e)e^{SC}, \quad (1.60)$$

where  $e^{SC}$  is strongly contextual. Note that this decomposition might not be unique. Thus the contextual fraction might be formally defined as:

**Definition 1.24** (Noncontextual fraction). *Let  $e$  be an empirical model on the scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ . The noncontextual fraction of  $e$ , written  $\text{NCF}(e)$ , is defined as*

$$\sup \{ \mathbf{1} \cdot \mathbf{b} \mid \mathbf{b} \text{ a subprobability distribution on } \mathcal{O}_{\mathcal{X}} \text{ such that } \forall C \in \mathcal{M}, \mathbf{b}|_C \leq e_C \} . \quad (1.61)$$

### Computing the contextual fraction as a linear program

A very convenient property is that the computation of the noncontextual fraction can be phrased as a linear program (see Subsection 1.4.1) [ABM17]. Fix a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  (with a finite number of observables and finite spaces of outcomes). Let  $e$  be an empirical model on that scenario. Let  $n := |\mathcal{O}_{\mathcal{X}}|$  be the number of possible global assignments and  $m := \sum_{C \in \mathcal{M}} |\mathcal{O}_C|$  be the number of total local assignments ranging over all contexts. It can

also be expressed as  $m := |\{\langle C, s \rangle \text{ s.t. } C \in \mathcal{M} \text{ and } s \in \mathcal{O}_C\}|$ . We define the incidence matrix  $M$  as the  $m \times n$  (0,1)-matrix such that:

$$M[\langle C, s \rangle, g] := \begin{cases} 1 & \text{if } g|_C = s \\ 0 & \text{otherwise,} \end{cases} \quad (1.62)$$

ranging over contexts  $C \in \mathcal{M}$ , local sections  $s \in \mathcal{O}_C$ , and global sections  $g \in \mathcal{O}_{\mathcal{X}}$ . Each column corresponds to a global section and each row to a local section. To understand its action, read the matrix column-by-column by fixing a global section  $g \in \mathcal{O}_{\mathcal{X}}$ . For that column, the incidence matrix assigns a 1 in a row corresponding to a local section  $s$  in context  $C$  whenever the global section under consideration marginalises to  $s$  i.e. if  $g|_C = s$ . Thus it can be seen as a table of possible restrictions from global to local assignments.

The empirical model  $e$  can be represented as a vector  $\mathbf{v}^e \in [0, 1]^m$  where for a given context  $C \in \mathcal{M}$  and a local assignment  $s \in \mathcal{O}_C$ ,  $\mathbf{v}^e[\langle C, s \rangle] = e_C(s)$ . It is a flattened version of the tables that are usually used to represent empirical models (see table 1.2). Then  $e$  is noncontextual (see Definition 1.20) if there exists a global probability distribution  $\mathbf{d} \in [0, 1]^n$  such that  $M\mathbf{d} = \mathbf{v}^e$ . The computation of the noncontextual fraction of  $e$  defined in Definition 1.24 is given by the value of the following program:

$$(P\text{-CF}^{\text{DV}}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{b} \in \mathbb{R}^n \\ \text{maximising } \mathbf{1} \cdot \mathbf{b} \\ \text{subject to:} \\ \quad M\mathbf{b} \leq \mathbf{v}^e \\ \quad \mathbf{b} \geq \mathbf{0} . \end{array} \right.$$

### Computing generalised Bell inequality

**Definition 1.25** (Generalised Bell inequality). *A generalised Bell inequality for the measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  is given by  $\langle \mathbf{a}, R \rangle$  where  $\mathbf{a} \in \mathbb{R}^m$  is a real vector indexed by local assignments  $\langle C, s \rangle$  for  $C \in \mathcal{M}$  and  $s \in \mathcal{O}_C$ , and  $R \in \mathbb{R}$  is a bound. For all noncontextual empirical model  $e$  on  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ , it must hold that  $\mathbf{a} \cdot \mathbf{v}^e \leq R$ .  $\langle \mathbf{a}, R \rangle$  is said to be tight if there exists a noncontextual behaviour  $e$  that saturates the bound i.e. if  $\mathbf{a} \cdot \mathbf{v}^e = R$ .*

For no-signalling empirical models, the quantity  $\mathbf{a} \cdot \mathbf{v}^e$  is upper bounded by

$$\|a\| := \sum_{C \in \mathcal{M}} \max\{a[\langle C, s \rangle] \mid s \in \mathcal{O}_C\}, \quad (1.63)$$

which amounts, for each context  $C \in \mathcal{M}$ , to having a behaviour which is a Kronecker delta at the maximal component of  $\mathbf{a}$  for each context. Note that this behaviour might be signalling (i.e. it might not respect the compatibility condition). We will only consider inequalities for which  $\|a\| > R$  for otherwise the inequality would be trivially satisfied by all models.

Now the amount by which an empirical model  $e$  violates a generalised Bell inequality is  $\max\{0, \mathbf{a} \cdot \mathbf{v}^e - R\}$ . It is usually desirable to normalise the previous quantity in order to have a quantity lying in the unit interval. The *normalised violation* of a generalised Bell inequality  $\langle \mathbf{a}, R \rangle$  by a model  $e$  is:

$$\frac{\max\{0, \mathbf{a} \cdot \mathbf{v}^e - R\}}{\|\mathbf{a}\| - R}. \quad (1.64)$$

Importantly, an optimal solution of the dual program of (P-CF<sup>DV</sup>) (with a clever change of variables) describes a generalised Bell inequality with maximal normalised violation by the empirical model with bound 0. Thus a generalised Bell inequality with maximal normalised violation for  $e$  is described by  $\langle \mathbf{a}^*, 0 \rangle$  where  $\mathbf{a}^* \in \mathbb{R}^m$  is an optimal feasible plan of the following program with value  $\text{CF}(e)$ :

$$(B\text{-CF}^{\text{DV}}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{a} \in \mathbb{R}^m \\ \text{maximising } \mathbf{a} \cdot \mathbf{v}^e \\ \text{subject to:} \\ \mathbf{M}^T \mathbf{a} \leq \mathbf{0} \\ \mathbf{a} \geq |\mathcal{M}|^{-1} \mathbf{1} . \end{array} \right.$$

We now recall the main theorem from [ABM17]:

**Theorem 1.6** (Th.1 [ABM17]). *Let  $e$  be an empirical model on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ . (i) The normalised violation of any Bell inequality is at most  $\text{CF}(e)$ ; (ii) if  $\text{CF}(e) > 0$  then there exists a Bell inequality whose normalised violation is exactly  $\text{CF}(e)$ ; (iii) for any convex decomposition of a contextual model  $e$  as  $e = \text{NCF}(e)e^{NC} + \text{CF}(e)e^{SC}$ , a Bell inequality, for which the normalised violation is  $\text{CF}(e)$ , is tight at  $e^{NC}$  (provided  $\text{NCF}(e) > 0$ ) and maximally violated at  $e^{SC}$ .*

What happens for all the results presented above if we now wish to extend the support of empirical models to spaces with continuous outcomes? This is the question we address in the next chapter. For instance our set of measurement labels  $\mathcal{X}$  may refer to observables like position and momentum which in principle have a continuous spectrum (being the reals  $\mathbb{R}$ ). For a context  $C$  including one of these observables,  $e_C$  naturally becomes a probability measure over the joint outcome set  $\mathcal{O}_C$  which necessarily comprises a continuum. Hence the need for the frugal dose of measure theory introduced in Subsection 1.3.

## CONTINUOUS-VARIABLE NONLOCALITY AND CONTEXTUALITY

**T**O date the study of contextuality has largely focused on discrete-variable scenarios. As a consequence, the main frameworks and approaches to contextuality are tailored to modelling these, e.g. [AB11, CSW14, AFLS15, DKC15, Spe05]. In such scenarios, observables can only take values in discrete, finite sets. Discrete-variable scenarios typically arise in finite-dimensional quantum mechanics, e.g. when modelling quantum computations with qubit systems.

Since quantum mechanics itself is infinite-dimensional, it also makes sense from a foundational perspective to extend analyses of the key concept of contextuality to the continuous-variable setting. Furthermore, continuous variables can be useful when dealing with iteration, even when attention is restricted to finite-variable actions at discrete time steps, as is traditional in informatics. An interesting question, for example, is whether contextuality arises and is of interest in such situations as the infinite behaviour of quantum random walks.

What kind of experiments are we trying to model? Suppose that we can interact with a system by performing measurements on it and observing their outcomes. A feature of quantum systems is that not all observables commute, so that certain combinations of measurements are incompatible. At best, we can obtain empirical data for contexts in which only compatible measurements are performed, which can be collected by running the experiment repeatedly. Contextuality then arises when the empirical data obtained is inconsistent with the assumption that for each run of the experiment the system has a global and context-independent assignment of values to all of its observable properties. To take an operational perspective, a typical example of an experimental setup or scenario that we consider in this chapter is the one depicted in

Figure 2.1 [left]. In this standard scenario extended to the continuous-variable realm, a system is prepared in some fixed bipartite state, following which parties  $A$  and  $B$  may each choose between two measurement settings,  $m_A \in \{a, a'\}$  for  $A$  and  $m_B \in \{b, b'\}$  for  $B$ . We assume that outcomes of each measurement live in  $\mathbf{R}$ , which typically will be a measurable subspace of the real numbers  $\mathbb{R}$  (with its Borel  $\sigma$ -algebra). Depending on which choices of inputs—which specify a context—were made, the empirical data might for example be distributed according to one of the four hypothetical probability density plots in  $\mathbf{R}^2$  depicted in Figure 2.1 [right]. This scenario and hypothetical empirical behaviour has been considered elsewhere [KLFA18] as a continuous-variable version of the Popescu–Rohrlich (PR) box [PR94].

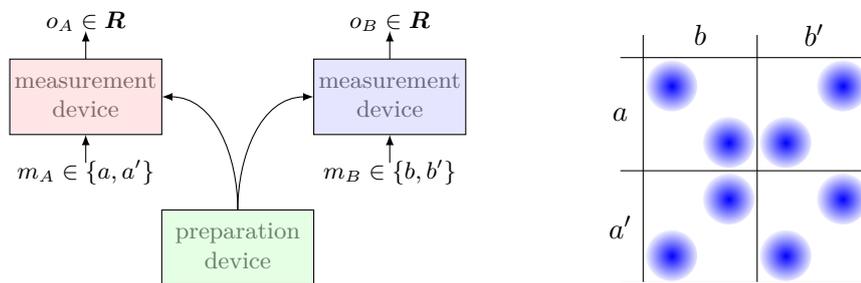


Figure 2.1: [Left] operational depiction of a typical bipartite experimental scenario. [Right] Hypothetical probability density plots for empirical data arising from such an experiment for a continuous generalisation of the PR box [KLFA18]. Cf. the discrete-variable probability tables of [MF12, Man17].

**Related work.** Note that we are specifically interested in scenarios involving observables with continuous spectra, or in more operational language, measurements having an outcome space which is a continuum. For quantifying contextuality, we still consider scenarios featuring only discrete sets of observables or measurements, as is typical in continuous-variable quantum computing, but we prove the Fine–Abramsky–Brandenburger (FAB) theorem for uncountable sets of measurement settings. This will be useful for Chapter 3. The possibility of considering contextuality in settings with outcome spaces being a continuum has also been evoked in [TC19]. We also note that several prior works have explicitly considered contextuality in continuous-variable systems [PC10, HCRD10, MPP11, SCW<sup>+</sup>12, ABS<sup>+</sup>15, LFKB<sup>+</sup>17, KLFA18]. In particular [CFRD07] presents a genuine continuous-variable Bell inequality that uses a second-order moment approach.

Our approach is distinct from these in that it provides a genuinely continuous-variable treatment of contextuality itself as opposed to embedding discrete-variable contextuality arguments into, or extracting them from, continuous-variable systems. This chapter is based on [BDE<sup>+</sup>19].

## 2.1 Framework

In this section, we follow closely the discrete-variable framework of [AB11] which is presented in Section 1.5 to develop a continuous-variable framework that allows to describe and analyse the empirical behaviours that arise with continuous-variable systems, although some extra care is required for dealing with continuous variables.

### Measurement scenarios

We recall that an abstract description of an experimental setup is formalised as a measurement scenario. Below we give its description for continuous-variable systems.

**Definition 2.1** (Measurement scenario). *A measurement scenario is a triple  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  whose elements are specified as follows.*

- $\mathcal{X}$  is a (possibly infinite) set of measurement labels.
- $\mathcal{M}$  is a covering family of subsets of  $\mathcal{X}$ , i.e. such that  $\bigcup \mathcal{M} = \mathcal{X}$ . The elements  $C \in \mathcal{M}$  are called maximal contexts and represent maximal sets of compatible observables. We therefore require that  $\mathcal{M}$  be an anti-chain with respect to subset inclusion, i.e. that no element of this family is a proper subset of another. Any subset of a maximal context also represents a set of compatible measurements.
- $\mathcal{O} = (\mathcal{O}_x)_{x \in \mathcal{X}}$  specifies a measurable space of outcomes  $\mathcal{O}_x = \langle \mathcal{O}_x, \mathcal{F}_x \rangle$  for each measurement  $x \in \mathcal{X}$ . If some set of measurements  $U \subseteq \mathcal{X}$  is considered together, there is a joint outcome space given by the product of the respective outcome spaces  $\mathcal{O}_U := \prod_{x \in U} \mathcal{O}_x = \langle \mathcal{O}_U, \mathcal{F}_U \rangle$  (see definition 1.3 for the finite case and definition 1.4 for the infinite case).

We consider  $\mathcal{X}$  as general as possible (since we will need  $\mathcal{X}$  of the form  $\mathbb{R}^M \times \mathbb{R}^M$  for some  $M \in \mathbb{N}^*$  in Chapter 3). Note that, later in the current chapter, we will restrict to finite set of measurement labels for the quantification of contextuality.

**Example.** The setup represented in Figure 2.1 is described by the measurement scenario:

$$\mathcal{X} = \{a, a', b, b'\}, \quad \mathcal{M} = \{\{a, b\}, \{a, b'\}, \{a', b\}, \{a', b'\}\}, \quad \forall x \in \mathcal{X}, \mathcal{O}_x = \mathbf{R}, \quad (2.1)$$

where  $\mathbf{R}$  is a measurable subspace of  $\langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$ .

As for the discrete-variable case, the *sheaf*  $\mathcal{E}$  that maps  $U \subseteq \mathcal{X}$  to  $\mathcal{E}(U) = \mathcal{O}_U$  is called the *event sheaf*. We recall that it comes with the notion of restriction, i.e. , for  $U \subseteq V \in \mathcal{P}(\mathcal{X})$  there is a restriction map  $\rho_U^V : \mathcal{O}_V \rightarrow \mathcal{O}_U$  which is continuous in the product topology.

### Empirical models

**Definition 2.2** (Empirical model). *An empirical model on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  is a family  $e = (e_C)_{C \in \mathcal{M}}$ , where  $e_C$  is a probability measure on the space  $\mathcal{E}(C) = \mathcal{O}_C$  for each*

maximal context  $C \in \mathcal{M}$ . It satisfies the compatibility conditions:

$$\forall C, C' \in \mathcal{M}, \quad e_C|_{C \cap C'} = e_{C'}|_{C \cap C'} .$$

Empirical models capture in a precise way the probabilistic *behaviours* that may arise upon performing measurements on physical systems. The compatibility condition ensures that the empirical behaviour of a given measurement or compatible subset of measurements is independent of which other compatible measurements might be performed along with them. As mentioned for the discrete-variable framework in Subsection 1.5.1, this is known as the *no-disturbance condition*. A special case is *no-signalling*, which applies in multi-party or Bell scenarios such as that of Figure 2.1 and Eq. (2.1). In that case, contexts consist of measurements that are supposed to occur in space-like separated locations, and compatibility ensures for instance that the choice of performing  $a$  or  $a'$  at the first location does not affect the empirical behaviour at the second location, i.e.  $e_{\{a,b\}}|_{\{b\}} = e_{\{a',b\}}|_{\{b\}}$ .

Note also that while empirical models may arise from the predictions of quantum theory, their definition is theory-independent. This means that empirical models can just as well describe hypothetical behaviours beyond what can be achieved by quantum mechanics such as the well-studied Popescu–Rohrlich box [PR94]. This can be useful in probing the limits of quantum theory and in singling out what distinguishes and characterises quantum theory within larger spaces of probabilistic theories, both well-established lines of research in quantum foundations.

### Extendability and contextuality

**Definition 2.3** (Noncontextuality). *An empirical model  $e$  on a scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  is extendable or noncontextual if there is a probability measure  $\mu$  on the space  $\mathcal{E}(\mathcal{X}) = \mathcal{O}_{\mathcal{X}}$  such that  $\mu|_C = e_C$  for every  $C \in \mathcal{M}$ .*

Recall that  $\mathcal{O}_{\mathcal{X}}$  is the space of global assignments. Of course, it is not always the case that  $\mathcal{X}$  is a valid context, and if it were then  $\mu = e_{\mathcal{X}}$  would trivially extend the empirical model. The question of the existence of such a probability measure that recovers the context-wise empirical content of  $e$  is particularly significant. When it exists, it amounts to a way of modelling the observed behaviour as arising stochastically from the behaviours of underlying states, identified with the elements of  $\mathcal{O}_{\mathcal{X}}$ . These elements deterministically assigns outcomes to *all* the measurements in  $\mathcal{X}$  independently of the measurement context that is actually performed. If an empirical model is not extendable it is said to be *contextual*.

## 2.2 A continuous-variable FAB theorem

As in the discrete-variable setting, we characterise contextuality of an empirical model by the absence of a global section for that empirical model. In this section, we prove a Fine–Abramsky–

Brandenburger (FAB) theorem in the continuous-variable setting. It establishes that in this setting there is also an unambiguous, unified description of locality and noncontextuality, which is captured in a canonical way by the notion of extendability.

We will begin by introducing hidden-variable models in a more precise way following the treatment given in Section 1.5.2.

**Definition 2.4** (Hidden-variable model). *A hidden-variable model on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  consists of the triple  $\langle \mathbf{\Lambda}, p, (k_C)_{C \in \mathcal{M}} \rangle$  where:*

- $\mathbf{\Lambda} = \langle \Lambda, \mathcal{F}_\Lambda \rangle$  is the measurable space of hidden variables,
- $p$  is a probability distribution on  $\mathbf{\Lambda}$ ,
- for each context  $C \in \mathcal{M}$ ,  $k_C$  is a probability kernel<sup>1</sup> between the measurable spaces  $\mathbf{\Lambda}$  and  $\mathcal{E}(C) = \mathcal{O}_C$  satisfying the following compatibility condition:

$$\forall C, C' \in \mathcal{M}, \forall \lambda \in \Lambda, \quad k_C(\lambda, -)|_{C \cap C'} = k_{C'}(\lambda, -)|_{C \cap C'} . \quad (2.2)$$

The last ingredient can be understood as defining a function  $\underline{k}$  from the set of hidden variables  $\Lambda$  to the set of empirical models over  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ . The function assigns to each  $\lambda \in \Lambda$  the empirical model  $\underline{k}(\lambda) := (\underline{k}(\lambda)_C)_{C \in \mathcal{M}}$ , where the correspondence with the definition above is via  $\underline{k}(\lambda)_C = k_C(\lambda, -)$ . This function must be ‘measurable’ in  $\mathbf{\Lambda}$  in the sense that  $\underline{k}(-)_C(B) : \Lambda \rightarrow [0, 1]$  is a measurable function for all  $C \in \mathcal{M}$  and  $B \in \mathcal{F}_C$ . Hence the need for Markov kernels.

A hidden-variable model  $\langle \mathbf{\Lambda}, p, (k_C)_{C \in \mathcal{M}} \rangle$  on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  gives rise to an empirical behaviour. The corresponding empirical model  $e$  is such that for any maximal context  $C \in \mathcal{M}$  and measurable set of joint outcomes  $B \in \mathcal{F}_C$ ,

$$e_C(B) = \int_{\Lambda} k_C(-, B) dp = \int_{\lambda \in \Lambda} k_C(\lambda, B) dp(\lambda) . \quad (2.3)$$

Note that this definition of hidden-variable model still assumes  $\lambda$ -independence and parameter-independence. As in the discrete-variable case, we may impose further constraints to ensure that the HVM behaves *classically*.

**Definition 2.5** (Deterministic HVM). *A hidden-variable model  $\langle \mathbf{\Lambda}, p, (k_C)_{C \in \mathcal{M}} \rangle$  is deterministic if  $k_C(\lambda, -) : \mathcal{F}_C \rightarrow [0, 1]$  is a Dirac measure for every  $\lambda \in \Lambda$  and for every maximal context  $C \in \mathcal{M}$ ; in other words, there is an assignment  $\mathbf{o} \in \mathcal{O}_C$  such that  $k_C(\lambda, -) = \delta_{\mathbf{o}}$ .*

In general discussions on hidden-variable models (e.g. [BY08]), the condition above, requiring that each hidden variable determines a unique joint outcome for each measurement context, is sometimes referred to as weak determinism. This is contraposed to strong determinism, which demands not only that each hidden variable fix a deterministic outcome to each individual

<sup>1</sup>Recall from Section 1.3 that a probability kernel  $k_C$  between  $\mathbf{\Lambda}$  and  $\mathcal{E}(C)$  is a function  $k_C : \Lambda \times \mathcal{F}_C \rightarrow [0, 1]$  which is: (i) a measurable function in the first argument, i.e.  $k_C(-, B) : \Lambda \rightarrow [0, 1]$  is measurable for every  $B \in \mathcal{F}_C$ ; and (ii) a probability measure in the second argument, i.e.  $k_C(\lambda, -) : \mathcal{F}_C \rightarrow [0, 1]$  is a probability measure for all  $\lambda \in \Lambda$ .

measurement, but that this outcome be independent of the context in which the measurement is performed. Note, however, that since our definition of hidden-variable models assumes the compatibility condition (2.2), i.e. parameter-independence, both notions of determinism coincide [BK13].

**Definition 2.6** (Factorisable HVM). *A hidden-variable model  $\langle \Lambda, p, (k_C)_{C \in \mathcal{M}} \rangle$  is factorisable if  $k_C(\lambda, -): \mathcal{F}_C \rightarrow [0, 1]$  factorises as a product measure for every  $\lambda \in \Lambda$  and for every maximal context  $C \in \mathcal{M}$ . That is, for any family of measurable sets  $(B_x \in \mathcal{F}_x)_{x \in C}$ ,*

$$k_C(\lambda, \prod_{x \in C} B_x) = \prod_{x \in C} k_C|_{\{x\}}(\lambda, B_x)$$

where  $k_C|_{\{x\}}(\lambda, -)$  is the marginal of the probability measure  $k_C(\lambda, -)$  on  $\mathcal{O}_C = \prod_{x \in C} \mathcal{O}_x$  to the space  $\mathcal{O}_{\{x\}} = \mathcal{O}_x$ .<sup>2</sup>

In other words, if we consider elements of  $\Lambda$  as inaccessible ‘empirical’ models – i.e. if we use the alternative definition of hidden-variable models using the map  $\underline{k}$  (see the remark below Definition 2.4) – then factorisability is the requirement that each of these be factorisable in the sense that

$$\underline{k}_C(\lambda) \left( \prod_{x \in C} B_x \right) = \prod_{x \in C} \underline{k}_{\{x\}}(\lambda)(B_x)$$

where  $\underline{k}_{\{x\}}(\lambda)$  is the marginal of the probability measure  $\underline{k}_C(\lambda)$  on  $\mathcal{O}_C = \prod_{x \in C} \mathcal{O}_x$  to the space  $\mathcal{O}_x$ .

We now prove the continuous-variable analogue of the theorem proved in the discrete probability setting by Abramsky and Brandenburger [AB11, Proposition 3.1 and Theorem 8.1] (see Section 1.5.2), generalising the result of Fine [Fin82] to arbitrary measurement scenarios with a possibly infinite number of measurement labels.

**Theorem 2.1.** *Let  $e$  be an empirical model on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ . The following are equivalent:*

- (1)  $e$  is extendable;
- (2)  $e$  admits a realisation by a deterministic hidden-variable model;
- (3)  $e$  admits a realisation by a factorisable hidden-variable model.

*Proof.* We prove the sequence of implications (1)  $\Rightarrow$  (2)  $\Rightarrow$  (3)  $\Rightarrow$  (1).

(1)  $\Rightarrow$  (2). The idea is that  $\mathcal{E}(\mathcal{X}) = \mathcal{O}_{\mathcal{X}}$  provides a canonical deterministic hidden-variable space. Suppose that  $e$  is extendable to a global probability measure  $\mu$ . Let us set  $\Lambda := \mathcal{O}_{\mathcal{X}}$ , set  $p := \mu$ , and set  $k_C(\mathbf{g}, -) := \delta_{\mathbf{g}|_C}$  for all global outcome assignments  $\mathbf{g} \in \mathcal{O}_{\mathcal{X}}$ . This is by construction a deterministic hidden-variable model, which we claim gives rise to the empirical model  $e$ . Let us verify it.

<sup>2</sup> Note that, due to the assumption of parameter independence (Eq. (2.2)), we can unambiguously write  $k_x(\lambda, -)$  for  $k_C|_{\{x\}}(\lambda, -)$ , as this marginal is independent of the context  $C$  from which one is restricting.

Let  $C \in \mathcal{M}$  and write  $\rho_C^{\mathcal{X}}: \mathcal{O}_{\mathcal{X}} \rightarrow \mathcal{O}_C$  for the measurable projection which, in the event sheaf, is the restriction map  $\rho_C^{\mathcal{X}} = \mathcal{E}(C \subseteq \mathcal{X}): \mathcal{E}(\mathcal{X}) \rightarrow \mathcal{E}(C)$ . For any  $E \in \mathcal{F}_C$ , we have

$$k_C(\mathbf{g}, E) = \delta_{\mathbf{g}|_C}(E) = \delta_{\rho_C^{\mathcal{X}}(\mathbf{g})}(E) = \mathbf{1}_E(\rho_C^{\mathcal{X}}(\mathbf{g})) = (\mathbf{1}_E \circ \rho_C^{\mathcal{X}})(\mathbf{g}) \quad (2.4)$$

which will assign 1 if  $\rho_C^{\mathcal{X}}(\mathbf{g}) = \mathbf{g}|_C \in E$ . Then, as required

$$\int_{\mathbf{\Lambda}} k_C(-, E) \, dp \quad (2.5)$$

$$= \{ \mathbf{\Lambda} = \mathcal{O}_{\mathcal{X}}; p = \mu; k_C(-, E) = \mathbf{1}_E \circ \rho_C^{\mathcal{X}} \text{ by Eq. (2.4)} \}$$

$$\int_{\mathcal{O}_{\mathcal{X}}} \mathbf{1}_E \circ \rho_C^{\mathcal{X}} \, d\mu \quad (2.6)$$

$$= \{ \text{change of variables, Eq. (1.38)} \}$$

$$\int_{\mathcal{O}_C} \mathbf{1}_E \, d\rho_{C*}^{\mathcal{X}} \mu \quad (2.7)$$

$$= \{ \text{marginalisation for probability measures} \}$$

$$\int_{\mathcal{O}_C} \mathbf{1}_E \, d\mu|_C \quad (2.8)$$

$$= \{ \text{integral of indicator function, Eq. (1.37)} \}$$

$$\mu|_C(E) \quad (2.9)$$

$$= \{ \mu \text{ extends the empirical model } e \}$$

$$e_C(E) .$$

**(2)  $\Rightarrow$  (3).** It is enough to show that if a hidden-variable model  $\langle \mathbf{\Lambda}, p, k \rangle$  is deterministic then it is also factorisable. For this, it is sufficient to notice that a Dirac measure  $\delta_{\mathbf{o}}$  with  $\mathbf{o} \in \mathcal{O}_C$  on a product space  $\mathcal{O}_C = \prod_{x \in C} \mathcal{O}_x$  factorises as the product of Dirac measures

$$\delta_{\mathbf{o}} = \prod_{x \in C} \delta_{\mathbf{o}|_{\{x\}}} = \prod_{x \in C} \delta_{\rho_{\{x\}}^C(\mathbf{o})} , \quad (2.10)$$

as the projection map  $\rho_{\{x\}}^C$  is well-defined and continuous. Then for any  $\prod_{x \in C} B_x \in \mathcal{F}_C$  (such that  $B_x \subseteq \mathcal{O}_x$  only for a finite number of  $x \in C$ ) we have:

$$\delta_{\mathbf{o}} \left( \prod_{x \in C} B_x \right) = \prod_{x \in C} \delta_{\mathbf{o}|_{\{x\}}}(B_x) . \quad (2.11)$$

**(3)  $\Rightarrow$  (1).** Suppose that  $e$  is realised by a factorisable hidden-variable model  $\langle \mathbf{\Lambda}, p, k \rangle$ . Write  $k_x$  for  $k_C|_{\{x\}}$ . Define a measure  $\mu$  on  $\mathcal{O}_{\mathcal{X}}$  as follows: for any measurable set  $\prod_{x \in \mathcal{X}} E_x \in \mathcal{F}_{\mathcal{X}}$  where  $E_x \subseteq \mathcal{O}_x$  only for a finite number of  $x \in \mathcal{X}$ , the value of  $\mu$  on it is given by

$$\mu \left( \prod_{x \in \mathcal{X}} E_x \right) := \int_{\mathbf{\Lambda}} \left( \prod_{x \in \mathcal{X}} k_x(-, E_x) \right) \, dp = \int_{\lambda \in \mathbf{\Lambda}} \left( \prod_{x \in \mathcal{X}} k_x(\lambda, E_x) \right) \, d\rho(\lambda) , \quad (2.12)$$

where the product  $\prod$  on the right-hand side is of real numbers in the interval  $[0, 1]$ . Note that the  $\sigma$ -algebra of  $\mathcal{O}_{\mathcal{X}}$  is given by the product topology and it is generated by such measurable sets; hence the equation above uniquely determines  $\mu$  as a measure on  $\mathcal{O}_{\mathcal{X}}$ .

Now, we show that this is a global section for the empirical probabilities. Let  $C \in \mathcal{M}$  and consider a measurable set  $F = \prod_{x \in C} F_x \in \mathcal{F}_C$  with  $F_x \subseteq \mathcal{O}_x$  only for a finite number of  $x \in C$ . Then

$$\mu|_C(F) \tag{2.13}$$

$$= \{ \text{definition of marginalisation} \}$$

$$\mu(F \times \mathcal{O}_{\mathcal{X} \setminus C}) \tag{2.14}$$

$$= \{ \text{definition of } F \text{ and } \mathcal{O}_U \}$$

$$\mu \left( \prod_{x \in C} F_x \times \prod_{x \in \mathcal{X} \setminus C} \mathcal{O}_x \right) \tag{2.15}$$

$$= \{ \text{definition of } \mu, \text{ Eq. (2.12)} \}$$

$$\int_{\Lambda} \left( \prod_{x \in C} k_x(-, F_x) \right) \left( \prod_{x \in \mathcal{X} \setminus C} k_x(-, \mathcal{O}_x) \right) dp \tag{2.16}$$

$$= \{ k_x(\lambda, -) \text{ probability measure so } k_x(\lambda, \mathcal{O}_x) = 1 \}$$

$$\int_{\Lambda} \left( \prod_{x \in C} k_x(-, F_x) \right) dp \tag{2.17}$$

$$= \{ \text{factorisability of the hidden-variable model} \}$$

$$\int_{\Lambda} k_C(-, \prod_{x \in C} F_x) dp \tag{2.18}$$

$$= \{ \text{definition of } F \}$$

$$\int_{\Lambda} k_C(-, F) dp \tag{2.19}$$

$$= \{ e \text{ is the empirical model corresponding to } \langle \Lambda, p, k \rangle \}$$

$$e_C(F) .$$

Since the  $\sigma$ -algebra  $\mathcal{F}_C$  of  $\mathcal{O}_C$  is generated by such measurable sets  $F$  above and we have seen that  $\mu|_C$  agrees with  $e_C$  on these sets, we conclude that  $\mu|_C = e_C$  as required.  $\square$

### 2.3 Quantifying contextuality

Beyond questioning whether a given empirical behaviour is contextual or not, it is also interesting to ask to what *degree* it is contextual. In discrete-variable scenarios, a very natural

measure of contextuality is the contextual fraction [AB11] (see Section 1.5.3). This measure was shown in [ABM17] to have a number of very desirable properties. It can be calculated using linear programming, an approach that subsumes the more traditional approach to quantifying nonlocality and contextuality using Bell and noncontextual inequalities. Indeed we can understand the (dual) linear program as optimising over *all* such inequalities for the scenario in question and returning the maximum normalised violation of *any* Bell or noncontextuality inequality achieved by the given empirical model. Crucially, the contextual fraction was also shown to *quantifiably* relate to quantum-over-classical advantages in specific informatic tasks [ABM17, MK18]. Moreover it has been demonstrated to be a monotone with respect to the free operations of resource theories for contextuality [ABM17, DA18, ABKM19].

In this section, we consider how to carry those ideas to the continuous-variable setting. The formulation as a linear optimisation problem and the attendant correspondence with Bell inequality violations requires special care as one needs to use infinite-dimensional linear programming, necessitating some extra assumptions on the measurable outcome spaces.

### 2.3.1 The continuous-variable contextual fraction

Asking whether a given behaviour is noncontextual amounts to asking whether the empirical model is extendable, or in other words whether it admits a deterministic hidden-variable model. However, a more refined question to pose is:

*what fraction of the behaviour admits a deterministic hidden-variable model?*

This quantity is what we call the noncontextual fraction. Similarly, the fraction of the behaviour that is left over and that can thus be considered *irreducibly* contextual is what we call the contextual fraction.

**Definition 2.7** (Noncontextual fraction). *Let  $e$  be an empirical model on the measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ . The noncontextual fraction of  $e$ , written  $\text{NCF}(e)$ , is defined as*

$$\sup \{ \mu(\mathcal{O}_{\mathcal{X}}) \mid \mu \in \mathbb{M}(\mathcal{O}_{\mathcal{X}}), \forall C \in \mathcal{M}, \mu|_C \leq e_C \} . \quad (2.20)$$

*Note that since  $e_C$  is a probability measure on  $\mathcal{O}_C$  for all  $C \in \mathcal{M}$  it follows that  $\text{NCF}(e) \in [0, 1]$ . The contextual fraction of  $e$ , written  $\text{CF}(e)$ , is given by  $\text{CF}(e) := 1 - \text{NCF}(e)$ .*

This is the continuous-variable analogue of Definition 1.24.

### 2.3.2 Monotonicity under free operations including binning

In the discrete-variable setting, the contextual fraction was shown to be a monotone under a number of natural classical operations that transform and combine empirical models and control their use as resources, therefore constituting the ‘free’ operations of a resource theory of contextuality [ABM17, DA18, ABKM19].

All of the operations on empirical models defined for discrete variables in [ABM17] – viz. translations of measurements, transformation of outcomes, probabilistic mixing, product, and choice – carry almost verbatim to our current setting. One detail is that one must insist that the coarse-graining of outcomes be achieved by (a family of) measurable functions. A particular example of practical importance is *binning*, which is widely used in continuous-variable quantum information as a method of discretising data by partitioning the outcome space  $\mathcal{O}_x$  for each measurement  $x \in \mathcal{X}$  into a finite number of ‘bins’, i.e. measurable sets. Note that a binned empirical model is obtained by pushing forward along a family  $(t_x)_{x \in \mathcal{X}}$  of outcome translations  $t_x: \mathcal{O}_x \rightarrow \mathcal{O}'_x$  where  $\mathcal{O}'_x$  is finite for all  $x \in \mathcal{X}$ .

**Proposition 2.1.** *If  $e$  is an empirical model, and  $e^{bin}$  is any discrete-variable empirical model obtained from  $e$  by binning, then contextuality of  $e^{bin}$  witnesses contextuality of  $e$ , and quantifiably gives a lower bound  $\text{CF}(e^{bin}) \leq \text{CF}(e)$ .*

For the conditional measurement operation introduced in [ABKM19], which allows for adaptive measurement protocols such as those used in measurement-based quantum computation [RB01], one must similarly insist that the map determining the next measurement to perform based on the observed outcome of a previous measurement be a measurable function. Since we are, for now on, only considering scenarios with a finite number of measurements, this amounts to a partition of the outcome space  $\mathcal{O}_x$  of the first measurement,  $x$ , into measurable subsets labelled by measurements compatible with  $x$ , indicating which will be subsequently performed depending on the outcome observed for  $x$ .

The inequalities establishing monotonicity from [ABM17, Theorem 2] also hold for continuous variables. There is a caveat for the equality formula for the product of two empirical models:

$$\text{NCF}(e_1 \otimes e_2) = \text{NCF}(e_1)\text{NCF}(e_2).$$

Whereas the inequality establishing monotonicity ( $\geq$ ) still holds in general, the proof establishing the other direction ( $\leq$ ) makes use of duality of linear programs. Therefore, it only holds under the assumptions we will impose in the remainder of this section.

### 2.3.3 Assumptions on the outcome spaces

In order to phrase the problem of contextuality as a linear programming problem and establish the connection with violations of Bell inequalities, we need to impose some conditions on the measurable spaces of outcomes. First we suppose that we have a *finite number of measurement labels* i.e. that  $\mathcal{X}$  is finite.

From now on, we restrict attention to the case where the outcome space  $\mathcal{O}_x$  for each measurement  $x \in \mathcal{X}$  is the Borel measurable space for a *compact* Hausdorff space, i.e. that the set  $\mathcal{O}_x$  is a compact space and  $\mathcal{F}_x$  is the  $\sigma$ -algebra generated by its open sets, written  $\mathcal{B}(\mathcal{O}_x)$ . Note that this includes most situations of interest in practice. In particular, it includes the

case of measurements with *outcomes in a bounded subspace of  $\mathbb{R}$  or  $\mathbb{R}^n$* . This is experimentally motivated since measurement devices are *energetically bounded*.

One may wish to relax this assumption to allow for the case of a locally compact outcome space in order to include the case of measurements with outcomes in  $\mathbb{R}$  or  $\mathbb{R}^n$ . Indeed  $\mathbb{R}$  would be the canonical outcome space for the quadratures of the electromagnetic field. We address this issue in Subsection 2.3.6 and show that it can be reduced to the case of a compact outcome space. Thus the assumptions laid out here are not restrictive for the cases of interest, even theoretical ones. To summarise we essentially have the two following assumptions for the rest of this chapter:

- (i)  $\mathcal{X}$  is a finite set of measurement labels,
- (ii) for each  $x \in \mathcal{X}$ , the outcome set  $\mathcal{O}_x$  is compact.

To obtain an infinite-dimensional linear program, we need to work with vector spaces (see Section 1.4). However as explained in Section 1.3, probability measures, or even finite or arbitrary measures, do not form one. We will therefore consider the set  $\mathbb{M}_{\pm}(\mathbf{Y})$  of *finite signed measures* on a measurable space  $\mathbf{Y} = \langle Y, \mathcal{F}_Y \rangle$ . Recall that these are functions  $\mu: \mathcal{F}_Y \rightarrow \mathbb{R}$  such that  $\mu(\emptyset) = 0$  and  $\mu$  is  $\sigma$ -additive. The set  $\mathbb{M}_{\pm}(\mathbf{Y})$  forms a real vector space which includes the probability measures  $\mathbb{P}(\mathbf{Y})$ . When  $Y$  is a compact Hausdorff space and  $\mathbf{Y} = \langle Y, \mathcal{B}(Y) \rangle$ , the Riesz–Markov–Kakutani representation theorem [Kak41] says that  $\mathbb{M}_{\pm}(\mathbf{Y})$  is a concrete realisation of the topological dual space of  $C(Y)$ , the space of continuous real-valued functions on  $Y$ . The duality is given by  $\langle \mu, f \rangle := \int_Y f \, d\mu$  for  $\mu \in \mathbb{M}_{\pm}(\mathbf{Y})$  and  $f \in C(Y)$ .

### 2.3.4 Infinite-dimensional linear programming for computing the contextual fraction

Consider an empirical model  $e = (e_C)_{C \in \mathcal{M}}$  on a scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  satisfying the assumptions discussed above. Calculation of its noncontextual fraction can be expressed as the infinite-dimensional linear programming problem (P-CF<sup>CV</sup>). This is our primal linear program; its dual linear program is given by (D-CF<sup>CV</sup>). In what follows, we will see how to derive the dual and show that the optimal values of both programs coincide.

$$\text{(P-CF}^{\text{CV}}) \quad \left\{ \begin{array}{l} \text{Find } \mu \in \mathbb{M}_{\pm}(\mathcal{O}_{\mathcal{X}}) \\ \text{maximising } \mu(\mathcal{O}_{\mathcal{X}}) \\ \text{subject to:} \\ \quad \forall C \in \mathcal{M}, \mu|_C \leq e_C \\ \quad \mu \geq 0. \end{array} \right.$$

$$\text{(D-CF}^{\text{CV}}) \quad \left\{ \begin{array}{l} \text{Find } (f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(\mathcal{O}_C) \\ \text{minimising } \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d\epsilon_C \\ \text{subject to:} \\ \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \geq \mathbf{1}_{\mathcal{O}_X} \\ \forall C \in \mathcal{M}, f_C \geq \mathbf{0}_{\mathcal{O}_C} . \end{array} \right.$$

We have written  $\rho_C^X$  for the projection  $\mathcal{O}_X \rightarrow \mathcal{O}_C$  as before, and  $\mathbf{1}_D$  (resp.  $\mathbf{0}_D$ ) for the constant function  $D \rightarrow \mathbb{R}$  that assigns the number 1 (resp. 0) to all elements of its domain  $D$ ; in the above instance, to all  $\mathbf{g} \in \mathcal{O}_X$  (resp. all  $\mathbf{o} \in \mathcal{O}_C$ ).

Analogues of these programs have been studied in the discrete-variable setting [ABM17]. Note however that, in general, these continuous-variable linear programs are over infinite-dimensional spaces and thus not practical to compute directly. For this reason, in Section 2.5 we will introduce a hierarchy of finite-dimensional semidefinite programs that approximate the solution of (P-CF<sup>CV</sup>) to arbitrary precision.

### Retrieving the standard form of infinite-dimensional LPs

Here we express the programs (P-CF<sup>CV</sup>) and (D-CF<sup>CV</sup>) in the standard form of infinite-dimensional linear programs as introduced in Section 1.4.1 following [Bar02]. We define the following spaces:

- $E_1 := \mathbb{M}_{\pm}(\mathcal{O}_X)$ .
- $F_1 := C(\mathcal{O}_X)$ , the dual space of  $E_1$ .
- $E_2 := \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\mathcal{O}_C)$ .
- $F_2 := \prod_{C \in \mathcal{M}} C(\mathcal{O}_C)$ , the dual space of  $E_2$ .

The dualities  $\langle -, - \rangle_1 : E_1 \times F_1 \rightarrow \mathbb{R}$  and  $\langle -, - \rangle_2 : E_2 \times F_2 \rightarrow \mathbb{R}$  are defined as follows:

$$\begin{aligned}
 \forall \mu \in E_1, \forall f \in F_1, \quad \langle \mu, f \rangle_1 &:= \int_{\mathcal{O}_X} f \, d\mu \\
 \forall (\nu_C) \in E_2, \forall (f_C) \in F_2, \quad \langle (\nu_C), (f_C) \rangle_2 &:= \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d\nu_C,
 \end{aligned}$$

where, for simplicity, we have omitted  $C \in \mathcal{M}$  as a subscript for the families of functions. We fix  $K_1$  to be the convex cone of positive measures in  $E_1 = \mathbb{M}_{\pm}(\mathcal{O}_X)$  and  $K_2$  to be the convex cone of families of positive measures in  $E_2 = \prod_{C \in \mathcal{M}} \mathbb{M}_{\pm}(\mathcal{O}_C)$ . Then  $K_1^*$  is the convex cone of positive function in  $F_1 = C(\mathcal{O}_X)$  and  $K_2^*$  is the convex cone of families of positive functions in  $F_2 = \prod_{C \in \mathcal{M}} C(\mathcal{O}_C)$ .

Let  $A: E_1 \rightarrow E_2$  be the following linear transformation:

$$\forall \mu \in E_1, \quad A(\mu) := (\mu|_C)_{C \in \mathcal{M}} \in E_2 .$$

We also define the linear transformation  $A^*: F_2 \rightarrow F_1$  as:

$$\forall (f_C) \in F_2, \quad A^*((f_C)) := \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \in F_1 .$$

We can verify that  $A^*$  is the dual transformation of  $A$ :  $\forall \mu \in E_1, \forall (f_C) \in F_2$ , we have

$$\langle A(\mu), (f_C) \rangle_2 = \langle (\mu|_C), (f_C) \rangle_2 \quad (2.21)$$

$$= \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d\mu|_C \quad (2.22)$$

$$= \int_{\mathcal{O}_X} \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \, d\mu \quad (2.23)$$

$$= \langle \mu, \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \rangle_1 \quad (2.24)$$

$$= \langle \mu, A^*((f_C)) \rangle_1 . \quad (2.25)$$

Now fixing the vector function in the objective to be  $c := -\mathbf{1}_{\mathcal{O}_X} \in F_1$  and the vector in the constraints to be  $b := -(e_C)_{C \in \mathcal{M}} \in E_2$ , the program (P-CF<sup>CV</sup>) (resp. (D-CF<sup>CV</sup>)) can be expressed as in the standard form given in (LP) (resp. (D-LP)). Note that the minus sign in the vectors  $c$  and  $b$  was added because we choose the primal program in the standard form to be a minimisation problem while the primal program at hand is a maximisation problem.

### Deriving the dual via the Lagrangian

We now give an explicit derivation of (D-CF<sup>CV</sup>) as the dual of (P-CF<sup>CV</sup>) via the Lagrangian method which is widely used in optimisation theory. We do not take into account positivity constraints as constraints in the Lagrangian because they are already satisfied in the convex cones in which one optimises. To derive the dual program we introduce  $|\mathcal{M}|$  dual variables—one continuous map  $f_C \in C(\mathcal{O}_C)$  for each  $C \in \mathcal{M}$ —to account for the constraints  $\mu|_C \leq e_C$ . We then define the Lagrangian  $\mathcal{L}: K_1 \times K_2^* \rightarrow \mathbb{R}$  as

$$\mathcal{L}(\mu, (f_C)) := \underbrace{\mu(\mathcal{O}_X)}_{\text{objective}} + \underbrace{\sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d(e_C - \mu|_C)}_{\text{constraints}} . \quad (2.26)$$

The primal program (P-CF<sup>CV</sup>) corresponds to

$$\sup_{\mu \in K_1} \inf_{(f_C) \in K_2^*} \mathcal{L}(\mu, (f_C)) . \quad (2.27)$$

The infimum here imposes the constraints that  $\mu|_C \leq e_C$  for all  $C \in \mathcal{M}$ , for otherwise the associated optimisation problem is not well-defined (or rather it necessarily has a value of

minus infinity). If these constraints are satisfied, then because of the infimum over  $(f_C) \in K_2^*$ , the second term of the Lagrangian vanishes yielding, as expected, the objective of the primal problem. Expressing the dual problem amounts to permuting the infimum and the supremum. Thus we need to rewrite the Lagrangian as:

$$\mathcal{L}(\mu, (f_C)) = \mu(\mathcal{O}_X) + \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C d(e_C - \mu|_C) \quad (2.28)$$

$$= \int_{\mathcal{O}_X} \mathbf{1}_{\mathcal{O}_X} d\mu + \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C de_C - \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C d\mu|_C \quad (2.29)$$

$$= \int_{\mathcal{O}_X} \mathbf{1}_{\mathcal{O}_X} d\mu + \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C de_C - \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_X} f_C \circ \rho_C^X d\mu \quad (2.30)$$

$$= \int_{\mathcal{O}_X} \mathbf{1}_{\mathcal{O}_X} d\mu + \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C de_C - \int_{\mathcal{O}_X} \left( \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \right) d\mu \quad (2.31)$$

$$= \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C de_C + \int_{\mathcal{O}_X} \left( \mathbf{1}_{\mathcal{O}_X} - \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \right) d\mu. \quad (2.32)$$

where Eq. (2.30) is obtained by the push-forward operation and Eq. (2.31) is obtained by linearity of the integral (and the finiteness of  $\mathcal{M}$ ). The dual program (D-CF<sup>CV</sup>) indeed corresponds to

$$\inf_{(f_C) \in K_2^*} \sup_{\mu \in K_1} \mathcal{L}(\mu, (f_C)). \quad (2.33)$$

The supremum imposes that  $\sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \geq \mathbf{1}_{\mathcal{O}_X}$  on  $\mathcal{O}_X$  since otherwise the Lagrangian diverges. If this constraint is satisfied, then because of the supremum the second term vanishes yielding, as expected, the objective of the dual problem.

### 2.3.5 Zero duality gap

A key result about the noncontextual fraction, which is essential in establishing the connection to Bell inequality violations, is that (P-CF<sup>CV</sup>) and (D-CF<sup>CV</sup>) are strongly dual, in the sense that no gap exists between their optimal values i.e.  $\text{val}(\text{P-CF}^{\text{CV}}) = \text{val}(\text{D-CF}^{\text{CV}})$ . Strong duality holds in finite linear programming when Slater's conditions are met—which will be the case for examples of interest. However it does not hold in general in the infinite-dimensional case. Below we prove a strong duality result for our programs.

**Proposition 2.2.** *Problems (P-CF<sup>CV</sup>) and (D-CF<sup>CV</sup>) have zero duality gap. Thus their optimal values satisfy:*

$$\text{val}(\text{P-CF}^{\text{CV}}) = \text{val}(\text{D-CF}^{\text{CV}}) = \text{NCF}(e) \quad (2.34)$$

and there exists a primal optimal plan for the primal program (P-CF<sup>CV</sup>).

*Proof.* This proof relies on [Bar02, Theorem 7.2] (recalled in Section 1.4.1, Theorem 1.1). Because  $\mu_0 = \mathbf{0}_{\mathcal{O}_X}$ —the measure that assigns 0 to every measurable set of  $\mathcal{O}_X$ —is a feasible

solution for (P-CF<sup>CV</sup>) and the noncontextual fraction lies between 0 and 1, (P-CF<sup>CV</sup>) is consistent with finite value. Thus it suffices to show that the following cone

$$\mathcal{K} = \{(A(\mu), \langle \mu, c \rangle_1) : \mu \in K_1\} = \{((\mu|_C)_C, \mu(\mathcal{O}_X)) : \mu \in K_1\} \quad (2.35)$$

is weakly closed in  $E_2 \oplus \mathbb{R}$  (i.e. closed in the weak topology of  $K_1$ ) where we recall that  $K_1$  is the convex cone of positive measures in  $E_1 = \mathbb{M}_\pm(\mathcal{O}_X)$ .

We first notice that the linear transformation  $A$  is a bounded linear operator and thus continuous. Boundedness comes from the fact that,  $\forall \mu \in K_1$ ,

$$\|A(\mu)\|_{E_2} = \|(\mu|_C)_C\|_{E_2} \quad (2.36)$$

$$= \sum_{C \in \mathcal{M}} \|\mu|_C\|_{\mathbb{M}_\pm(\mathcal{O}_C)} \quad (2.37)$$

$$\leq \sum_{C \in \mathcal{M}} \|\mu\|_{E_1} \quad (2.38)$$

$$= |\mathcal{M}| \|\mu\|_{E_1}, \quad (2.39)$$

where we take the strong topology—i.e. the norm induced by the total variation distance—on finite-signed measure spaces. It is defined as:

$$\|\mu\|_{\mathbb{M}_\pm(U)} = |\mu|(U).$$

We also equip the finite product space  $E_2 = \prod_{C \in \mathcal{M}} \mathbb{M}_\pm(\mathcal{O}_C)$  with the norm obtained by summing<sup>3</sup> the individual total variation norms. Eq. (2.38) is due to the fact that  $\mu \in K_1$  so this is a positive measure and thus  $\|\mu|_C\|_{\mathbb{M}_\pm(\mathcal{O}_C)} \leq \|\mu\|_{E_1}$ . This, of course, extends to the weak topology.

Secondly, we consider a sequence  $(\mu^k)_{k \in \mathbb{N}}$  in  $K_1$  and we want to show that the accumulation point  $((\Theta_C)_C, \lambda) = \lim_{k \rightarrow \infty} (A(\mu^k), \langle \mu^k, c \rangle_1)$  belongs to  $\mathcal{K}$ , where  $\Theta = (\Theta_C)_C \in E_2$  and  $\lambda \in \mathbb{R}$ . If we consider the product of indicator functions  $(\mathbf{1}_{\mathcal{O}_C})_C \in F_2$  then  $\langle A(\mu^k), (\mathbf{1}_{\mathcal{O}_C}) \rangle_2 = \sum_{C \in \mathcal{M}} \mu^k|_C(\mathcal{O}_C) \xrightarrow{k} \sum_{C \in \mathcal{M}} \Theta_C(\mathcal{O}_C) < \infty$  as  $\Theta_C$  is a finite measure for all maximal contexts  $C \in \mathcal{M}$ . Then because  $\mathcal{M}$  is a covering family of  $\mathcal{X}$ ,  $\forall k \in \mathbb{N}$ ,  $\mu^k(\mathcal{O}_X) \leq \sum_{C \in \mathcal{M}} \mu^k|_C(\mathcal{O}_C) < \infty$ . Since  $(\mu^k) \in K_1^{\mathbb{N}}$  is a sequence of positive measures, this implies that  $(\mu^k)$  is bounded. Next, by weak-\* compactness of the unit ball (Alaoglu's theorem [Lue97]), there exists a subsequence  $(\mu^{k_i})_{k_i}$  that converges weakly to an element  $\omega \in K_1$ . By continuity of  $A$ , it yields that the accumulation point is such that  $((\Theta_C)_C, \lambda) = (A(\omega), \langle \omega, c \rangle_1) \in \mathcal{K}$ .  $\square$

### 2.3.6 The case of local compactness

We now focus on cases where the outcome space might be only locally compact which include most theoretical situations that are of interest in practice; for instance  $\mathbb{R}$  could be the outcome space for the position and momentum operators.

<sup>3</sup>Categorically, this is a coproduct.

For each measurement  $x \in \mathcal{X}$ ,  $\mathcal{O}_x$  is supposed to be the Borel measurable space for a second-countable locally compact Hausdorff space, i.e. that the set  $\mathcal{O}_x$  is equipped with a second-countable locally compact Hausdorff topology and  $\mathcal{F}_x$  is still the  $\sigma$ -algebra generated by its open sets, written  $\mathcal{B}(\mathcal{O}_x)$ . Second countability and Hausdorffness of two spaces  $Y$  and  $Z$  suffice to show that the Borel  $\sigma$ -algebra of the product topology is the tensor product of the Borel  $\sigma$ -algebras, i.e.  $\mathcal{B}(Y \times Z) = \mathcal{B}(Y) \otimes \mathcal{B}(Z)$  [Bog07, Lemma 6.4.2 (Vol. 2)]. Hence, these assumptions guarantee that  $\mathcal{O}_U$  for  $U \in \mathcal{P}(\mathcal{X})$  is the Borel  $\sigma$ -algebra of the product topology on  $\mathcal{O}_U = \prod_{x \in U} \mathcal{O}_x$ . These product spaces are also second-countable, locally compact, and Hausdorff as all three properties are preserved by finite products. When  $Y$  is a second-locally compact Hausdorff space and  $\mathbf{Y} = \langle Y, \mathcal{B}(Y) \rangle$ , the Riesz–Markov–Kakutani representation theorem [Kak41] says that  $\mathbb{M}_{\pm}(\mathbf{Y})$  is a concrete realisation of the topological dual space of  $C_0(Y)$ , the space of continuous real-valued functions on  $Y$  that vanish at infinity.<sup>4</sup> The duality is given by  $\langle \mu, f \rangle := \int_{\mathbf{Y}} f \, d\mu$  for  $\mu \in \mathbb{M}_{\pm}(\mathbf{Y})$  and  $f \in C_0(Y)$ <sup>5</sup>. Note that when  $Y$  is compact (as treated above), we only need to consider functions in  $C(Y)$ .

Next, we show that we can approximate the linear program (P-CF<sup>CV</sup>)<sup>6</sup> by a slightly modified linear program defined on the space of finite measures on a measurable compact subspace of  $\mathcal{O}_{\mathcal{X}}$ . The idea is to approximate to any desired error the mass of a finite measure on a locally compact set by the mass of the same measure on a compact subset. This naturally comes from the notion of *tightness* of a measure.

**Definition 2.8** (Tightness of a measure). *A measure  $\mu$  on a metric space  $U$  is said to be tight if for each  $\epsilon > 0$  there exists a compact set  $U_{\epsilon} \subseteq U$  such that  $\mu(U \setminus U_{\epsilon}) < \epsilon$ .*

Then we need to argue that every measure we will consider is tight. This is a result of the following theorem:

**Theorem 2.2** ([Par67]). *If  $S$  is a complete separable metric space, then every finite measure on  $S$  is tight.*

For  $x \in \mathcal{X}$ ,  $\mathcal{O}_x$  is a second-countable locally compact Hausdorff space, thus a Polish space i.e. a separable completely metrisable topological space. For this reason, the above theorem applies. We are now ready to state and prove the main theorem of this subsection.

**Theorem 2.3.** *The linear program (P-CF<sup>CV</sup>) defined over finite-signed measures on a locally compact space can be approximated to any desired precision  $\epsilon$  by a linear program (P-CF<sup>CV, $\epsilon$</sup> ) defined over finite signed measures on a compact space.*

<sup>4</sup>A function  $f: Y \rightarrow \mathbb{R}$  on a locally compact space  $Y$  is said to *vanish at infinity* if the set  $\{y \in Y \mid \|f(x)\| \geq \epsilon\}$  is compact for all  $\epsilon > 0$ .

<sup>5</sup>This theorem holds more generally for locally compact Hausdorff spaces if one considers only (finite signed) Radon measures, which are measures that play well with the underlying topology. However, second-countability, together with local compactness and Hausdorffness, guarantees that every Borel measure is Radon [Fol84, Theorem 7.8].

<sup>6</sup>Here we will still use the form of the program (P-CF<sup>CV</sup>) though throughout this subsection one has to keep in mind that it is defined over finite-signed measures on a *locally compact* space rather than a compact space.

*Proof.* Fix  $\epsilon > 0$ . Let  $C \in \mathcal{M}$  be a given context and  $x \in C$  a given measurement label within that context. Because  $e_C$  is a probability measure on  $\mathcal{O}_C$ , the marginal measure  $e_C|_{\{x\}}$  is a finite measure on  $\mathcal{O}_x$ . Following Theorem 2.2,  $e_C|_{\{x\}}$  is tight and there exists a compact subset  $K_x^{\epsilon, C} \subseteq \mathcal{O}_x$  such that:  $e_C|_{\{x\}}(\mathcal{O}_x \setminus K_x^{\epsilon, C}) \leq \epsilon$ . We apply this procedure<sup>7</sup> for every context and for all measurements in a context. We now define the compact set:

$$\mathcal{O}_x^\epsilon := \bigcup_{C|x \in C} K_x^{\epsilon, C}.$$

The previous definition is essential to ensure a noncontextual cutoff of the outcome set which, in turn, ensures the good definition of a compact subset for each measurement label independent of the context. For some subset of measurement labels  $U \subseteq \mathcal{X}$ , we define the compact set  $\mathcal{O}_U^\epsilon := \prod_{x \in U} \mathcal{O}_x^\epsilon$ . For every context  $C \in \mathcal{M}$  and for every measurement label  $x \in C$ , we now have that  $K_x^{\epsilon, C} \subseteq \mathcal{O}_x^\epsilon$  and thus  $e_C|_{\{x\}}(\mathcal{O}_x \setminus \mathcal{O}_x^\epsilon) \leq \epsilon$ . Note that due to the compatibility condition, we can write  $e_C|_{\{x\}}$  as  $e_{\{x\}}$  for any context.

Let  $\mu$  be any feasible solution of (P-CF<sup>CV</sup>) defined over finite-signed measures on a locally compact space. Due to the constraints of (P-CF<sup>CV</sup>) we have that  $\forall x \in \mathcal{X}, \mu|_{\{x\}} \leq e_{\{x\}}$ . Then:

$$\mu(\mathcal{O}_{\mathcal{X}} \setminus \mathcal{O}_{\mathcal{X}}^\epsilon) = \mu \left( \prod_{x \in \mathcal{X}} \mathcal{O}_x \setminus \prod_{x \in \mathcal{X}} \mathcal{O}_x^\epsilon \right) \quad (2.40)$$

$$= \mu \left( \prod_{x \in \mathcal{X}} (\mathcal{O}_x \setminus \mathcal{O}_x^\epsilon) \right) \quad (2.41)$$

$$= \prod_{x \in \mathcal{X}} \mu|_{\{x\}}(\mathcal{O}_x \setminus \mathcal{O}_x^\epsilon) \quad (2.42)$$

$$\leq \prod_{x \in \mathcal{X}} e_{\{x\}}(\mathcal{O}_x \setminus \mathcal{O}_x^\epsilon) \quad (2.43)$$

$$\leq \epsilon^{|\mathcal{X}|}. \quad (2.44)$$

We now define the linear program (P-CF<sup>CV,ε</sup>) which has the same form as (P-CF<sup>CV</sup>) though the unknown measures are taken from  $\mathbb{M}_\pm(\mathcal{O}_{\mathcal{X}}^\epsilon)$  where  $\mathcal{O}_{\mathcal{X}}^\epsilon = \langle \mathcal{O}_{\mathcal{X}}^\epsilon, \mathcal{B}(\mathcal{O}_{\mathcal{X}}^\epsilon) \rangle$ . We would like to state that (P-CF<sup>CV,ε</sup>) approximates (P-CF<sup>CV</sup>) up to  $\epsilon$ ; i.e. that their values are  $\epsilon$ -close. The missing ingredient from the previous chain of inequalities is that given an optimal measure  $\mu^*$  satisfying (P-CF<sup>CV</sup>), we do not know whether an optimal solution  $\mu_\epsilon^*$  of (P-CF<sup>CV,ε</sup>) is necessarily the restriction of  $\mu^*$  to  $\mathcal{O}_{\mathcal{X}}^\epsilon$ . In fact, it is possible that we do not even have a unique optimal solution. However we only need to prove that they have the same mass on  $\mathcal{O}_{\mathcal{X}}^\epsilon$ , i.e.  $\mu_\epsilon^*(\mathcal{O}_{\mathcal{X}}^\epsilon) = \mu^*|_{\mathcal{O}_{\mathcal{X}}^\epsilon}(\mathcal{O}_{\mathcal{X}}^\epsilon)$ . By contradiction, suppose it does not hold. Then because  $\mu_\epsilon^*$  is an optimal value of (P-CF<sup>CV,ε</sup>), we must have  $\mu_\epsilon^*(\mathcal{O}_{\mathcal{X}}^\epsilon) > \mu^*|_{\mathcal{O}_{\mathcal{X}}^\epsilon}(\mathcal{O}_{\mathcal{X}}^\epsilon)$ . From this we construct a new measure  $\tilde{\mu}$  on  $\mathcal{O}_{\mathcal{X}}$  which equals to  $\mu_\epsilon^*$  on  $\mathcal{O}_{\mathcal{X}}^\epsilon$  and  $\mu^*$  on  $\mathcal{O}_{\mathcal{X}} \setminus \mathcal{O}_{\mathcal{X}}^\epsilon$ . It satisfies all constraints

<sup>7</sup>Note that there exists proofs that explicitly constructs the approximating sets  $K_x^{\epsilon, C}$  (see [Orb11]) based on the separability of the underlying spaces. It makes this construction feasible in practice and justifies this approach.

and furthermore  $\tilde{\mu}(\mathcal{O}_X) > \mu^*(\mathcal{O}_X)$ . This contradicts that  $\mu^*$  is an optimal solution of (P-CF<sup>CV</sup>). Thus necessarily  $\mu_\varepsilon^*(\mathcal{O}_X^\varepsilon) = \mu^*|_{\mathcal{O}_X^\varepsilon}(\mathcal{O}_X^\varepsilon)$ .

The linear program (P-CF<sup>CV,ε</sup>) defined on a compact space has indeed a value  $\varepsilon$ -close to the original program (P-CF<sup>CV</sup>).  $\square$

In conclusion of this subsection, we can approximate the problem of finding the noncontextual fraction with the assumption that the outcome spaces are locally compact spaces by the same problem defined on compact subspaces and it suffices to restrict the study to the case of compact outcome spaces. Thus for the rest of this chapter, we indeed restrict ourselves to compact outcome spaces.

## 2.4 Continuous generalisation of Bell inequalities

The dual program (D-CF<sup>CV</sup>) is of particular interest in its own right. As we now show, it can essentially be understood as computing a continuous-variable ‘Bell inequality’ that is optimised to the empirical data. Making the change of variables  $\beta_C := |\mathcal{M}|^{-1}\mathbf{1}_{\mathcal{O}_C} - f_C \in C(\mathcal{O}_C)$  for each  $C \in \mathcal{M}$ , the dual program (D-CF<sup>CV</sup>) transforms to:

$$(B-CF^{CV}) \quad \left\{ \begin{array}{l} \text{Find } (\beta_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(\mathcal{O}_C) \\ \text{maximising } \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} \beta_C \, de_C \\ \text{subject to:} \\ \sum_{C \in \mathcal{M}} \beta_C \circ \rho_C^X \leq \mathbf{0}_{\mathcal{O}_X} \\ \forall C \in \mathcal{M}, \beta_C \leq |\mathcal{M}|^{-1}\mathbf{1}_{\mathcal{O}_C} . \end{array} \right.$$

This program directly computes the contextual fraction  $CF(e)$  instead of the noncontextual fraction. It maximises, subject to constraints, the total value obtained by integrating these functionals context-wise against the empirical model in question. The first set of constraints—a generalisation of a system of linear inequalities determining a Bell inequality—ensures that, for noncontextual empirical models, the value of the program is at most 0, since any such model extends to a measure  $\mu$  on  $\mathcal{O}_X$  such that  $\mu(\mathcal{O}_X) = 1$ . The final set of constraints acts as a normalisation condition on the value of the program, ensuring that it takes values in the interval  $[0, 1]$  for any empirical model. Any family of functions  $\beta = (\beta_C) \in F_2$  satisfying the constraints will thus result in what can be regarded as a generalised Bell inequality,

$$\sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} \beta_C \, de_C \leq 0, \quad (2.45)$$

which is satisfied by all noncontextual empirical models.

**Definition 2.9** (Form on a measurement scenario). *A form  $\beta$  on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  is a family  $\beta = (\beta_C)_{C \in \mathcal{M}}$  of functions  $\beta_C \in C(\mathcal{O}_C)$  for all  $C \in \mathcal{M}$ . Given an empirical model  $e$  on  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ , the value of  $\beta$  on  $e$  is*

$$\langle e, \beta \rangle_2 := \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} \beta_C \, \mathrm{d}e_C . \quad (2.46)$$

The norm of  $\beta$  is given by

$$\|\beta\| := \sum_{C \in \mathcal{M}} \|\beta_C\| = \sum_{C \in \mathcal{M}} \sup \{ \beta_C(\mathbf{o}) \mid \mathbf{o} \in \mathcal{O}_C \} . \quad (2.47)$$

**Definition 2.10** (Inequality). *An inequality  $(\beta, R)$  on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  is given by  $\langle \beta, R \rangle$  where  $\beta$  is a form on  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  together with a bound  $R \in \mathbb{R}$ . An empirical model  $e$  is said to satisfy the inequality  $\langle \beta, R \rangle$  if the value of  $\beta$  on  $e$  is below the bound, i.e.  $\langle e, \beta \rangle_2 \leq R$ .*

**Definition 2.11** (Continuous Bell inequality). *An inequality  $(\beta, R)$  is said to be a continuous Bell inequality if it is satisfied by all noncontextual empirical models, i.e. if for any noncontextual model  $d$  on  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ , it holds that  $\langle d, \beta \rangle_2 \leq R$ .*

A continuous Bell inequality  $(\beta, R)$  establishes a bound  $\langle e, \beta \rangle_2$  amongst noncontextual models  $e$ . For more general models, the value of  $\beta$  on  $e$  is only limited by the algebraic bound  $\|\beta\|$ . A model  $e$  reaching  $\|\beta\|$  is likely to be signalling. In the following, we will only consider inequalities  $(\beta, R)$  for which  $R < \|\beta\|$  excluding inequalities trivially satisfied by all empirical models.

**Definition 2.12** (Normalised violation of a continuous Bell inequality). *The normalised violation of a generalised Bell inequality  $(\beta, R)$  by an empirical model  $e$  is*

$$\frac{\max\{0, \langle e, \beta \rangle_2 - R\}}{\|\beta\| - R} , \quad (2.48)$$

*the amount by which its value  $\langle \beta, e \rangle_2$  exceeds the bound  $R$  normalised by the maximal ‘algebraic’ violation.*

An optimal plan  $\beta \in F_2$  for (B-CF<sup>CV</sup>) defines a continuous Bell inequality  $\langle \beta, 0 \rangle$  that is maximally violated by the empirical model  $e$ . Note however that an optimal solution might not be feasible—for instance the optimum might be achieved by a discontinuous function—in which case there will exist a sequence of feasible solutions converging to such a discontinuous function.

The above definition restricts to the usual notions of Bell inequality and noncontextual inequality in the discrete-variable case and is particularly close to the presentation in [ABM17]. The following theorem also generalises to continuous variables the main result of [ABM17] (see Theorem 1.6).

**Theorem 2.4.** *Let  $e$  be an empirical model on a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$ . (i) The normalised violation by  $e$  of any Bell inequality is at most  $\text{CF}(e)$ ; (ii) if  $\text{CF}(e) > 0$  then for every  $\varepsilon > 0$  there exists a Bell inequality whose normalised violation by  $e$  is at least  $\text{CF}(e) - \varepsilon$ .*

Item (ii) is slightly modified compared to the discrete analogue version because there is no guarantee that there exists an optimal plan for the dual program (D-CF<sup>CV</sup>). In particular, its optimal solution might be achieved by a discontinuous function that can be approximated by continuous ones. Hence the modification of (ii) with a normalised violation  $\varepsilon$ -close to  $\text{CF}(e)$ .

*Proof.* As in the discrete-variable case (see [ABM17] for a detailed proof), the proof follows directly from the definitions of the linear programs, and from strong duality, i.e. the fact that their optimal values coincide (Proposition 2.2 above).  $\square$

## 2.5 Approximating the contextual fraction with SDPs

In Section 2.3, we presented the problem of computing the noncontextual fraction as an infinite-dimensional linear program. Although of theoretical importance, it cannot be run to perform the actual, numerical computation of this quantity. Here we exploit the link between measures and their sequence of moments to derive a hierarchy of truncated finite-dimensional semidefinite programs which are a relaxation of the original primal problem (P-CF<sup>CV</sup>). Dual to this vision, we can equivalently exploit the link between positive polynomials and their sum-of-squares representation to derive a hierarchy of semidefinite programs which are a restriction of the dual problem (D-CF<sup>CV</sup>). We further prove that the optimal values of the truncated programs converge monotonically to the noncontextual fraction. This makes use of global optimisation techniques developed by Lasserre and Parrilo [Las09, Par03] and further developed in [Las11]. These techniques were introduced in Subsection 1.4.3 and we will use the same notations throughout this section. Another extensive and well presented reference on the subject is [Lau09]. In Subsection 2.5.1, we derive a hierarchy of SDPs to approximate the contextual fraction and we show its convergence in Subsection 2.5.2.

### 2.5.1 Hierarchy of semidefinite relaxations for computing $\text{NCF}(e)$

We fix a measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  and an empirical model  $e$  on this scenario. We will restrict our attention to outcome spaces of the form detailed in subsection 2.3.3. Let  $d = |\mathcal{X}| \in \mathbb{N}^*$  so that  $\mathcal{O}_{\mathcal{X}}$  is a Borel subset of  $\mathbb{R}^d$ . As a prerequisite, we first need to compute the sequences of moments associated to measures  $(e_C)_{C \in \mathcal{M}}$  derived from the empirical model. For  $C \in \mathcal{M}$ , let  $\mathbf{y}^{e,C} = (y_{\alpha}^{e,C})_{\alpha \in \mathbb{N}^d}$  be the sequence of all moments of  $e_C$ . For a given  $k \in \mathbb{N}$  which will fix the level of the hierarchy, we only need to compute a finite number  $s(k)$  of moments for all contexts. These will be the inputs of the program.

Below, we derive a hierarchy of SDP relaxations for the primal program (P-CF<sup>CV</sup>) such that their optimal values converge monotonically to  $\text{val}(\text{P-CF}^{\text{CV}}) = \text{NCF}(e)$ . We start by discussing the assumptions we have to make on the outcome space. Then we derive the hierarchy based on, first, the primal program and then the dual program and we further show that these formulations are indeed dual. Finally we prove the convergence of the hierarchy.

### Further assumptions on the outcome space?

We already made the assumptions mentioned in Subsection 2.3.3 for the outcome spaces  $\mathcal{O} = (\mathcal{O}_x)_{x \in \mathcal{X}}$  noting that they are not restrictive when considering actual experimental or theoretical applications. However we would like to meet the assumptions detailed in Assumption 1.1 for the global outcome space  $\mathcal{O}_{\mathcal{X}}$  so that both Theorems 1.2 and 1.4 apply in our setting.

Assumption 1.1 (ii) is de facto met because we already assumed that for all  $x \in \mathcal{X}$ ,  $\mathcal{O}_x \subset \mathbb{R}$  is compact. Remember that, at worst,  $\mathcal{O}_x$  is locally compact and we saw that this can be reduced to the compact case in Subsection 2.3.6.

Let us discuss Assumption 1.1 (i). We have that  $\mathcal{O}_{\mathcal{X}} = \prod_{x \in \mathcal{X}} \mathcal{O}_x$  with  $\mathcal{O}_x \subset \mathbb{R}$  compact. If  $\mathcal{O}_x$  is disconnected, we can always complete it into a connected space by attributing a measure zero to the added parts for all measures  $e_C$  whenever  $x \in C$ . Then because  $\mathcal{O}_x$  is compact, it is bounded and it can be described by two constant polynomials: there exists  $a_x, b_x \in \mathbb{R}$  such that  $\mathcal{O}_x = [a_x, b_x]$ . This makes  $\mathcal{O}_{\mathcal{X}}$  a polytope so in particular, it is semi-algebraic. We write it as

$$\mathcal{O}_{\mathcal{X}} = \left\{ \mathbf{x} \in \mathbb{R}^d \mid \forall j = 1, \dots, m, g_j(\mathbf{x}) \geq 0 \right\} \quad (2.49)$$

for some polynomials  $g_j \in \mathbb{R}[\mathbf{x}]$  of degree 1.

As noted in [Las09], Assumption 1.1 (iii) is not very restrictive. For instance, it is satisfied when the set is a polytope. This is the case for  $\mathcal{O}_{\mathcal{X}}$ .

Thus there is no need for further assumptions than what we already assumed in Subsection 2.3.3 to apply the main results we presented in Section 1.4.3.

### Relaxation of the primal program

The program (P-CF<sup>CV</sup>) can be relaxed so that a converging hierarchy of SDPs can be derived. The program (P-CF<sup>CV</sup>) is essentially a maximisation problem on finite-signed Borel measures with additional constraints such as the fact that these are proper measures (i.e. they are nonnegative). We will represent a measure by its moment sequence and use conditions for which this moment sequence has a (unique) representing Borel measure (see Subsection 1.4.3.2). We

recall the expression of the primal program (P-CF<sup>CV</sup>):

$$\text{(P-CF}^{\text{CV}}) \quad \left\{ \begin{array}{l} \text{Find } \mu \in \mathbb{M}_{\pm}(\mathcal{O}_{\mathcal{X}}) \\ \text{maximising } \mu(\mathcal{O}_{\mathcal{X}}) \\ \text{subject to:} \\ \quad \forall C \in \mathcal{M}, \mu|_C \leq e_C \\ \quad \mu \geq 0 . \end{array} \right.$$

From Subsection 1.4.3.2 which culminates at Theorem 1.4, it can be relaxed for  $k \in \mathbb{N}^*$  as:

$$\text{(PS-CF}^{\text{CV},k}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} \in \mathbb{R}^{s(2k)} \\ \text{maximising } y_0 \\ \text{subject to:} \\ \quad \forall C \in \mathcal{M}, M_k(\mathbf{y}^{e,C} - \mathbf{y}|_C) \succeq 0, \\ \quad \forall j = 1, \dots, m, M_{k-1}(g_j \mathbf{y}) \succeq 0, \\ \quad M_k(\mathbf{y}) \succeq 0 . \end{array} \right.$$

We consider localising matrices of order  $k - 1$  rather than  $k$  because all  $g_j$ 's are of exact degree 1. In this way, the maximum degree with the moment matrices matches. In general we have to deal with localising matrices of order  $k - \lceil \frac{\deg(g_j)}{2} \rceil$ . If  $\mu$  is a representing measure on  $\mathcal{O}_{\mathcal{X}}$  for  $\mathbf{y}$  then for all contexts  $C \in \mathcal{M}$ ,  $\mathbf{y}|_C$  can be defined through  $\mathbf{y}$  by requiring that  $\mathbf{y}|_C$  has representing measure  $\mu|_C$ . The two last constraints state necessary conditions on the variable  $\mathbf{y}$  to be moments of some finite Borel measure supported on  $\mathcal{O}_{\mathcal{X}}$ . The first constraint is a relaxation of the constraint  $\mu|_C \leq e_C$  for  $C \in \mathcal{M}$ . As expected, (PS-CF<sup>CV,k</sup>) is a semidefinite relaxation of problem (P-CF<sup>CV</sup>) so that  $\forall k \in \mathbb{N}^*$ ,  $\text{NCF}(e) = \text{val}(\text{LP-CF}^{\text{CV}}) \leq \text{val}(\text{PS-CF}^{\text{CV},k})$ . Moreover  $(\text{val}(\text{PS-CF}^{\text{CV},k}))_k$  form a monotone nonincreasing sequence because more constraints are added as  $k$  increases (so that the relaxations are tighter and tighter).

### Restriction of the dual program

The program (D-CF<sup>CV</sup>) can be restricted so that we can derive a converging hierarchy of SDPs. It is essentially the minimisation of continuous functions for which we require additional constraints such as the fact that they are nonnegative. We will exploit the link between positive polynomials and sum-of-squares representation that was presented in Subsection 1.4.3.1. We

recall the expression of the dual program (D-CF<sup>CV</sup>):

$$(D-CF^{CV}) \quad \left\{ \begin{array}{l} \text{Find } (f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(\mathcal{O}_C) \\ \text{minimising } \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, d\epsilon_C \\ \text{subject to:} \\ \sum_{C \in \mathcal{M}} f_C \circ \rho_C^X \geq \mathbf{1}_{\mathcal{O}_X} \\ \forall C \in \mathcal{M}, f_C \geq \mathbf{0}_{\mathcal{O}_C} . \end{array} \right.$$

As this point we could derive the dual of program (PS-CF<sup>CV,k</sup>) and show that this is indeed a restriction of the above program. For a more symmetric treatment, we restrict the dual program building on Subsection 1.4.3.1 and Theorem 1.2. Instead of optimising over positive continuous functions, we restrict them to belong to the quadratic module  $Q(g)$  and then  $Q_k(g)$  for some  $k \in \mathbb{N}^*$  further requiring that the degrees of SOS polynomials are fixed. For  $k \in \mathbb{N}^*$ , we have

$$(DS-CF^{CV,k}) \quad \left\{ \begin{array}{l} \text{Find } (p_C)_{C \in \mathcal{M}} \subset \Sigma^2\mathbb{R}[\mathbf{x}]_k \text{ and } (\sigma_j)_{j=1,\dots,m} \subset \Sigma^2\mathbb{R}[\mathbf{x}]_{k-1} \\ \text{maximising } \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} p_C \, d\epsilon_C \\ \text{subject to:} \\ \sum_{C \in \mathcal{M}} p_C \circ \rho_C^X - \mathbf{1}_{\mathcal{O}_X} = \sum_{j=0}^m \sigma_j g_j . \end{array} \right.$$

(DS-CF<sup>CV,k</sup>) is a restriction of (D-CF<sup>CV</sup>) so that for all  $k \in \mathbb{N}^*$ , we have that  $\text{NCF}(e) = \text{val}(D-CF^{CV}) \leq \text{val}(DS-CF^{CV,k})$ . Furthermore,  $(\text{val}(DS-CF^{CV,k}))_k$  form a monotone nonincreasing sequence.

### Proving duality for the semidefinite programs

As mentioned above, we chose to derive programs (PS-CF<sup>CV,k</sup>) and (DS-CF<sup>CV,k</sup>) using dual arguments. These programs should therefore be dual to one another, which will immediately provide weak duality. We prove this for completeness.

**Proposition 2.3.** *The program (DS-CF<sup>CV,k</sup>) is the dual formulation of the program (PS-CF<sup>CV,k</sup>).*

*Proof.* We start by rewriting  $M_k(\mathbf{y})$  as  $\sum_{\alpha \in \mathbb{N}_k^d} \mathbf{y}_\alpha A_\alpha$  and  $M_{k-1}(g_j \mathbf{y})$  as  $\sum_{\alpha \in \mathbb{N}_k^d} y_\alpha B_\alpha^j$  for  $1 \leq j \leq m$  and for appropriate real symmetric matrices  $A_\alpha$  and  $(B_\alpha^j)_j$ . For instance, in the basis  $(\mathbf{x}^\alpha)$ :

$$(A_\alpha)_{s,t} = \left( \begin{array}{c} 1 \text{ if } \mathbf{s} + \mathbf{t} = \alpha \\ 0 \text{ otherwise} \end{array} \right)_{s,t} .$$

From  $A_\alpha$ , we also extract  $A_\alpha^C$  for  $C \in \mathcal{M}$  in order to rewrite  $M_k(\mathbf{y}|_C)$  as  $\sum_{\alpha \in \mathbb{N}_k^d} y_\alpha A_\alpha^C$ . This amounts to identifying which matrices  $(A_\alpha)$  contribute to a given context  $C \in \mathcal{M}$ . Then (PS-CF<sup>CV,k</sup>) can be rewritten as:

$$(\text{SDP-CF}^{\text{CV},k}) \left\{ \begin{array}{l} \text{Find } \mathbf{y} \in \mathbb{R}^{s(2k)} \\ \text{maximising } y_0 \\ \text{subject to:} \\ \forall C \in \mathcal{M}, M_k(\mathbf{y}^{e,C}) - \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha A_\alpha^C \succeq 0, \\ \forall j = 1, \dots, m, \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha B_\alpha^j \succeq 0, \\ \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha A_\alpha \succeq 0. \end{array} \right.$$

Via the Lagrangian, this is equivalent to:

$$\sup_{\mathbf{y} \in \mathbb{R}^{s(k)}} \inf_{\substack{(X^C), (Y_j), Z \\ \text{SDP matrices}}} \mathcal{L}(\mathbf{y}, (X_C), Y, (Z_j)), \quad (2.50)$$

with

$$\begin{aligned} \mathcal{L}(\mathbf{y}, (X^C), (Y_j), Z) &= y_0 \\ &+ \sum_{C \in \mathcal{M}} \text{Tr}(M_k(\mathbf{y}^{e,C}) X^C) - \sum_{C \in \mathcal{M}} \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \text{Tr}(A_\alpha^C X^C) \\ &+ \sum_{j=1}^m \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \text{Tr}(B_\alpha^j Y_j) \\ &+ \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \text{Tr}(A_\alpha Z). \end{aligned} \quad (2.51)$$

The dual program corresponds to

$$\inf_{\substack{(X^C), (Y_j), Z \\ \text{SDP matrices}}} \sup_{\mathbf{y} \in \mathbb{R}^{s(k)}} \mathcal{L}(\mathbf{y}, (X^C), (Y_j), Z). \quad (2.52)$$

We rewrite the Lagrangian as:

$$\begin{aligned} \mathcal{L}(\mathbf{y}, (X_C), Y, (Z_j)) &= \sum_{C \in \mathcal{M}} \text{Tr}(M_k(\mathbf{y}^{e,C}) X^C) \\ &+ \sum_{\alpha \in \mathbb{N}_k^d} y_\alpha \left( \delta_{\alpha, \mathbf{0}} - \sum_{C \in \mathcal{M}} \text{Tr}(A_\alpha^C X^C) + \sum_{j=1}^m \text{Tr}(B_\alpha^j Y_j) + \text{Tr}(A_\alpha Z) \right). \end{aligned} \quad (2.53)$$

Then the dual program of (PS-CF<sup>CV,k</sup>) reads:

$$\left\{ \begin{array}{l} \text{Find } (X^C)_{C \in \mathcal{M}}, (Y_j)_{j=1, \dots, m} \text{ and } Z \text{ SDP matrices} \\ \text{maximising } \sum_{C \in \mathcal{M}} \text{Tr}(M_k(\mathbf{y}^{e,C})X^C) \\ \text{subject to:} \\ \forall \boldsymbol{\alpha} \in \mathbb{N}_k^d, \sum_{C \in \mathcal{M}} \text{Tr}(A_\alpha^C X^C) - \sum_{j=1}^m \text{Tr}(B_\alpha^j Y_j) - \text{Tr}(A_\alpha Z) = \delta_{\boldsymbol{\alpha}, \mathbf{0}}. \end{array} \right.$$

We finally show that the above program exactly corresponds to (DS-CF<sup>CV,k</sup>). We start with the objective. For all  $C \in \mathcal{M}$ , with  $X^C$  a positive semidefinite matrix:

$$\text{Tr}(M_k(\mathbf{y}^{e,C})X^C) = \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} (M_k(\mathbf{y}^{e,C}))_{\boldsymbol{\alpha}\boldsymbol{\beta}} (X^C)_{\boldsymbol{\beta}\boldsymbol{\alpha}} \quad (2.54)$$

$$= \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} \mathbf{y}_{\boldsymbol{\alpha}+\boldsymbol{\beta}}^{e,C} (X^C)_{\boldsymbol{\alpha}\boldsymbol{\beta}} \quad (2.55)$$

$$= \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\beta}} \int_{\mathcal{O}_C} \mathbf{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \text{de}_C(X^C)_{\boldsymbol{\alpha}\boldsymbol{\beta}} \quad (2.56)$$

$$= \int_{\mathcal{O}_C} \mathbf{v}_k(\mathbf{x})^T X^C \mathbf{v}_k(\mathbf{x}) \text{de}_C \quad (2.57)$$

$$= \int_{\mathcal{O}_C} p_C \text{de}_C, \quad (2.58)$$

with for all  $C \in \mathcal{M}$ ,  $p_C \in \Sigma^2 \mathbb{R}[\mathbf{x}]_k$  a sum-of-squares polynomial via Proposition 1.1 and where we used  $\mathbf{v}_k(\mathbf{x})$  the vectors of monomials of maximal total degree  $k$ .

Now, to retrieve the constraint, we multiply each side by  $\mathbf{x}^\alpha$  and we sum for all  $\boldsymbol{\alpha}$ :

$$\sum_{C \in \mathcal{M}} \text{Tr}(\sum_{\boldsymbol{\alpha}} \mathbf{x}^\alpha A_\alpha^C X^C) - \mathbf{1} = \sum_{j=1}^m \text{Tr}(\sum_{\boldsymbol{\alpha}} \mathbf{x}^\alpha B_\alpha^j Y_j) + \text{Tr}(\sum_{\boldsymbol{\alpha}} \mathbf{x}^\alpha A_\alpha Z) \quad (2.59)$$

Recalling the definition of moment and localising matrices:

$$\sum_{\boldsymbol{\alpha}} A_\alpha^C \mathbf{x}^\alpha = \mathbf{v}_k(\mathbf{x}) \mathbf{v}_k(\mathbf{x})^T \quad (2.60)$$

$$\sum_{\boldsymbol{\alpha}} B_\alpha^j \mathbf{v}_k(\mathbf{x}) \mathbf{v}_k(\mathbf{x})^T = g_j(\mathbf{x}) \mathbf{v}_{k-1}(\mathbf{x}) \mathbf{v}_{k-1}(\mathbf{x})^T, \quad \forall j = 1, \dots, m \quad (2.61)$$

Thus, by Proposition 1.1, for appropriate sum-of-squares polynomials  $(\sigma_j)_{j=0,1,\dots,m} \subset \mathbb{R}[\mathbf{x}]_{k-1}$ :

$$\text{Tr}(\sum_{\boldsymbol{\alpha}} \mathbf{x}^\alpha A_\alpha^C X^C) = p_C \circ \rho_C^X(\mathbf{x}) \quad (2.62)$$

$$\text{Tr}(\sum_{\boldsymbol{\alpha}} \mathbf{x}^\alpha B_\alpha^j Y_j) = g_j(\mathbf{x}) \sigma_j(\mathbf{x}) \quad (2.63)$$

$$\text{Tr}(\sum_{\boldsymbol{\alpha}} \mathbf{x}^\alpha A_\alpha Z) = \sigma_0(\mathbf{x}) \quad (2.64)$$

This is exactly the constraint in (DS-CF<sup>CV,k</sup>) with, for convenience,  $g_0 = 1$ .

□

### 2.5.2 Convergence of the hierarchy of SDPs

Finally, we prove that the constructed hierarchy provides a sequence of objective values that converges monotonically to the noncontextual fraction  $\text{NCF}(e)$ .

**Theorem 2.5.** *The optimal values of the hierarchy of semidefinite programs (PS-CF<sup>CV,k</sup>) (resp. (DS-CF<sup>CV,k</sup>)) provide monotonically decreasing upper bounds converging to the noncontextual fraction  $\text{NCF}(e)$  which is the value of (P-CF<sup>CV</sup>). That is*

$$\text{val}(\text{PS-CF}^{\text{CV},k}) \downarrow \text{val}(\text{P-CF}^{\text{CV}}) = \text{NCF}(e) \quad \text{as } k \rightarrow \infty, \quad (2.65)$$

$$\text{val}(\text{DS-CF}^{\text{CV},k}) \downarrow \text{val}(\text{D-CF}^{\text{CV}}) = \text{NCF}(e) \quad \text{as } k \rightarrow \infty. \quad (2.66)$$

*Proof.* Because of the strong duality between the original infinite-dimensional linear programs we have:

$$\text{val}(\text{P-CF}^{\text{CV}}) = \text{val}(\text{D-CF}^{\text{CV}}) = \text{NCF}(e). \quad (2.67)$$

Moreover,  $\forall k \geq 1$ , (PS-CF<sup>CV,k</sup>) is a relaxation of (P-CF<sup>CV</sup>):

$$\text{val}(\text{PS-CF}^{\text{CV},k}) \geq \text{val}(\text{P-CF}^{\text{CV}}). \quad (2.68)$$

And  $\forall k \geq 1$ , (DS-CF<sup>CV,k</sup>) is a restriction of (D-CF<sup>CV</sup>):

$$\text{val}(\text{DS-CF}^{\text{CV},k}) \geq \text{val}(\text{D-CF}^{\text{CV}}). \quad (2.69)$$

Also  $\forall k \geq 1$ , we have weak duality between (PS-CF<sup>CV,k</sup>) and (DS-CF<sup>CV,k</sup>) (by Proposition 2.3):

$$\text{val}(\text{DS-CF}^{\text{CV},k}) \geq \text{val}(\text{PS-CF}^{\text{CV},k}). \quad (2.70)$$

Thus for all  $k \geq 1$ :

$$\text{val}(\text{DS-CF}^{\text{CV},k}) \geq \text{val}(\text{PS-CF}^{\text{CV},k}) \geq \text{NCF}(e). \quad (2.71)$$

We already saw that  $(\text{val}(\text{PS-CF}^{\text{CV},k}))_k$  and  $(\text{val}(\text{DS-CF}^{\text{CV},k}))_k$  form monotone nonincreasing sequences. We now show that  $(\text{val}(\text{DS-CF}^{\text{CV},k}))_k$  converges to  $\text{NCF}(e)$ . This is equivalent to showing that we can approximate any feasible solution<sup>8</sup> of program (D-CF<sup>CV</sup>) with a solution of (DS-CF<sup>CV,k</sup>) for a high enough rank  $k$ .

Fix  $\varepsilon > 0$  and a feasible solution  $(f_C)_{C \in \mathcal{M}} \in \prod_{C \in \mathcal{M}} C(\mathcal{O})$  of (D-CF<sup>CV</sup>). Then  $\forall C \in \mathcal{M}$ ,  $f_C + \frac{\varepsilon}{|\mathcal{M}|}$  is a positive continuous function on  $\mathcal{O}_C$ . Because  $\mathcal{O}_C$  is compact (see Subsection 2.3.3) by the Stone-Weierstrass theorem,  $f_C + \frac{\varepsilon}{|\mathcal{M}|}$  can be approximated by a positive polynomial. Thus there exist positive polynomials  $p_C^\varepsilon \in \mathbb{R}[\mathbf{x}]$  such that for all contexts  $C \in \mathcal{M}$  (in sup norm):

$$\left\| f_C + \frac{\varepsilon}{|\mathcal{M}|} - p_C^\varepsilon \right\| \leq \frac{\varepsilon}{|\mathcal{M}|} \quad (2.72)$$

<sup>8</sup>Note that program (D-CF<sup>CV</sup>) might not have an optimal solution in which case it only has an optimal solution in the closure of the feasible set. In that case, we can always find a sequence of feasible solutions converging to an optimal solution.

and also:

$$\left\| \left( f_C + \frac{\varepsilon}{|\mathcal{M}|} - p_C^\varepsilon \right) \circ \rho_C^\mathcal{X} \right\| < \frac{1}{|\mathcal{M}|} \min_{\mathbf{x} \in \mathcal{O}_\mathcal{X}} \left( \sum_{C \in \mathcal{M}} \left( f_C + \frac{\varepsilon}{|\mathcal{M}|} \right) \circ \rho_C^\mathcal{X}(\mathbf{x}) - \mathbf{x} \right) \quad (2.73)$$

where the minimum is strictly positive as  $\sum_C (f_C + \frac{\varepsilon}{|\mathcal{M}|}) \circ \rho_C^\mathcal{X} > \mathbf{1}_{\mathcal{O}_\mathcal{X}}$ .

From Eq. (2.72), the objective derived with  $(p_C^\varepsilon)_C$  is  $\varepsilon$ -close to the original objective:

$$\left| \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} f_C \, de_C - \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} p_C^\varepsilon \, de_C \right| \leq \sum_{C \in \mathcal{M}} \int_{\mathcal{O}_C} \left| f_C + \frac{\varepsilon}{|\mathcal{M}|} - p_C^\varepsilon \right| \, de_C \quad (2.74)$$

$$\leq \varepsilon. \quad (2.75)$$

Also from Eq. (2.73):

$$\sum_{C \in \mathcal{M}} p_C^\varepsilon \circ \rho_C^\mathcal{X} - \mathbf{1} \quad (2.76)$$

$$> \sum_{C \in \mathcal{M}} \left( f_C + \frac{\varepsilon}{|\mathcal{M}|} \right) \circ \rho_C^\mathcal{X} - \min_{\mathbf{x} \in \mathcal{O}_\mathcal{X}} \left( \sum_{C \in \mathcal{M}} \left( f_C + \frac{\varepsilon}{|\mathcal{M}|} \right) \circ \rho_C^\mathcal{X}(\mathbf{x}) - \mathbf{x} \right) - \mathbf{1} \quad (2.77)$$

$$\geq 0, \quad (2.78)$$

so that  $\sum_{C \in \mathcal{M}} p_C^\varepsilon \circ \rho_C^\mathcal{X} - \mathbf{1}_{\mathcal{O}_\mathcal{X}}$  is a positive polynomial on  $\mathcal{O}_\mathcal{X}$ . Next, because  $\mathcal{O}_\mathcal{X}$  is of the form required in Assumption 1.1, by Putinar's Positivellensatz (see Theorem 1.2),  $\sum_{C \in \mathcal{M}} p_C^\varepsilon \circ \rho_C^\mathcal{X} - \mathbf{1}_{\mathcal{O}_\mathcal{X}}$  belongs to the quadratic module  $Q(g)$ . Therefore, for a high enough rank  $k \in \mathbb{N}$ , it is a feasible solution of  $(\text{DS-CF}^{\text{CV},k})$  and thus:

$$|\text{NCF}(e) - \text{val}(\text{DS-CF}^{\text{CV},k})| \leq \varepsilon. \quad (2.79)$$

□

## 2.6 Discussion and open problems

In this chapter, we have presented a framework formalising measurement contextuality for continuous-variable systems. We further extended the FAB theorem to those kind of scenarios. We also studied how to quantify contextuality via the extension of the contextual fraction to the continuous-variable realm. Its computation can be phrased as an infinite-dimensional linear program that we can relax to derive a converging hierarchy of finite dimensional semidefinite programs.

*Logical* forms of contextuality, which are present at the level of the *possibilistic* rather than *probabilistic* information contained in an empirical model, remain to be considered (e.g. in discrete-variable scenarios [Fri09, AB11, Abr13b, MF12]). In the discrete setting, these can be treated by analysing ‘possibilistic’ empirical models obtained by considering the supports of the discrete-variable probability distributions [AB11], which indicate the elements of an outcome

space that occur with non-zero probability. It can often be easily pictured on a bundle diagram. However the notion of support of a measure is not as straightforward, and the naïve approach is not viable since typically all singletons have measure 0. Nevertheless, supports can be defined in the setting of Borel measurable spaces, for instance, which in any case are the kind of spaces in which we are practically interested in Sections 2.3 and 2.5.

Approaches to contextuality that characterise obstructions to global sections using cohomology have had considerable success [AMB12, ABK<sup>+</sup>15, Car15, Car17, Rau16, Rou17, ORBR17, Car18, OTR18, Man13] and typically apply to logical forms of contextuality. An interesting prospect is to explore how the present framework may be employed to these ends, and to see whether the continuous-variable setting can open the door to new techniques that can be applied, or whether qualitatively new forms of contextual behaviour may be uncovered. A related direction to be developed is to understand how our treatment of contextuality can be further extended to continuous measurement spaces as proposed in [TC19].

Another direction to be explored is how our continuous-variable framework for contextuality can be extended to apply to more general notions of contextuality that relate not only to measurement contexts but also more broadly to contexts of preparations and transformations as well [Spe05, MK18], noting that these also admit quantifiable relationships to quantum advantage e.g. [MK18, HCB<sup>+</sup>18]. Indeed, a major motivation to study contextuality is for its connections to quantum-over-classical advantages in informatic tasks. An important line of questioning is to ask what further connections can be found in the continuous-variable setting, and whether continuous-variable contextuality might offer advantages that outstrip those achievable with discrete-variable contextual resources. Note that it is known that infinite-dimensional quantum systems can offer certain additional advantages beyond finite-dimensional ones [Slo16], though the empirical model that arises in that example is still a discrete-variable one in our sense.

The present work sets the theoretical basis for computational exploration of continuous-variable contextuality in quantum-mechanical empirical models. This, we hope, can provide new insights and inform all other avenues to be developed in future work. It can also be useful in verifying the non-classicality of empirical models. Numerical implementation of the programs of Section 2.5 is of particular interest. The hierarchy of semidefinite programs can be used numerically to witness contextuality in continuous-variable experiments. Even if the time-complexity of the semidefinite program may increase drastically with its degree, a low-degree program can already provide a first witness of contextual behaviour as it will provide a lower bound on the contextual fraction.

Since our framework for continuous-variable contextuality is independent of quantum theory itself, it can equally be applied to ‘empirical models’ that arise in other, non-physical settings. The discrete-variable framework of [AB11] has led to a number of surprising connections and cross-fertilisations with other fields [Abr15], including natural language [AS14b],

relational databases [Abr13a, Bar15], logic [AH12, ABK<sup>+</sup>15, Kis16], constraint satisfaction [AGK13, ABdSZ17] and social systems [DZK16]. It may be hoped that similar connections and applications can be found for the present framework to fields in which continuous-variable data is of central importance. For instance probability kernels of the kind we have used are also widely employed in machine learning (e.g. [HSS08]), inviting intriguing questions about how our framework might be used or what advantages contextuality may confer in that setting.

## EQUIVALENCE OF WIGNER NEGATIVITY AND CONTEXTUALITY FOR CONTINUOUS-VARIABLE PAULI MEASUREMENTS

FOR discrete-variable systems of odd power-of-a-prime dimension, Howard *et al.* [HWVE14] showed that negativity of the Wigner function actually corresponds to contextuality with respect to Pauli measurements (with an additional ancilla system), thereby establishing the operational utility of contextuality for the gate-based model of quantum computation—particularly in a fault-tolerant setting. The equivalence of Wigner negativity and contextuality was established by deriving a noncontextuality inequality using the graph-theoretic techniques of Cabello, Severini and Winter [CSW14] which extends Kochen-Specker type state-independent proofs to the state-dependent realm. CSW inequalities have been shown to be equivalent to the logical Bell inequalities appearing in the sheaf theoretic approach [AB11] in [dS17]. This proof of equivalence [HWVE14] and subsequent alternate proof requires [DAGBR15, DOBV<sup>+</sup>17] that, as well as the system displaying Wigner negativity, a second ancillary system must be present in order to have a sufficiently rich set of available *Pauli measurements*. The equivalence was generalised to odd dimensions in [DOBV<sup>+</sup>17] and also established for qubit systems [RBD<sup>+</sup>17, DAGBR15].

Still the Wigner function and the phase-space formulation associated were initially introduced for continuous-variable systems and, to date, there is no link between Wigner negativity and contextuality in the continuous-variable setting. In [BB87, BG20] it is proven that the Wigner function is the *unique* phase-space quasiprobability distribution yielding the correct marginals for every quadrature. However the link to contextuality remains unclear as it is delicate to

exhibit the right measurement scenario [BW98]. In fact, generalising the discrete-variable approach to the sheaf-theoretic framework for continuous-variable contextuality runs into several problems of a functional-analytic nature. For example, since we only assume access to measurable properties of the quantum system, already the results of [BG20] are of no use since they implicitly assume point wise equality of marginals whereas we can only guarantee equality almost-everywhere. On the other hand, we assume a minimal structure in the nature of our model. In fact, assuming only the (pre-)sheaf structure for a set of measurements on quantum systems, we prove that there is a natural hidden-variable model which relates to the Wigner function on the phase space. This allows us to properly establish that Wigner negativity is equivalent to contextuality with respect to generalised quadrature measurements i.e. continuous-variable Pauli measurements. These are the most commonly used measurements in continuous-variable quantum information, in particular in quantum optics [ARL14, Wal21]. Our result directly relies on continuous-variable contextuality as presented in the previous chapter which generalises the well-established notion presented in [AB11].

We start by properly defining the measurement scenario we are focusing on in Section 3.1. We provide the suitable empirical models and crucially prove that they are family of probability measures on linear functions rather than any functions in Section 3.2. Then we prove the main theorem in Section 3.3 and conclude with a few open questions in Section 3.4. This chapter is based on [BCE21].

### 3.1 Measurement scenario under consideration

Hereafter we fix  $M \in \mathbb{N}^*$  to be the number of modes—that is,  $M$  continuous-variable systems—and  $\mathcal{H}$  the corresponding Hilbert space. As introduced in Section 1.2, the phase space for a single continuous-variable mode is  $V = \mathbb{R}^2$ . The Hilbert space associated with  $M$  modes is  $V^M$ . This a  $2M$ -dimensional symplectic  $\mathbb{R}$ -linear space. See [SSM88] for a concise introduction to the symplectic structure of the phase space and [DG06] for a detailed review.

From Subsection 1.2.3, the Wigner function representing the total system is a real-valued quasiprobability distribution  $V^M \rightarrow \mathbb{R}$ . It was already established in [ADVT09] that the Wigner function must be at least a  $L^1$ -function for the generalised quadrature measurements to provide a good characterisation. The proof strategy is to use the extendability property of a suitable noncontextual empirical model from Definition 2.3 to exhibit a global probability measure on global value assignments. Then we show that it corresponds to the Wigner function. This step is delicate since it requires careful functional-analytic considerations. Having achieved this, it proves that the Wigner function must be everywhere nonnegative since it corresponds to a global probability measure. For the other direction, a nonnegative Wigner function can be used directly as a hidden-variable model so that the corresponding empirical model is noncontextual [SKK+09].

We recall some background on the symplectic structure of  $V^M$  (see Subsection 1.2.3). The symplectic form is denoted  $\Omega: V^M \times V^M \rightarrow \mathbb{R}$ . For  $\mathbf{r}_1, \mathbf{r}_2 \in V^M$ ,

$$\Omega(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{r}_1 \cdot J\mathbf{r}_2 \quad \text{where} \quad J = \begin{pmatrix} 0 & \mathbb{1}_M \\ -\mathbb{1}_M & 0 \end{pmatrix} \quad (3.1)$$

We have that  $J^{-1} = J^T = -J$  and of course  $J$  can be seen as linear map from  $V^M$  to  $V^M$ . A Lagrangian vector subspace is a maximal isotropic subspace, that is, a maximal subspace on which the symplectic form  $\Omega$  vanishes. For a symplectic space of dimension  $2M$ , Lagrangian subspaces are of dimension  $M$ . The Lagrangian Grassmannian is the set of all Lagrangian subspaces.

### Measurement scenario

The Wigner function bears a close relationship to displacement operators as emphasised via its link with the characteristic function in Eq. (1.29). Wigner negativity will be shown to be equivalent to contextuality with respect to Pauli measurements in the same spirit as [HWVE14, DOBV<sup>+</sup>17]. Following Definition 2.1, this measurement scenario corresponds to the setting described below.

**Definition 3.1.** *We fix the measurement scenario  $\langle \mathcal{X}, \mathcal{M}, \mathcal{O} \rangle$  as follows:*

- *the set of measurement labels is  $\mathcal{X} := V^M$ ;*
- *the maximal contexts are Lagrangian subspaces of  $V^M$  so that the set of maximal contexts  $\mathcal{M}$  is the Lagrangian Grassmannian of  $\mathcal{X}$ ;*
- *for each  $\mathbf{x} \in \mathcal{X}$ ,  $\mathcal{O}_{\mathbf{x}} := \langle \mathbb{R}, \mathcal{B}_{\mathbb{R}} \rangle$  so that for any set of measurement labels  $U \subseteq \mathcal{X}$ ,  $\mathcal{O}_U \cong \mathbb{R}^U$  can be seen as the set of functions from  $U$  to  $\mathbb{R}$  with its product  $\sigma$ -algebra  $\mathcal{F}_U^1$ .*

Each  $\mathbf{x} = (q_1, \dots, q_M, p_1, \dots, p_M) \in \mathcal{X}$  specifies a point in phase-space which corresponds to measuring the associated displacement operator  $\hat{D}(\mathbf{x}) = \hat{D}(q_1, p_1) \otimes \dots \otimes \hat{D}(q_M, p_M)$ . Since a displacement operator is not self-adjoint (i.e. Hermitian) we detail below what the precise meaning of such a “measurement” is.

### Measuring a displacement operator

We will focus on a single mode since it can be straightforwardly extended to multimode quantum states. A quadrature operator such as the position operator  $\hat{q}$  (see Subsection 1.2.2) is self-adjoint and it can be expanded via the spectral theorem [Hal13] as:

$$\hat{q} = \int_{x \in \text{sp}(\hat{q})} x \, dP_{\hat{q}}(x), \quad (3.2)$$

---

<sup>1</sup>It is generated by collection of functions  $E$  from  $L$  to  $\mathbb{R}$  such that  $\pi_{\mathbf{x}}(E)$  is a real interval for a finite number of  $\mathbf{x} \in \mathcal{X}$  and  $\mathbb{R}$  for the rest where  $\pi_{\mathbf{x}}$  is the projection given later in Eq (3.13).

where the spectrum of  $\hat{q}$  is  $\text{sp}(\hat{q}) = \mathbb{R}$  and  $P_{\hat{q}}$  is the spectral measure of  $\hat{q}$  [Hal13, Th. 8.10]. For  $E \in \mathcal{B}(\mathbb{R})$ ,  $P_{\hat{q}}(E)$  is given by [Hal13, Def. 8.8]:

$$P_{\hat{q}}(E) = \mathbf{1}_E(\hat{q}) . \quad (3.3)$$

Informally, it assigns 1 whenever measuring  $\hat{q}$  yields an outcome that belongs to  $E$ . We can view  $P_{\hat{q}}(E)$  as the formal version of the projector  $\int_{x \in E} |x\rangle\langle x| dx$  (with  $|x\rangle$  a non-normalisable eigenvector of  $\hat{q}$ ). Its functional calculus [Hal13] can be expressed as:

$$f(\hat{q}) = \int_{x \in \text{sp}(\hat{q})} f(x) dP_{\hat{q}}(x) , \quad (3.4)$$

for  $f$  a bounded measurable function. Then we can write the spectral measure of  $f(\hat{q})$  via the push-forward operation:

$$\forall E \in \mathcal{B}(\mathbb{R}), P_{f(\hat{q})}(E) = P_{\hat{q}}(f^{-1}(E)) . \quad (3.5)$$

It follows immediately that, for  $s \in \mathbb{R}$ , the spectral measure for the diagonal phase operator  $e^{is\hat{q}}$  is given, for  $E \in \mathcal{B}(\mathbb{S}_1)$ , by

$$P_{\exp(is\hat{q})}(E) := P_{\hat{q}}(\{x \in \mathbb{R} \mid e^{isx} \in E\}) . \quad (3.6)$$

Define the rotated quadrature  $\hat{q}_\theta := \cos(\theta)\hat{q} + \sin(\theta)\hat{p}$  for  $\theta \in [0, 2\pi]$  and the phase-shift operator  $\hat{R}(\theta) := \exp(i\frac{\theta}{2}(\hat{q}^2 + \hat{p}^2))$ . Then

$$\hat{q}_\theta = \hat{R}(\theta)\hat{q}\hat{R}(-\theta) , \quad (3.7)$$

so that the spectral measure of  $\hat{q}_\theta$  is given by

$$P_{\hat{q}_\theta}(E) = \hat{R}(\theta)P_{\hat{q}}(E)\hat{R}(-\theta) . \quad (3.8)$$

Let  $(q, p) \in \mathbb{R}^2$ . We can find  $r \in \mathbb{R}_+$ ,  $\theta \in [0, 2\pi]$  such that  $(q, p) = (-r \sin(\theta), r \cos(\theta))$ . Then:

$$\hat{D}(q, p) = e^{i(p\hat{q} - q\hat{p})} = e^{ir(\cos(\theta)\hat{q} - \sin(\theta)\hat{p})} = e^{ir\hat{q}_\theta} . \quad (3.9)$$

This form allows us to deduce spectral measures for the displacement operators. For any  $E \in \mathcal{B}(\mathbb{S}_1)$ , we have

$$P_{\hat{D}(q, p)}(E) = P_{\exp(ir\hat{q}_\theta)}(E) \quad (3.10)$$

$$= P_{\hat{q}_\theta}(\{x \in \mathbb{R} \mid e^{irx} \in E\}) \quad (3.11)$$

$$= \hat{R}(\theta)P_{\hat{q}}(\{x \in \mathbb{R} \mid e^{irx} \in E\})\hat{R}(-\theta) . \quad (3.12)$$

In conclusion, “measuring a displacement operator” can be implemented with quadrature measurements. Now measuring any multimode displacement amounts to being able to measure arbitrary multimode quadratures; that is, any linear combination of quadratures, e.g.  $\hat{q}_1 +$

$2\hat{p}_{2\theta} + 5\hat{q}_{M\alpha}$  for arbitrary angles  $\theta, \alpha$ . First, we apply phase-shift operators  $\hat{R}(\theta)$  for each individual mode to obtain the right rotated quadratures. Then we apply CZ gates of the form  $e^{ig\hat{q}_i\hat{q}_j}$  for  $g \in \mathbb{R}$  to pairs of modes  $i$  and  $j$  to sum them. This permits construction of the desired linear combinations in a quadrature of a mode. It remains to measure it. This can be implemented with standard homodyne detection. It consists of a Gaussian measurement of a quadrature of the field, by mixing the state with a strong coherent state. Then the intensities of both output arms are measured with photodiode detectors. Their difference yields a value proportional to a quadrature of the input mode, which can be rotated depending on the phase of the local oscillator. The POVM elements for homodyne detection are given by  $|x\rangle_\phi \langle x|$  for all  $x \in \mathbb{R}$  where  $|x\rangle_\phi$  is the eigenstate of the rotated quadrature operator  $\hat{q}_\phi$  with eigenvalue  $x$ . This is represented in Figure 3.1. All of these steps can be implemented experimentally [FOP05, SHD<sup>+</sup>13].

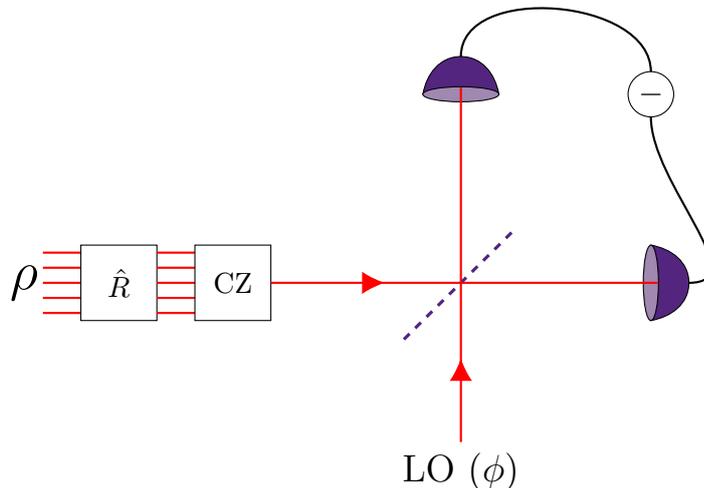


Figure 3.1: Experimental protocol corresponding to the measurement scenario of this chapter. It permits measurement of any linear combination of quadratures. After phase-shift operations on individual modes and CZ gates on pairs of modes, homodyne detection one mode of the state is implemented. The dashed line represents a balanced beamsplitter. The local oscillator (LO) is a strong coherent state. At the hand of each arm are photodiode detectors. The difference in the intensity yields a value proportional to a quadrature of the input mode, which can be rotated depending on the phase of the local oscillator.

### Maximal contexts

This measurement scenario is to be interpreted as follows. The measurement corresponding to the label  $\mathbf{r} \in \mathcal{X}$  is described by the spectral measure  $P_{\hat{r}}$  for the quadrature corresponding to  $\mathbf{r}$ . A pair of spectral measures of self-adjoint operators is compatible, in the sense that they admit a joint spectral measure if and only if they commute, which in turn is true if and only if the

operators themselves commute [Hal13]. In the case of our measurement scenario, two spectral measures associated to  $\mathbf{x}, \mathbf{y} \in \mathcal{X}$  commute if and only if  $\Omega(\mathbf{x}, \mathbf{y}) = 0$ . Thus, measurement labels are compatible only when they both belong to some isotropic subspace of  $\mathcal{X}$ . The maximal isotropic subspaces are the Lagrangian subspaces, so each context  $L \in \mathcal{M}$  corresponds to a Lagrangian subspace.  $\mathcal{M}$  must be the Lagrangian Grassmanian of  $\mathcal{X}$ .

## 3.2 Admissible empirical models

### Empirical models

We are interested in experiments arising from quadrature measurements of a quantum system. For each context  $L \in \mathcal{M}$ , the set  $\mathcal{O}_L = \prod_{\mathbf{x} \in L} \mathbb{R}$  can be seen as the set of functions from  $L$  to  $\mathbb{R}$  with the corresponding  $\sigma$ -algebra constructed with the product topology. For  $\mathbf{x} \in \mathcal{X}$ , the categorical projections are:

$$\begin{aligned} \pi_{\mathbf{x}} : \mathcal{O}_L &\longrightarrow \mathbb{R} \\ f &\longmapsto f(\mathbf{x}) . \end{aligned} \tag{3.13}$$

We restrict our attention to empirical models  $e = (e_L)_{L \in \mathcal{M}}$  which satisfy the Born rule; i.e. there exists some quantum state  $\rho \in \mathcal{D}(\mathcal{H})$  such that for all contexts  $L \in \mathcal{M}$  and measurable sets  $U \in \mathcal{F}_L$ :

$$e_L(U) = \text{Tr} \left( \rho \prod_{\mathbf{x} \in L} P_{\hat{\mathbf{x}}} \circ \pi_{\mathbf{x}}(U) \right) . \tag{3.14}$$

We will therefore use the notation  $e^\rho = (e_L^\rho)_{L \in \mathcal{M}}$  to make explicit the dependence with  $\rho$ . Because of the compatibility condition, we may unambiguously write, for each  $\mathbf{x} \in \mathcal{X}$  and for each  $U \in \mathcal{F}_{\mathbf{x}}$  (we write  $\mathcal{F}_{\mathbf{x}}$  for simplicity though we mean  $\mathcal{F}_{\{\mathbf{x}\}}$ ):

$$e_{\mathbf{x}}^\rho(U) = \text{Tr} (\rho P_{\hat{\mathbf{x}}} \circ \pi_{\mathbf{x}}(U)) . \tag{3.15}$$

This comes from the marginalisation  $e_L^\rho|_{\mathbf{x}}$  for each  $L \in \mathcal{M}$  such that  $\mathbf{x} \in L$ .

At this stage there is a mismatch. The Wigner function is a quasiprobability distribution over  $V^M = \mathcal{X}$ , while the extendability property of a noncontextual empirical model in the measurement scenario presented above provides a global probability measure on  $\mathcal{O}_{\mathcal{X}}$ , which can be seen as the set of functions  $\mathcal{X} \rightarrow \mathbb{R}$ . In general, the latter is much larger than the former. To solve this issue, we show that we can restrict to linear value assignments so that  $\mathcal{O}_{\mathcal{X}}$  can be taken as  $\mathcal{X}^*$ , the linear dual of  $\mathcal{X}$ . Because  $\mathcal{X}^* \cong \mathcal{X}$  then there is no more mismatch between the Wigner function and the global probability measure that exists for a noncontextual empirical behaviour.

We first show that we can restrict to linear value assignments on Lagrangian subspaces in Lemma 3.1. We do so by showing that the empirical model assigns a nonzero mass only to the linear functions  $L \rightarrow \mathbb{R}$ . We then lift this property to global value assignments in Proposition 3.1 in the same spirit as [DOBV<sup>+</sup>17].

**Lemma 3.1.** *Let  $L \subseteq \mathcal{X}$  be a Lagrangian subspace. Let  $U \in \mathcal{F}_L$  be a Lebesgue measurable set of functions  $L \rightarrow \mathbb{R}$  such that  $\pi_{\mathbf{x}}(U)$  is distinct from  $\mathbb{R}$  for a finite number of  $\mathbf{x} \in L$ . Then there exists a subset  $U_{\text{lin}}$  of linear functions  $L \rightarrow \mathbb{R}$  such that*

$$e_L^\rho(U_{\text{lin}}) = e_L^\rho(U) . \quad (3.16)$$

*Proof.* First let  $(\mathbf{e}_k)_{k=1,\dots,M}$  be a basis of  $L \cong \mathbb{R}^M$ . Let  $P$  be the joint spectral measure of  $\{P_{\hat{\mathbf{e}}_1}, \dots, P_{\hat{\mathbf{e}}_M}\}$ . For any  $\mathbf{y} \in L$ , define the function

$$\begin{aligned} f_{\mathbf{y}} : L &\longrightarrow \mathbb{R} \\ \mathbf{x} &\longmapsto \mathbf{x} \cdot \mathbf{y} , \end{aligned} \quad (3.17)$$

where  $\cdot$  is the usual Euclidean scalar product on  $L \cong \mathbb{R}^M$ . For any  $\mathbf{x} \in L$ ,  $P_{\hat{\mathbf{x}}}$  is the push-forward of  $P$  by the measurable function  $f_{\mathbf{x}}$  by definition of the functional calculus on  $M$  variables. Recall that for  $\mathbf{x} \in L$ ,  $\pi_{\mathbf{x}}(U) = \{f(\mathbf{x}) \mid f \in U\} \subseteq \mathbb{R}$ .

Then,

$$\text{Tr} \left( \rho \prod_{\mathbf{x} \in L} P_{\hat{\mathbf{x}}} \circ \pi_{\mathbf{x}}(U) \right) = \text{Tr} \left( \rho \prod_{\mathbf{x} \in L} P \left( f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U)) \right) \right) \quad (3.18)$$

$$= \text{Tr} \left( \rho P \left( \bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U)) \right) \right) \quad (3.19)$$

with

$$\bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U)) = \{\mathbf{y} \in L \mid \forall \mathbf{x} \in L, \mathbf{x} \cdot \mathbf{y} \in \pi_{\mathbf{x}}(U)\} . \quad (3.20)$$

Now define

$$U_{\text{lin}} := \left\{ \begin{array}{l} L \longrightarrow \mathbb{R} \\ \mathbf{x} \longmapsto \mathbf{x} \cdot \mathbf{y} \end{array} \mid \mathbf{y} \in \bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U)) \right\} . \quad (3.21)$$

By construction,

$$\bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U_{\text{lin}})) \quad (3.22)$$

$$= \bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\{\mathbf{x} \cdot \mathbf{y} \mid \mathbf{y} \in L \text{ s.t. } \forall \mathbf{z} \in L, \mathbf{y} \cdot \mathbf{z} \in \pi_{\mathbf{z}}(U)\}) \quad (3.23)$$

$$= \bigcap_{\mathbf{x} \in L} \{\boldsymbol{\alpha} \in L \mid \mathbf{x} \cdot \boldsymbol{\alpha} = \mathbf{x} \cdot \mathbf{y} \text{ with } \mathbf{y} \in L \text{ s.t. } \forall \mathbf{z} \in L, \mathbf{y} \cdot \mathbf{z} \in \pi_{\mathbf{z}}(U)\} \quad (3.24)$$

$$= \{\boldsymbol{\alpha} \in L \mid \forall \mathbf{x} \in L, \mathbf{x} \cdot \boldsymbol{\alpha} = \mathbf{x} \cdot \mathbf{y} \text{ with } \mathbf{y} \in L \text{ s.t. } \forall \mathbf{z} \in L, \mathbf{y} \cdot \mathbf{z} \in \pi_{\mathbf{z}}(U)\} \quad (3.25)$$

$$= \{\boldsymbol{\alpha} \in L \mid \forall \mathbf{z} \in L, \boldsymbol{\alpha} \cdot \mathbf{z} \in \pi_{\mathbf{z}}(U)\} \quad (3.26)$$

$$= \bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U)) , \quad (3.27)$$

where Eq. (3.26) follows from that fact that  $\forall \mathbf{x} \in L$ ,  $\mathbf{x} \cdot \boldsymbol{\alpha} = \mathbf{x} \cdot \mathbf{y}$  implies  $\boldsymbol{\alpha} = \mathbf{y}$ . Also for all  $\mathbf{x} \in L$ ,  $\pi_{\mathbf{x}}(U_{\text{lin}}) \subseteq \pi_{\mathbf{x}}(U)$  so that we are indeed reproducing all value assignments from linear functions of  $U$ .

Then, as claimed,

$$e_L^\rho(U_{\text{lin}}) = \text{Tr} \left( \rho \prod_{\mathbf{x} \in L} P_{\hat{\mathbf{x}}} \circ \pi_{\mathbf{x}}(U_{\text{lin}}) \right) \quad (3.28)$$

$$= \text{Tr} \left( \rho \prod_{\mathbf{x} \in L} P \left( f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U_{\text{lin}})) \right) \right) \quad (3.29)$$

$$= \text{Tr} \left( \rho P \left( \bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U_{\text{lin}})) \right) \right) \quad (3.30)$$

$$= \text{Tr} \left( \rho P \left( \bigcap_{\mathbf{x} \in L} f_{\mathbf{x}}^{-1}(\pi_{\mathbf{x}}(U)) \right) \right) \quad (3.31)$$

$$= \text{Tr} \left( \rho \prod_{\mathbf{x} \in L} P_{\hat{\mathbf{x}}}(\pi_{\mathbf{x}}(U)) \right) \quad (3.32)$$

$$= e_L^\rho(U) . \quad (3.33)$$

□

Now we prove that the set of global value assignments can be identified with  $\mathcal{X}^*$ , the linear dual space of  $\mathcal{X}$ .

**Proposition 3.1.** *If  $M \geq 2$  (i.e. for more than 2 modes), global value assignments are linear functions  $\mathcal{X} \rightarrow \mathbb{R}$ , and the set of global value assignments forms a  $\mathbb{R}$ -linear space of dimension  $2M$ , namely  $\mathcal{E}(\mathcal{X}) = \mathcal{X}^*$ .*

*Proof.* The sheaf-theoretic framework for contextuality describes value assignments as a sheaf  $\mathcal{E} : \mathcal{P}(\mathcal{X})^{\text{op}} \rightarrow \text{Meas}$ , where  $\mathcal{E}(U)$  is the set of value assignments for the measurement labels in  $U$ , which can be viewed as a set of functions  $U \rightarrow \mathbb{R}$ . For any Lagrangian  $L \in \mathcal{M}$ , there is a restriction map  $\rho_L^{\mathcal{X}} = \mathcal{E}(\mathcal{X}) \rightarrow \mathcal{E}(L) : f \mapsto f|_L$  that simply restricts the domain of any function from  $\mathcal{X}$  to  $L$ . Then  $\mathcal{E}(L)$  must coincide with the set of possible value assignments  $\mathcal{O}_L$ .

By Lemma 3.1,  $\mathcal{E}(L)$  consists in linear functions  $L \rightarrow \mathbb{R}$  so that  $\mathcal{E}(\mathcal{X})$  contains only functions  $\mathcal{X} \rightarrow \mathbb{R}$  whose restriction to any Lagrangian subspace is  $\mathbb{R}$ -linear. Then, following [DOBV<sup>+</sup>17, Lemma 1] (the lemma is proven for the discrete phase-space  $\mathbb{Z}_d^M \times \mathbb{Z}_d^M$  but its proof extends directly to  $\mathbb{R}^M \times \mathbb{R}^M$ ), we conclude that if  $M \geq 2$ ,  $\mathcal{E}(\mathcal{X})$  contains only  $\mathbb{R}$ -linear functions  $\mathcal{X} \rightarrow \mathbb{R}$ , i.e.  $\mathcal{E}(\mathcal{X}) = \mathcal{X}^*$ . □

Therefore, without loss of generality, for any  $U \subset \mathcal{X}$ , we can restrict  $\mathcal{O}_U$  to be the set of linear functions from  $U \rightarrow \mathbb{R}$ . Thus, an empirical model will be a collection of probability measures on  $L^*$  for each  $L \in \mathcal{M}$ . For a noncontextual empirical model  $e^\rho$ , the extendability

property yields a global probability measure on  $\mathcal{X}^*$  that we will identify with the Wigner function of  $\rho$  in the following section.

### 3.3 Equivalence between Wigner negativity and contextuality

We are now ready to tackle the main proof that Wigner negativity is equivalent to contextuality in our measurement scenario. We prove it by essentially identifying the Wigner function with the probability density of a carefully constructed hidden-variable model. We first set up the hidden-variable model via Proposition 3.2 and we prove the equivalence in Theorem 3.1. Crucially, for the identification with the Wigner function, we need to ensure that the hidden-variable model is realisable by a probability measure over hidden variables that has a density. We further require that this density is a  $L^1$  function.

**Proposition 3.2.** *If an empirical model  $e^\rho$  for the continuous-variable measurement scenario in Definition 3.1 is noncontextual, then  $e^\rho$  admits a realisation by a deterministic hidden-variable model  $\langle \mathcal{X}, \mu_r, (k_L)_{L \in \mathcal{M}} \rangle$  such that  $\mu_r$  has density  $w_\mu \in L^1(\mathcal{X})$  with respect to the Lebesgue measure.*

*Proof.* By our extension of the FAB theorem (see Theorem 2.1) and Lemma 3.1,  $e^\rho$  is realised by a canonical hidden-variable model (HVM)  $(\mathbf{\Lambda}, \mu, k)$  (see Definition 2.4), for which

- $\mathbf{\Lambda} = \mathcal{O}_{\mathcal{X}} = \mathcal{X}^*$  i.e. hidden variables are linear value assignments;
- each probability kernel  $k_L : \mathcal{X}^* \rightarrow L^*$  is deterministic and factorisable;
- $\mu$  is a probability measure on  $\mathcal{X}^*$ .

In the same spirit as the Riesz representation theorem [Rie09], we pick the following isomorphism constructed with the symplectic form (see Eq. (1.28)) to identify elements from  $\mathcal{X}$  to elements from  $\mathcal{X}^*$ :

$$\begin{aligned} \alpha : \mathcal{X} &\longrightarrow \mathcal{X}^* \\ \mathbf{x} &\longmapsto \Omega(\mathbf{x}, -) . \end{aligned} \tag{3.34}$$

This will be essential to take the hidden-variable space to be  $\mathcal{X}$  rather than  $\mathcal{X}^*$ .

For all  $L \in \mathcal{M}$ , let

$$\begin{aligned} \tilde{k}_L : \mathcal{X} \times \mathcal{F}_L &\longrightarrow \mathbb{R} \\ (\mathbf{x}, E) &\longmapsto k_L(\alpha(\mathbf{x}), E) = \delta_{\alpha(\mathbf{x})|_L}(E) . \end{aligned} \tag{3.35}$$

Note that for both  $(\tilde{k}_L)_{L \in \mathcal{M}}$  and  $(k_L)_{L \in \mathcal{M}}$  we can unambiguously write  $\tilde{k}_{\mathbf{x}}$  and  $k_{\mathbf{x}}$  for a measurement label  $\mathbf{x} \in \mathcal{X}$  because of the compatibility condition.

Fix  $\mathbf{x} \in \mathcal{X}$ . For  $E \in \mathcal{B}(\mathbb{R})$ , let

$$p_{\mathbf{x}}^\rho(E) := \text{Tr}(\rho P_{\tilde{\mathbf{x}}}(E)) . \tag{3.36}$$

Fix  $U \in \mathcal{F}_x$  (a measurable subset of linear functions from  $\{\mathbf{x}\}$  to  $\mathbb{R}$ ). We first evaluate  $p_x^\rho \circ \pi_x(E) = e_x^\rho(E)$  (see Eq. (3.15)) with the HVM above (see Eq. (2.3)).

$$p_x^\rho(\pi_x(U)) = e_x^\rho(U) \quad (3.37)$$

$$= \int_{\Lambda} k_x(-, U) \, d\mu \quad (3.38)$$

$$= \int_{f \in \mathcal{X}^*} k_x(f, U) \, d\mu(f) \quad (3.39)$$

$$= \int_{f \in \mathcal{X}^*} \delta_{f|_{\{\mathbf{x}\}}}(U) \, d\mu(f) \quad (3.40)$$

$$= \int_{f \in \mathcal{X}^*} \delta_{f(\mathbf{x})}(\pi_x(U)) \, d\mu(f) \quad (3.41)$$

$$= \int_{f \in \mathcal{X}^*} \delta_{\alpha^{-1}(f) \cdot \mathbf{x}}(\pi_x(U)) \, d\mu(f) \quad (3.42)$$

$$= \int_{\mathbf{y} \in \Omega(\mathbf{x}, -)^{-1}(\pi_x(U))} d\mu \circ \alpha(\mathbf{y}) \quad (3.43)$$

$$= \int_{\mathbf{y} \in (\mathbf{x} \cdot -)^{-1}(\pi_x(U))} d\mu \circ \alpha \circ J(\mathbf{y}) . \quad (3.44)$$

Thus  $p_x^\rho$  is the push-forward of the measure  $\mu \circ \alpha \circ J$  on  $\mathcal{X}$  by the linear functional  $(\mathbf{x} \cdot -)$  where  $J$  is defined in Eq. (3.1). By the Lebesgue decomposition theorem [Bil08], there is a decomposition  $\mu \circ \alpha = \mu_r + \mu_s$  where  $\mu_r$  is absolutely continuous with respect to the Lebesgue measure  $d\mathbf{x}$  on  $\mathcal{X}$  and  $\mu_s$  is singular with respect to  $d\mathbf{x}$ . It follows that, that for any  $\mathbf{x} \in \mathcal{X}$ , since  $A = (\mathbf{x} \cdot -)^{-1}(E)$  has non-zero  $d\mathbf{x}$ -measure for any Borel-measurable  $E \subseteq \mathbb{R}$  of non-zero Lebesgue measure,

$$\mu_r \circ J(A) = \mu \circ \alpha \circ J(A) - \mu_s \circ J(A) = \mu \circ \alpha \circ J(A) = p_x^\rho(E). \quad (3.45)$$

Then  $(\mathcal{X}, \mu_r, (\tilde{k}_L)_L)$  is a deterministic and factorisable hidden-variable model for the empirical model  $e^\rho$ . By the Radon-Nikodym theorem [Nik30],  $\mu_r$  has a density  $w_\mu$  with respect to the Lebesgue measure  $d\mathbf{x}$  on  $\mathcal{X}$ . Since  $\mu_r$  is a probability measure,  $w_\mu \in L^1(\mathcal{X})$ .  $\square$

The main result follows from identifying  $w_\mu$  (Proposition 3.2) and the Wigner function  $W_\rho$  as  $L^1(\mathcal{X})$  functions:

**Theorem 3.1.** *Assume  $\rho$  is a density operator such that its Wigner function  $W_\rho \in L^1(\mathcal{X})$  with respect to the Lebesgue measure. Let  $e^\rho$  be an empirical model on the measurement scenario in Definition 3.1 for  $\rho$  according to the Born rule i.e. for any  $L \in \mathcal{M}$ , for  $U \in \mathcal{F}_L$ ,  $e_L(U) = \text{Tr}(\rho \prod_{\mathbf{x} \in L} P_{\tilde{\mathbf{x}}} \circ \pi_{\mathbf{x}}(U))$ . Then  $e^\rho$  is noncontextual if and only if the Wigner function  $W_\rho$  of  $\rho$  is nonnegative, and in that case  $W_\rho$  describes a hidden variable model for  $e^\rho$ .*

*Proof.* The result holds by identifying the characteristic function of  $\rho$ , denoted  $\Phi_\rho$  (see Eq. (1.27)) with the Wigner function and with the density  $w_\mu$  from Proposition 3.2. For  $\mathbf{x} \in \mathcal{X}$ , we have

that  $\Phi_\rho(\mathbf{x}) = \text{FT}^{-1}[W_\rho](J\mathbf{x})$  by taking the inverse Fourier transform of Eq. (1.29) with the change of variables  $\mathbf{x} \rightarrow J\mathbf{x}$ .

On the other hand, fix a noncontextual empirical model  $e^\rho$  satisfying the Born rule associated to  $\rho$  and the measurement scenario in Definition 3.1. By Proposition 3.2 we have:

$$\Phi(\mathbf{x}) = \text{Tr}(\hat{D}(\mathbf{x})\rho) \quad (3.46)$$

$$= \text{Tr}\left(\rho \int_{\lambda \in \mathbb{R}} e^{i\lambda} dP_{\hat{\mathbf{x}}}(\lambda)\right) \quad (3.47)$$

$$= \int_{\mathbb{R}} e^{i\lambda} dp_{\hat{\mathbf{x}}}^\rho(\lambda) \quad (3.48)$$

$$= \int_{\mathcal{X}} e^{i\mathbf{x} \cdot \mathbf{y}} d\mu_r \circ J(\mathbf{y}) \quad (3.49)$$

$$= \int_{\mathcal{X}} e^{iJ\mathbf{x} \cdot \mathbf{y}} w_\mu(\mathbf{y}) d\mathbf{y} \quad (3.50)$$

$$= \text{FT}^{-1}[w_\mu](J\mathbf{x}). \quad (3.51)$$

where the second line comes from the spectral theorem; the third line via Eq. (3.36) and the fact that the integral and the trace may be inverted by the definition of the integral with respect to the spectral measure [Hal13]; the fourth line via the push-forward operation in Eq. (3.45); and the two last lines comes from Proposition 3.2 and the definition of the inverse Fourier transform.<sup>2</sup> As a result, we have  $\text{FT}^{-1}[w_\mu](J\mathbf{x}) = \text{FT}^{-1}[W_\rho](J\mathbf{x})$  and since  $w_\mu, W_\rho \in L^1(X)$ ,  $w_\mu = W_\rho$  almost everywhere [Fol92, Corollary 7.1]. We have that  $w_\mu$  is a density function of a probability measure, so it follows that both functions must be almost everywhere nonnegative. Because the Wigner function is a continuous function from  $\mathcal{X}$  to  $\mathbb{R}$  [CG69],  $W_\rho$  must be nonnegative.

Conversely, a nonnegative Wigner function provides a canonical noncontextual hidden variable model for  $e^\rho$  [SKK<sup>+</sup>09].  $\square$

### 3.4 Discussion and open problems

We have seen that Wigner negativity is equivalent to contextuality with respect to continuous-variable Pauli measurements. This raises important questions. First the present argument requires considering a measurement scenario that comprises an uncountable family of measurement labels (the entire phase-space  $V^M$ ). From an experimental perspective, it is crucial to wonder what happens if we restrict to a finite family of measurement labels and see whether we can derive a robust version of this theorem.

Another question concerns a quantifiable relationship between contextuality and Wigner negativity. In the previous chapter we saw that can formalise the contextual fraction [ABM17] in continuous variables. In the next chapter, we will derive witnesses for Wigner negativity whose violation lower bounds the distance to the set of states with a positive Wigner function. It

<sup>2</sup>The missing factor  $(\frac{1}{2\pi})^M$  can always be taken into account in the measure.

would be highly desirable to establish a precise and quantified link between those two measures of nonclassicality.

Homodyne detection is a standard detection method in continuous variables [YUA<sup>+</sup>13] and it is the basis of several computational models in continuous-variable quantum information [DMK<sup>+</sup>17, CDM<sup>+</sup>17, CC17, GKP01]. In [BPA<sup>+</sup>19], a continuous-variable model for fault-tolerant, universal quantum computation using only homodyne detection is developed. It combines Gottesman–Kitaev–Preskill [GKP01] Clifford quantum computation and Gaussian quantum computation. This justifies the measurement scenario under consideration. Homodyne detection is a Gaussian measurement, therefore any quantum advantage is due to Wigner negativity being present before the detection setup. This result shows that, just like in the discrete-variable case [HWVE14], continuous-variable contextuality supplies the necessary ingredients for continuous-variable quantum computing.

## WITNESSING WIGNER NEGATIVITY

**I**N Chapter 2 we formalised a framework for defining properly contextuality in continuous-variable systems. In Chapter 3, we showed that contextuality has a close connection with Wigner negativity, namely that they are equivalent in the setting we presented. We now turn to a more in-depth analysis of Wigner negativity.

As the Wigner function is a real-valued quasi-probability distribution [CG69]—a normalised distribution which can take negative values—it cannot be sampled directly experimentally in general. However, its marginals are proper probability distributions which can thus be sampled using homodyne detection in the optical setting [LR09]. Alternatively, heterodyne detection (also called double homodyne detection) allows for sampling from a smoothed version of the Wigner function [Hus40, Ric98]. In both cases, displacing a state in phase space before the detection is equivalent to measuring the undisplaced state directly with homodyne or heterodyne detection and then applying a classical post-processing procedure—namely, a translation of the classical outcome according to the displacement amplitude [CGKM20, CRW<sup>+</sup>21].

Reconstructing the Wigner function to assess whether it is negative or not is possible via full tomographic reconstruction [DPS03]. However, such reconstructions are very costly in terms of the number of measurements needed, and require performing a tomographically complete set of measurements—thus usually involving multiple measurement settings. In the continuous-variable setting the task is even more daunting, since the Hilbert space of quantum states is infinite-dimensional [LR09, CDG<sup>+</sup>20].

Instead, another strategy is to introduce *witnesses* for specific properties of quantum states [Ter01, LKCH00, MKN<sup>+</sup>11, KV12, GEH<sup>+</sup>21, CRW<sup>+</sup>21] that are more accessible experimentally. These witnesses should be observables that possess a *threshold expectation value* indicating whether the measured state exhibits the desired property or not. Intuitively, a witness

for a given property can be thought of as a separating hyperplane in the set of quantum states, such that any state on one side of this hyperplane has this property (see Fig. 4.1). In particular, some states with the sought property may remain unnoticed by the witness. In that regard, it is desirable to have a *complete* set of witnesses such that for each state exhibiting the desired property, there exists at least one witness in the set that captures it.

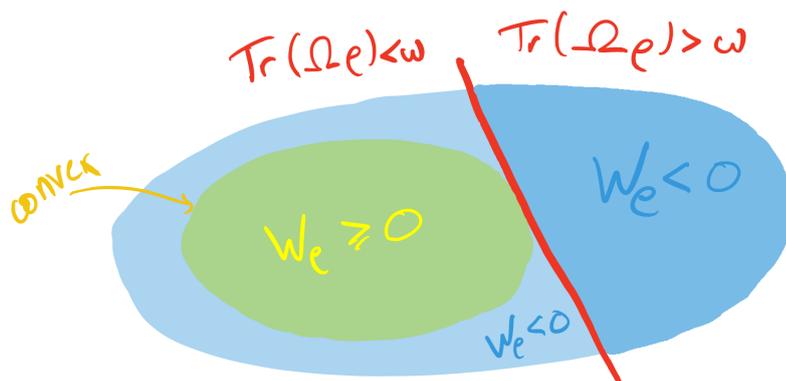


Figure 4.1: Pictorial representation of a witness  $\hat{\Omega}$  with threshold value  $\omega$  for a given property. In green: states without the property. In blue: states with the property. In red: witness threshold value. In light blue: states with the property undetected by the witness.

A natural choice for a witness of Wigner negativity is the fidelity with a pure state having a Wigner function with negative values, since it is a quantity that can be accessed experimentally by direct fidelity estimation [DPS03, CDG<sup>+</sup>20]. Building on this intuition, and given that all Fock states—with the exception of the (Gaussian) vacuum state  $|0\rangle$ —have a negative Wigner function, we introduce a broad family of Wigner negativity witnesses for single-mode and multimode continuous-variable quantum states based on fidelities with Fock states. Furthermore this family is *complete*, that is, for each state exhibiting Wigner negativity, there exists at least one witness that will detect it. The expectation values of the witnesses are linear functions of the state which may be *efficiently estimated experimentally* using standard homodyne or heterodyne detection, thus providing a reliable method for detecting Wigner negativity with certifiable bounds. Additionally, we show that the amount by which the measured expectation value exceeds the threshold value of the witness provides an operational measure of Wigner negativity as it directly lower bounds the distance between the measured state and the set of states with positive Wigner function.

We cast the computation of the threshold values of the witnesses as infinite-dimensional linear programs. To perform an actual numerical implementation, we can either relax or restrict these programs. *Upper and lower bounds for the threshold values* of our witnesses are then given by two converging hierarchies of finite-dimensional semidefinite programs. While we first study

the case of single-mode quantum states and witnesses for simplicity, we then show that our results extend to the multimode setting—when there are more than one quantum system.

This chapter may be of interest for physicists interested in characterising Wigner negativity of quantum states—either theoretically or experimentally—and mathematicians interested in infinite-dimensional convex optimisation theory. It is structured as follows: we give a detailed exposition of our witnesses in Section 4.1. Section 4.2 describes the experimental procedure for witnessing Wigner negativity of a quantum state using these witnesses, together with use-case examples. The following Section 4.3—which deals with infinite-dimensional optimisation techniques of independent interest—is devoted to estimating the threshold values of the witnesses: after some technical background in Section 4.3.2, Section 4.3.3 reformulates the problem of finding the threshold value of a witness as an infinite-dimensional linear optimisation, while Section 4.3.4 derives two hierarchies of semidefinite relaxations and restrictions for this linear program, yielding numerical upper and lower bounds for the threshold value. Section 4.3.5 establishes the proof of convergence of these hierarchies of upper and lower bounds to the threshold values in their respective optimisation spaces. We introduce the generalisation to the multimode case in Section 4.4 and conclude with a few open questions in Section 4.5. This chapter is based on [CEG21].

## 4.1 Wigner negativity witnesses

We introduce the following Wigner negativity witnesses:

$$\hat{\Omega}_{\mathbf{a},\alpha} := \sum_{k=1}^n a_k \hat{D}(\alpha) |k\rangle\langle k| \hat{D}^\dagger(\alpha), \quad (4.1)$$

for  $n \in \mathbb{N}^*$ ,  $\mathbf{a} = (a_1, \dots, a_n) \in [0, 1]^n$ , with  $\max_k a_k = 1$ , and  $\alpha \in \mathbb{C}$ . These operators are *finite weighted sums of displaced Fock states projectors*. They can be thought of as Positive Operator-Valued Measure (POVM) elements, and their expectation value for a quantum state  $\rho$  is given by

$$\mathrm{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho) = \sum_{k=1}^n a_k F(\hat{D}^\dagger(\alpha) \rho \hat{D}(\alpha), |k\rangle), \quad (4.2)$$

where  $F$  is the fidelity (see Subsection 1.1.1). This quantity can be (efficiently) directly estimated from homodyne or heterodyne detection of multiple copies of the state  $\rho$  by translating the samples obtained by the amplitude  $\alpha$  in the classical postprocessing and performing fidelity estimation with the Fock states  $|1\rangle, \dots, |n\rangle$  [DPS03, CDG<sup>+</sup>20].

For  $n \in \mathbb{N}^*$ , each choice of  $(\mathbf{a}, \alpha) \in [0, 1]^n \times \mathbb{C}$  yields a different Wigner negativity witness. In particular, when  $\alpha = 0$  and only one entry of the vector  $\mathbf{a}$  is non-zero, the expectation value of the witness is given by the fidelity with a single Fock state.

To each witness  $\hat{\Omega}_{\mathbf{a},\alpha}$  is associated a threshold value defined as:

$$\omega_{\mathbf{a}} := \sup_{\substack{\rho \in \mathcal{D}(\mathcal{H}) \\ W_{\rho} \geq 0}} \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho). \quad (4.3)$$

Since negativity of the Wigner function is invariant under displacements, the threshold values do not depend on the value of the displacement amplitude  $\alpha$  and we thus write  $\omega_{\mathbf{a}}$  (rather than  $\omega_{\mathbf{a},\alpha}$ ) for the threshold value associated to the witness  $\hat{\Omega}_{\mathbf{a},\alpha}$ . This is sensible, given that the threshold value asks for nonnegativity anywhere in phase space, so a displacement in phase space should not change its value. Combining (i) that the threshold value associated to a witness does not depend on the displacement parameter  $\alpha$  and (ii) that we can always take into account displacement via classical post-processing if one uses homodyne or heterodyne detection associated to  $\hat{\Omega}_{\mathbf{a},0}$  [CGKM20, CRW<sup>+</sup>21], we can restrict the analysis to witnesses of the form  $\hat{\Omega}_{\mathbf{a},0}$  that will generate the family  $\{\hat{\Omega}_{\mathbf{a},\alpha}\}_{\alpha \in \mathbb{C}}$ . Note however that the choice of displacement amplitude can still play an important role for certifying negativity of certain quantum states. For instance, the most suitable witness for detecting negativity of the displaced Fock state  $\hat{D}(e^{i\frac{\pi}{4}})|1\rangle$  is itself i.e.  $\hat{\Omega}_{(1,0,\dots,0),e^{i\frac{\pi}{4}}}$  though  $\hat{\Omega}_{(1,0,\dots,0),0}$  will be used experimentally and then classical post-processing will be applied on the samples to retrieve the displaced witness.

If the measured expectation value for an experimental state is higher than the threshold value given by Eq. (4.3), this implies by definition that its Wigner function takes negative values. Moreover, the following result shows that the amount by which the expectation value exceeds the threshold value directly provides an operational quantification of Wigner negativity for that state:

**Lemma 4.1.** *Let  $\rho \in \mathcal{D}(\mathcal{H})$  Wigner negative, and fix a witness  $\hat{\Omega}_{\mathbf{a},\alpha}$  defined in Eq. (4.1), for  $n \in \mathbb{N}^*$ ,  $\mathbf{a} = (a_1, \dots, a_n) \in [0, 1]^n$ , and  $\alpha \in \mathbb{C}$ , with threshold value  $\omega_{\mathbf{a}}$ . Let us further assume that it violates the threshold value of the witness i.e.  $\text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho) > \omega_{\mathbf{a}}$  and denote the amount of violation as*

$$\delta_{\mathbf{a},\alpha}(\rho) := \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho) - \omega_{\mathbf{a}}. \quad (4.4)$$

Then,

$$\eta_{\rho} \geq \delta_{\mathbf{a},\alpha}(\rho), \quad (4.5)$$

where  $\eta_{\rho}$  is the distance between  $\rho$  and the set of states having a positive Wigner function, defined in Eq. (1.32).

*Proof.* We use the notations of the lemma. Let us consider the binary POVM  $\{\hat{\Omega}_{\mathbf{a},\alpha}, \mathbb{1} - \hat{\Omega}_{\mathbf{a},\alpha}\}$ . For all  $\sigma \in \mathcal{D}(\mathcal{H})$ , we write  $P_{\mathbf{a},\alpha}^{\sigma}$  the associated probability distribution:  $P_{\mathbf{a},\alpha}^{\sigma}(0) = 1 - P_{\mathbf{a},\alpha}^{\sigma}(1) = \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \sigma)$ .

Let  $\sigma$  be a state with a positive Wigner function, so that  $\text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha}\sigma) \leq \omega_{\mathbf{a}}$ , by definition of the threshold value. We have:

$$\begin{aligned}
 \delta_{\mathbf{a},\alpha}(\rho) &= \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha}\rho) - \omega_{\mathbf{a}} \\
 &\leq \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha}\rho) - \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha}\sigma) \\
 &= |P_{\mathbf{a},\alpha}^{\rho}(0) - P_{\mathbf{a},\alpha}^{\sigma}(0)| \\
 &= \|P_{\mathbf{a},\alpha}^{\rho} - P_{\mathbf{a},\alpha}^{\sigma}\| \\
 &\leq D(\rho, \sigma),
 \end{aligned} \tag{4.6}$$

where we used  $\delta_{\mathbf{a},\alpha}(\rho) \geq 0$  in the second line,  $\|P - Q\| = \frac{1}{2} \sum_x |P(x) - Q(x)|$  denotes the total variation distance, and we used the operational property of the trace distance in the last line [NC11]. We finally take the infimum over  $\sigma$  and with the definition of  $\eta_{\rho}$  (see Eq. (1.32)) that concludes the proof.  $\square$

This result directly extends to the case where only an upper bound of the threshold value is known: the amount by which the expectation value exceeds this upper bound is also a lower bound of the distance to the set of states having a positive Wigner function.

Importantly, the family of Wigner negativity witnesses  $\{\hat{\Omega}_{\mathbf{a},\alpha}\}$  is *complete*, i.e., for any quantum state with negative Wigner function there exists a choice of witness  $(\mathbf{a}, \alpha)$  such that the expectation value of  $\hat{\Omega}_{\mathbf{a},\alpha}$  for this state is higher than the threshold value. Indeed, by taking  $\mathbf{a} = (1, 0, 1, 0, 1, \dots)$ , this family includes as a subclass the complete family of witnesses from [CRW<sup>+</sup>21]. Indeed from Eq. (1.25) (definition of the parity operator), Eq. (1.26) (definition of the Wigner function) and the completeness relation  $\sum_{n \in \mathbb{N}} |n\rangle\langle n| = \mathbb{1}$ , the Wigner function of any density operator  $\rho \in \mathcal{D}(\mathcal{H})$  reads:

$$W_{\rho}(\alpha) = \frac{2}{\pi} \left( 1 - 2\text{Tr} \left( \hat{\Omega}_{(1,0,1,0,\dots),\alpha} \rho \right) \right).$$

Thus for any state with a negative Wigner function, there exists a choice of  $\alpha \in \mathbb{C}$  such that the witness  $\hat{\Omega}_{(1,0,1,0,\dots),\alpha}$  with threshold value  $\frac{1}{2}$  can detect its negativity.

The threshold value in Eq. (4.3) is given by an optimisation problem over quantum states having a positive Wigner function. This is a convex subset of an infinite-dimensional space that does not possess a well-characterised structure. While solving this optimisation problem thus seems unfeasible in general, it turns out that we can obtain increasingly good numerical upper and lower bounds for the threshold value using semidefinite programming. This is developed in Section 4.3.

The relevant programs are derived in Section 4.3.4.3 and 4.3.4.2 and their respective convergence is proven in 4.3.5. Since these proofs introduce several intermediate forms of the programs, we explicitly give them below to avoid confusion on which programs to implement numerically. For  $n \in \mathbb{N}^*$ ,  $\mathbf{a} = (a_1, \dots, a_n) \in [0, 1]^n$ , and  $m \geq n$ , the hierarchies of semidefinite programs that respectively provide lower bounds and upper bounds for the threshold value  $\omega_{\mathbf{a}}$

associated to the witnesses  $\{\hat{\Omega}_{\mathbf{a},\alpha}\}_{\alpha \in \mathbb{C}}$  are:

$$(\text{SDP}_{\mathbf{a}}^{m,\leq}) \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{m+1} \text{ and } \mathbf{F} \in \mathbb{R}^{m+1} \\ \text{maximising } \sum_{k=1}^n a_k F_k \\ \text{subject to} \\ \sum_{k=0}^m F_k = 1 \\ \forall k \in \llbracket 0, m \rrbracket, F_k \geq 0 \\ \forall l \in \llbracket 1, m \rrbracket, \sum_{i+j=2l-1} Q_{ij} = 0 \\ \forall l \in \llbracket 0, m \rrbracket, \sum_{i+j=2l} Q_{ij} = \sum_{k=l}^m \frac{(-1)^{k+l}}{l!} \binom{k}{l} F_k \\ Q \succeq 0, \end{array} \right.$$

and

$$(\text{SDP}_{\mathbf{a}}^{m,\geq}) \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{m+1} \text{ and } \mathbf{F} \in \mathbb{R}^{m+1} \\ \text{maximising } \sum_{k=1}^n a_k F_k \\ \text{subject to} \\ \sum_{k=0}^m F_k = 1 \\ \forall k \in \llbracket 0, m \rrbracket, F_k \geq 0 \\ \forall l \in \llbracket 1, m \rrbracket, \forall i+j=2l-1, A_{ij} = 0 \\ \forall l \leq m, \forall i+j=2l, A_{ij} = \sum_{k=0}^l F_k \binom{l}{k} l! \\ A \succeq 0, \end{array} \right.$$

Let  $\omega_{\mathbf{a}}^{m,\geq} = \text{val}(\text{SDP}_{\mathbf{a}}^{m,\geq})$  be the optimal value of  $(\text{SDP}_{\mathbf{a}}^{m,\geq})$ . We show in Section 4.3 that the sequence  $(\omega_{\mathbf{a}}^{m,\geq})_{m \geq n}$  is a decreasing sequence of upper bounds of  $\omega_{\mathbf{a}}$ , which converges to  $\omega_{\mathbf{a}}$ . Similarly, let  $\omega_{\mathbf{a}}^{m,\leq}$  be the optimal value of  $(\text{SDP}_{\mathbf{a}}^{m,\leq})$ . We show that the sequence  $(\omega_{\mathbf{a}}^{m,\leq})_{m \geq n}$  is an increasing sequence of lower bounds of  $\omega_{\mathbf{a}}$ , which converges to  $\omega_{\mathbf{a}}^{\text{S}}$ —a modified threshold value computed with Schwartz functions rather than square-integrable functions; we have  $\omega_{\mathbf{a}}^{\text{S}} \leq \omega_{\mathbf{a}}$ , and the equality between the two values is still open.

In particular, the numerical upper bounds  $\omega_{\mathbf{a}}^{m,\geq}$  can be used instead of the threshold value  $\omega_{\mathbf{a}}$  to witness Wigner negativity, while the numerical lower bounds  $\omega_{\mathbf{a}}^{m,\leq}$  may be used to control how much the upper bounds differ from the threshold value. We give a detailed procedure in the following section, together with use-case examples and details on the numerical implementation.

## 4.2 Procedure for witnessing Wigner negativity

### 4.2.1 Procedure

In this section, we describe an experimental procedure to check whether a continuous-variable quantum state exhibits Wigner negativity using our witnesses.

#### Witnessing Wigner negativity:

1. Choose a suitable fidelity-based witness  $\hat{\Omega}_{\mathbf{a},\alpha}$  defined in Eq. (4.1) by picking  $n \in \mathbb{N}^*$ ,  $\mathbf{a} \in [0, 1]^n$  and  $\alpha \in \mathbb{C}$ .
2. Run the upper bounding semidefinite program ( $\text{SDP}_{\mathbf{a}}^{m,\geq}$ ) for  $m \geq n$  to get a numerical estimate  $\omega_{\mathbf{a}}^{m,\geq}$ .
3. Run the lower bounding semidefinite program ( $\text{SDP}_{\mathbf{a}}^{m,\leq}$ ) for  $m \geq n$  to get a numerical estimate  $\omega_{\mathbf{a}}^{m,\leq}$ .
4. Estimate the expectation value for that witness of the experimental state from samples of homodyne or heterodyne detection by performing fidelity estimation with the corresponding Fock states and translating the samples by  $\alpha$ . This yields an experimental witness value denoted  $\omega_{\text{exp}}$ .
5. Compare the value obtained experimentally with the numerical bounds: if it is greater than the numerical upper bound, i.e.  $\omega_{\text{exp}} \geq \omega_{\mathbf{a}}^{m,\geq}$ , then the state displays Wigner negativity, and its distance to the set of Wigner positive states is lower bounded by  $\omega_{\text{exp}} - \omega_{\mathbf{a}}^{m,\geq}$ . Otherwise, if it is lower than the numerical lower bound, i.e.  $\omega_{\text{exp}} \leq \omega_{\mathbf{a}}^{m,\leq}$ , then the witness cannot detect Wigner negativity for this state.

The main subroutine of this procedure is to estimate fidelities with displaced Fock states using classical samples from homodyne or heterodyne detection<sup>1</sup>, in order to compute the experimental value for a Wigner negativity witness. As already mentioned, displacement can be achieved with classical post-processing by translating the classical samples according to the displacement amplitude, and performing direct fidelity estimation with Fock states [LR09, DPS03, CDG<sup>+</sup>20].

Upper and lower bounds on the threshold value of the witness are then obtained using semidefinite programming, and comparing the experimental witness value to these bounds gives insight about the Wigner negativity of the measured quantum state.

We give a detailed procedure for using our witnesses for detecting Wigner negativity in the framed box. This procedure starts by the choice of a specific witness, and we explain hereafter a heuristic method for picking a good witness.

If the experimental state is anticipated to have negativity at  $\alpha$ , then one may use the witness with parameters  $(n, \mathbf{a}, \alpha)$  with  $\mathbf{a} = (1, 0, 1, 0, \dots)$ , which will detect negativity for  $n$  large enough [CRW<sup>+</sup>21]. However, this may imply having to estimate fidelities with Fock states having a high photon number with homodyne or heterodyne detection, which requires a lot of

<sup>1</sup>Actually, using a fidelity witness rather than a fidelity estimate is sufficient for our purpose.

samples, while simpler witnesses can suffice for the task and be more efficient, as we show in the next section. Moreover, there are cases where the state to be characterised is fully unknown. Instead, a simple heuristic for picking a good witness for Wigner negativity is the following:

- From samples of homodyne or heterodyne detection of multiple copies of an experimental state, estimate the expected values of witnesses in Eq. (4.1) for a small value of  $n$  and a large set of values  $\mathbf{a}$  and  $\alpha$ , using the same samples for all witnesses.
- Based on these values, pick the simplest witness possible—with the smallest value of  $n$ —that is able to witness Wigner negativity with a reasonable violation. This is done by comparing the estimated expected values with the upper bounds on the corresponding threshold values computed numerically. These bounds depend only the choice of the witness parameters  $n, \mathbf{a}$  and can even be precomputed using (SDP $_{\mathbf{a}}^{m, \geq}$ ). To facilitate the use of our methods, we have collected such bounds for  $\mathbf{a} = (0, \dots, 0, 1)$  and  $n \leq 10$  in Table 4.1. We also precomputed these bounds for  $n = 3$  and a large number of values of  $\mathbf{a}$  [CE].
- Then, estimate the expected value for that witness using a new collection of samples—thus obtaining proper error bars and avoiding the accumulation of statistical errors.

In what follows, we give a few theoretical examples for using our witnesses to detect negativity of the Wigner function of single-mode quantum states.

### 4.2.2 Examples

We identify three levels of generality within our family of single-mode witnesses in Eq. (4.1): (i) fidelities with single Fock states, (ii) linear combinations of fidelities with Fock states, and (iii) displaced linear combinations of fidelities with Fock states.

Fidelities with Fock states are the most practical of our witnesses, since they require the estimation of only one diagonal element of the density matrix of the measured state. The corresponding values in Table 4.1 can be used directly by experimentalists: if an estimate of  $\langle n|\rho|n\rangle$  for appropriate  $n$  is above one of these numerical upper bounds then it ensures that  $\rho$  has a Wigner function with negative values. Moreover, by Lemma 4.1, the amount by which the estimate of  $\langle n|\rho|n\rangle$  exceeds the numerical upper bound directly provides a lower bound on the distance between  $\rho$  and the set of states having a positive Wigner function.

For instance, if we focus on  $n = 3$  in Table 4.1, the threshold value  $\omega_3$  satisfies  $0.378 \leq \omega_3 \leq 0.427$ . Having a state  $\rho$  such that  $\langle 3|\rho|3\rangle > 0.427$  guarantees that  $\rho$  has Wigner negativity. If  $\langle 3|\rho|3\rangle < 0.378$  then we conclude that the witness cannot detect negativity for this state. When the experimental state is close to a Fock state (different from the vacuum), a natural choice for the witness thus is the fidelity with the corresponding Fock state. For instance, consider a photon-subtracted squeezed vacuum state [OTBLG06]

$$|\text{p-ssvs}(r)\rangle = \frac{1}{\sinh r} \hat{a} \hat{S}(r) |0\rangle, \quad (4.7)$$

$n$	Lower bound	Upper bound
1	0.5	0.5
2	0.5	0.5
3	0.378	0.427
4	0.375	0.441
5	0.314	0.385
6	0.314	0.378
7	0.277	0.344
8	0.280	0.348
9	0.256	0.341
10	0.262	0.334

Table 4.1: Table of numerical upper and lower bounds for the threshold value  $\omega_n$  of various Wigner negativity witnesses obtained using our hierarchies of semidefinite programs at rank  $m$  up to around 30. The witnesses considered here are Fock states projectors (photon-number states  $|n\rangle$ ) from 1 to 10. Note that the gap between the lower and upper bounds never exceeds 0.1. Additionally, the bounds in the first two lines are analytical (see Section 4.3.3) and for the upper bounds, the corresponding values obtained numerically are 0.528 and 0.551, respectively. See Section 4.2.3 and [CE] for the numerical implementation.

where  $\hat{S}(r) = e^{\frac{r}{2}(\hat{a}^2 - \hat{a}^{\dagger 2})}$  is a squeezing operator with parameter  $r \in \mathbb{R}$ . Its fidelity with the single-photon Fock state  $|1\rangle$  is given by:

$$\frac{1}{(\sinh r)^2} \left| \langle 1 | \hat{a} \hat{S}(r) | 0 \rangle \right|^2 = \frac{1}{(\cosh r)^3}. \quad (4.8)$$

When the squeezing parameter is small, this state is close to a single-photon Fock state. In particular, for  $0 < r < 0.70$ , the fidelity  $F(\text{p-ssvs}(r), 1)$  in Eq. (4.8) is greater than  $\omega_1^{\geq} = \frac{1}{2}$  and our witness can be used to detect Wigner negativity of this state (see Fig. 4.2).

Another example is given by superpositions of coherent states: we consider the cat state [San92]

$$|\text{cat}_2(\alpha)\rangle = \frac{|\alpha\rangle + |-\alpha\rangle}{\sqrt{2(1 + e^{-2|\alpha|^2})}}, \quad (4.9)$$

and the compass state [Zur01]

$$|\text{cat}_4(\alpha)\rangle = \frac{|\alpha\rangle + |-\alpha\rangle + |i\alpha\rangle + |-i\alpha\rangle}{2\sqrt{1 + e^{-|\alpha|^2}(2\cos(|\alpha|^2) + 1)}}, \quad (4.10)$$

where  $|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{k \geq 0} \frac{\alpha^k}{\sqrt{k!}} |k\rangle$  is the coherent state of amplitude  $\alpha \in \mathbb{C}$ . We have

$$|\langle 2 | \text{cat}_2(\alpha) \rangle|^2 = \frac{|\alpha|^4}{2 \cosh(|\alpha|^2)}, \quad (4.11)$$

and

$$|\langle 4 | \text{cat}_4(\alpha) \rangle|^2 = \frac{|\alpha|^8/12}{\cosh(|\alpha|^2) + \cos(|\alpha|^2)}. \quad (4.12)$$

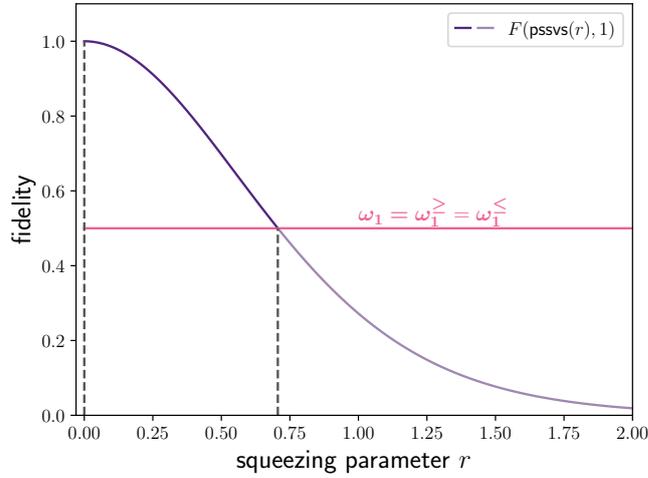


Figure 4.2: Fidelities of photon-subtracted squeezed vacuum states  $|\text{p-ssvs}(r)\rangle$  with the Fock state  $|1\rangle$  with respect to the squeezing parameter  $r \in \mathbb{R}$ . The dashed grey lines delimit the intervals of parameter values where the witness  $|1\rangle\langle 1|$  can be used to detect Wigner negativity, i.e. when the fidelity (violet curve) is above the witness upper bound (pink line). The height difference when the violet curve is above the pink line directly provides a lower bound on the distance between the corresponding state and the set of states having a positive Wigner function.

For  $1.63 \leq |\alpha|^2 \leq 2.59$ , the fidelity  $F(\text{cat}_2(\alpha), 2)$  in Eq. (4.11) is greater than  $\omega_2^{\geq} = \frac{1}{2}$  and our witness corresponding to  $n = 2$  can be used to detect Wigner negativity of this state. Similarly, for  $2.10 \leq |\alpha|^2 \leq 6.53$ , the fidelity  $F(\text{cat}_4(\alpha), 4)$  in Eq. (4.12) is greater than  $\omega_4^{\geq} = 0.441$  and our witness corresponding to  $n = 4$  can be used to detect Wigner negativity of this state (see Fig. 4.3).

Some quantum states will remain unnoticed by all single Fock state negativity witnesses. For example, the state  $\rho_{0,1,2} := \frac{1}{9} |0\rangle\langle 0| + \frac{4}{9} |1\rangle\langle 1| + \frac{4}{9} |2\rangle\langle 2|$  has a negative Wigner function but is not detected by any of the single Fock state negativity witnesses, since the lower bounds for  $n = 1, 2$  in Table 4.1 are higher than  $\frac{4}{9}$ , and this state has fidelity 0 with higher Fock states. However, it is detected by simple witnesses based on linear combinations of fidelities. For example, with  $n = 2$  and  $\mathbf{a} = (1, 1)$  we find numerically that the threshold value of the witness  $|1\rangle\langle 1| + |2\rangle\langle 2|$  is less than 0.875 when running the corresponding program  $(\text{SDP}_{\mathbf{a}}^{m, \geq})$  for  $m = 7$ . And since  $\text{Tr}(\rho_{0,1,2}(|1\rangle\langle 1| + |2\rangle\langle 2|)) = \frac{8}{9} > 0.875$ , this linear combination of Fock state fidelities can indeed detect Wigner negativity for this state.

However, some quantum states with a negative Wigner function will always go unnoticed with witnesses of the previous form because these witnesses are invariant under phase-space rotations while the Wigner function of those states becomes positive under random dephasing. Consider for instance the superposition  $\sqrt{1 - \frac{1}{s}} |0\rangle + \frac{1}{\sqrt{s}} |1\rangle$ , for  $s > 2$  (which under random dephasing is mapped to  $(1 - \frac{1}{s}) |0\rangle\langle 0| + \frac{1}{s} |1\rangle\langle 1|$ ). In that case, the Wigner negativity of such

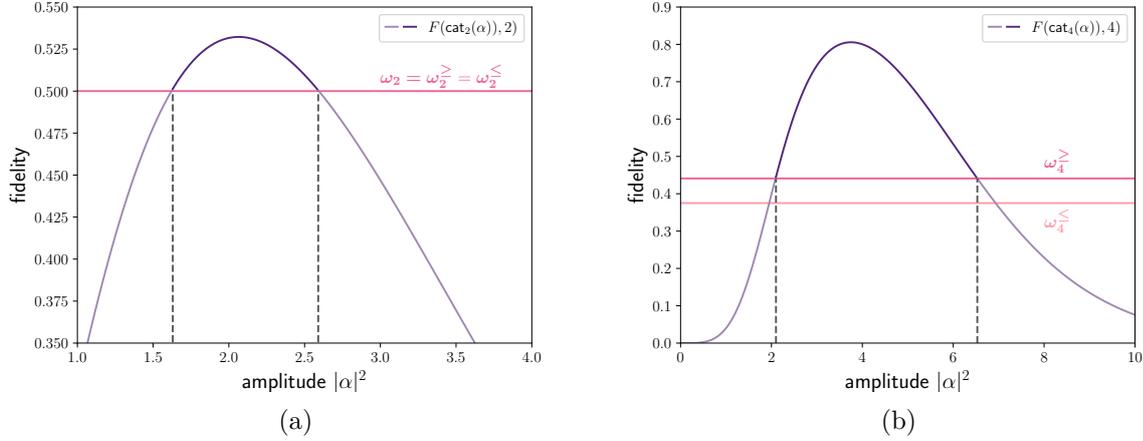


Figure 4.3: (a) Fidelities of cat states  $|\text{cat}_2(\alpha)\rangle$  with amplitude  $\alpha \in \mathbb{C}$  with the Fock state  $|2\rangle$ . (b) Fidelities of compass states  $|\text{cat}_4(\alpha)\rangle$  with amplitude  $\alpha \in \mathbb{C}$  with the Fock state  $|4\rangle$ . The dashed grey lines delimit the intervals of amplitude where our witnesses from Table 4.1 can be used to detect Wigner negativity of the corresponding state. When it is below the witness lower bound (light pink line), we are guaranteed that the witness (here  $|4\rangle\langle 4|$ ) cannot be used to detect Wigner negativity of the state.

states can still be witnessed by using the displaced version of our witnesses. In particular, if any single-mode quantum state has a Wigner function negative at  $\alpha \in \mathbb{C}$ , then there is a choice of  $n \in \mathbb{N}^*$  such that the witness in Eq. (4.1) defined by  $\mathbf{a} = (1, 0, 1, 0, 1, \dots)$  and the displacement amplitude  $\alpha$  detects its negativity [CRW<sup>+</sup>21]. In practice, simpler witnesses may suffice to detect negativity, and the choice of witness will ultimately depend on the experimental state at hand. Hereafter we discuss the heuristics for picking a good witness, with the theoretical example of the lossy 3-photon Fock state:

$$\begin{aligned} \rho_{3,\eta} := & (1 - \eta)^3 |3\rangle\langle 3| + 3\eta(1 - \eta)^2 |2\rangle\langle 2| \\ & + 3\eta^2(1 - \eta) |1\rangle\langle 1| + \eta^3 |0\rangle\langle 0|, \end{aligned} \quad (4.13)$$

where  $0 \leq \eta \leq 1$  is the loss parameter. Setting  $\eta = 0$  gives  $\rho_{3,\eta} = |3\rangle\langle 3|$  while setting  $\eta = 1$  gives  $\rho_{3,\eta} = |0\rangle\langle 0|$ . This state has a nonnegative Wigner function for  $\eta \geq \frac{1}{2}$ . The fidelities of  $\rho_{3,\eta}$  with displaced Fock states  $\hat{D}(\alpha)|l\rangle$  are given by:

$$\begin{aligned} F(\rho_{3,\eta}, \hat{D}(\alpha)|l\rangle) &= (1 - \eta)^3 |\langle 3|\hat{D}(\alpha)|l\rangle|^2 \\ &+ 3\eta(1 - \eta)^2 |\langle 2|\hat{D}(\alpha)|l\rangle|^2 \\ &+ 3\eta^2(1 - \eta) |\langle 1|\hat{D}(\alpha)|l\rangle|^2 \\ &+ \eta^3 |\langle 0|\hat{D}(\alpha)|l\rangle|^2. \end{aligned} \quad (4.14)$$

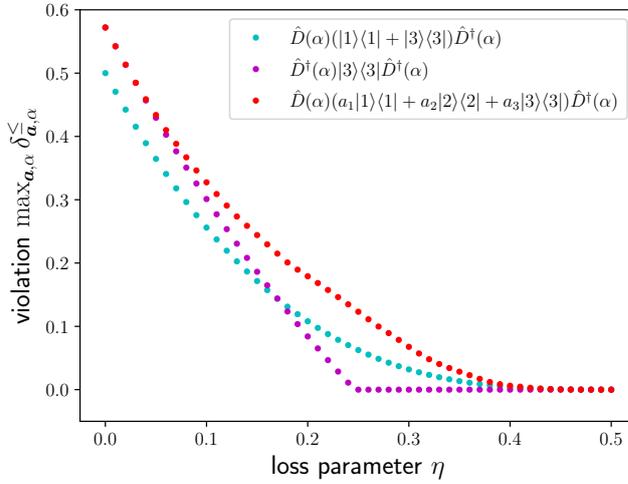


Figure 4.4: Lower bounds on the violation  $\delta_{\mathbf{a},\alpha} = \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho_{3,\eta}) - \omega_{\mathbf{a}}$  for a lossy 3-photon Fock state  $\rho_{3,\eta} = \eta^3 |3\rangle\langle 3| + 3\eta^2(1-\eta) |2\rangle\langle 2| + 3\eta(1-\eta)^2 |1\rangle\langle 1| + (1-\eta)^3 |0\rangle\langle 0|$  for  $\mathbf{a} = (a_1, a_2, a_3)$  with respect to the loss parameter  $\eta$ . Precomputed bounds on threshold values for witnesses of the form  $\hat{\Omega}_{\mathbf{a}} = a_1 |1\rangle\langle 1| + a_2 |2\rangle\langle 2| + a_3 |3\rangle\langle 3|$  can be found in [CE]. We use these values to find the witness  $\hat{\Omega}_{\mathbf{a},\alpha}$  giving the maximum lower bound  $\delta_{\mathbf{a},\alpha}^{\leq} = \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho_{3,\eta}) - \omega_{\mathbf{a}}^{\geq}$  on the violation  $\delta_{\mathbf{a},\alpha}$ , for different values of the loss parameter  $\eta$ . These optimised lower bounds are represented in red. In blue is the maximal violation that can be detected using the witnesses  $\hat{\Omega}_{(1,0,1),\alpha} = \hat{D}(\alpha)(|1\rangle\langle 1| + |3\rangle\langle 3|)\hat{D}^\dagger(\alpha)$  [CRW<sup>+</sup>21]. In violet is the maximal violation that can be detected using the more naive witness  $\hat{\Omega}_{(0,0,1),\alpha} = D(\alpha) |3\rangle\langle 3| D^\dagger(\alpha)$ . Note that  $\rho_{3,\eta}$  has a nonnegative Wigner function for  $\eta \geq 0.5$ .

where the coefficients of the displacement operator in Fock basis are given in Eq. (1.24). In an experimental scenario, the state would be unknown and these fidelities should be estimated using samples from a homodyne or heterodyne detection of the state translated by  $\alpha$ , and estimating the fidelities with Fock states [CGKM20, CRW<sup>+</sup>21].

Following the heuristic detailed in the previous section, we have determined suitable Wigner negativity witnesses for 50 values of the loss parameter  $\eta$  between 0 and 0.5 as follows: for each value  $\eta$ , we have computed numerically the values of the fidelities in Eq. (4.14) for  $l = 1, 2, 3$ , and for displacement parameters  $\alpha = q/10 + ip/10$  for all  $q, p \in \llbracket 0, 10 \rrbracket$ . Using these values, we have computed the expectation value of the witnesses  $\hat{\Omega}_{\mathbf{a},\alpha}$  for multiple choices of  $\mathbf{a} = (a_1, a_2, a_3)$  with  $\max_i a_i = 1$ . We have used the corresponding precomputed bounds  $\omega_{\mathbf{a}}^{\geq}$  on the threshold values of  $\hat{\Omega}_{\mathbf{a},\alpha}$  in [CE] to determine the witness leading to the maximal lower bound  $\delta_{\mathbf{a},\alpha}^{\leq} := \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho_{3,\eta}) - \omega_{\mathbf{a}}^{\geq}$  on the violation  $\delta_{\mathbf{a},\alpha} = \text{Tr}(\hat{\Omega}_{\mathbf{a},\alpha} \rho_{3,\eta}) - \omega_{\mathbf{a}}$  over the choice of  $(\mathbf{a}, \alpha)$ .

We have represented these violations for each value of the loss parameter  $\eta$  in Fig. 4.4.

For all values of  $\eta$ , we find that the optimal displacement parameter is  $\alpha = 0$ . On the other hand, we find different optimal choices of  $\mathbf{a}$  for different values of  $\eta$ . To illustrate the usefulness of the optimisation over the choice of witnesses parametrised by  $(\mathbf{a}, \alpha)$ , we have also represented the violations obtained when using the witnesses  $\hat{\Omega}_{(1,0,1),\alpha} = \hat{D}(\alpha)(|1\rangle\langle 1| + |3\rangle\langle 3|)\hat{D}^\dagger(\alpha)$  from [CRW<sup>+</sup>21] for all values of  $\eta$ . In that setting, the violations obtained quantify how hard it is to detect the Wigner negativity of the state: a larger violation implies that a less precise estimate of the witness expectation value is needed to witness Wigner negativity. In particular, we obtain that our optimised witnesses always provide a greater violation to detect negativity than the previous witnesses which will result in an easier experimental detection. We also represented the violation obtained when using the more naive witnesses  $\hat{\Omega}_{(0,0,1),\alpha} = \hat{D}(\alpha)|3\rangle\langle 3|\hat{D}^\dagger(\alpha)$  and we see that it is only useful when the loss parameter is smaller than 0.25, while the optimised witnesses may detect negativity of the state  $\rho_{3,\eta}$  up to  $\eta = 0.5$ —when the Wigner functions becomes nonnegative—provided the estimates of the fidelities are precise enough.

Overall, this procedure only amounts to a simple classical post-processing of samples from homodyne or heterodyne detection and yields a good witness for detecting Wigner negativity.

### 4.2.3 Numerical implementation

Here we discuss numerical implementations of the semidefinite programs ( $\text{SDP}_{\mathbf{a}}^{m,\leq}$ ) and ( $\text{SDP}_{\mathbf{a}}^{m,\geq}$ ). All codes are available [here](#) [CE].

We implemented the semidefinite programs with Python through the interface provided by PICOS [SS12]. We first used the solver Mosek [ApS19] to solve these problems but, while the size of the semidefinite programs remains relatively low for small values of  $n$  and  $m$ , binomial terms grow rapidly and numerical precision issues arise quickly (usually around  $m = 12$ ,  $n \leq m$  on a laptop). The linear constraints involving  $Q_{ij}$  in the semidefinite programs come from a polynomial equality (see Lemma 4.5). While polynomial equalities are usually written in the canonical basis, a first trick is to express them in a different basis—for instance the basis  $(1, \frac{X}{1!}, \frac{X^2}{2!}, \dots)$ —to counterbalance the binomial terms.

However, this may not be sufficient to probe larger values of  $m$ . Instead, we used the solver SDPA-GMP [Nak10, FKNY02] which allows arbitrary precision arithmetic. While much slower, this solver is dedicated to solve problems requiring a lot of precision. Because our problems remain rather small, time efficiency is not an issue and this solver is particularly well-suited. All problems were initially solved on a regular laptop as warning flags on optimality were raised before the problems were too large. A high-performance computer<sup>2</sup> handling floating point arithmetic more accurately was later used to compute further ranks in the hierarchy.

Using the semidefinite programs ( $\text{SDP}_{\mathbf{a}}^{m,\geq}$ ) and ( $\text{SDP}_{\mathbf{a}}^{m,\leq}$ ) for values of  $m$  up to around 30 and  $\mathbf{a} = (0, 0, \dots, 0, 1)$  (where the size  $n$  of the vector  $\mathbf{a}$  is ranging from 1 to 10), we have

<sup>2</sup>DELL PowerEdge R440, 384 Gb RAM, Intel Xeon Silver 4216 processor, 64 threads from LIP6.

obtained upper and lower bounds for the threshold values of Wigner negativity witnesses corresponding to fidelities with Fock states from 1 to 10, reported in Table 4.1.

We also computed upper and lower bounds on the threshold values of witnesses of the form:

$$\hat{\Omega}_{(a_1, a_2, a_3)} = \sum_{k=1}^3 a_k |k\rangle\langle k|, \quad (4.15)$$

where  $\forall i \in \{1, 2, 3\}$ ,  $0 \leq a_i \leq 1$  and  $\max_i a_i = 1$ . We focused on these particular witnesses for experimental considerations as it is challenging to obtain fidelities with higher Fock states. We fix one coefficient equal to 1 and vary each other  $a_i$  from 0 to 1 with a step of 0.1. The resulting bounds on the threshold values can be found in [CE] and in [CEG21].

We now turn to the mathematical proofs of our results, i.e., that the threshold values in Eq. (4.3) can be upper bounded and lower bounded by the optimal values of the converging hierarchies of semidefinite programs  $(\text{SDP}_{\mathbf{a}}^{m, \geq})_{m \geq n}$  and  $(\text{SDP}_{\mathbf{a}}^{m, \leq})_{m \geq n}$ , respectively.

The following section is rather technical as we dive into infinite-dimensional optimisation techniques to derive the hierarchies of semidefinite programs and prove their convergence. Some readers may want to skip directly to Section 4.4.

### 4.3 Infinite-dimensional optimisation

In this section we use infinite-dimensional optimisation techniques: (i) to phrase the computation of the witness threshold value introduced in Eq. (4.3) as an infinite-dimensional linear program in Section 4.3.3, (ii) to derive two hierarchies of finite-dimensional semidefinite programs that upper bound and lower bound the threshold value in Section 4.3.4, and (iii) to show, in Section 4.3.5, that the sequence of upper bounds converges to the threshold value computed over  $L^2(\mathbb{R}_+)$  functions and the sequence of lower bounds converges to the threshold value computed over Schwartz functions (see Fig. 4.5). Given the technicalities of the proofs above, we sketch them in Section 4.3.1 before detailing them in the following sections.

As a convention, except if specifically mentioned, we will use the terminology ‘relaxation’ and ‘restriction’ from the point of view of the primal program. We will refer to the hierarchy of semidefinite programs providing the upper bounds as a *hierarchy of relaxations* because the obtained SDP programs are indeed relaxations of the primal program (while they are restrictions of the dual program). Likewise, we will refer to the hierarchy providing the lower bounds as a *hierarchy of restrictions*.

For clarity, we treat the case where the witnesses are given by the fidelity with a single Fock state, corresponding to the case where one entry of the vector  $\mathbf{a}$  is equal to 1 and all the other entries are 0. The generalisation to linear combinations of fidelities with Fock states is straightforward by linearity.

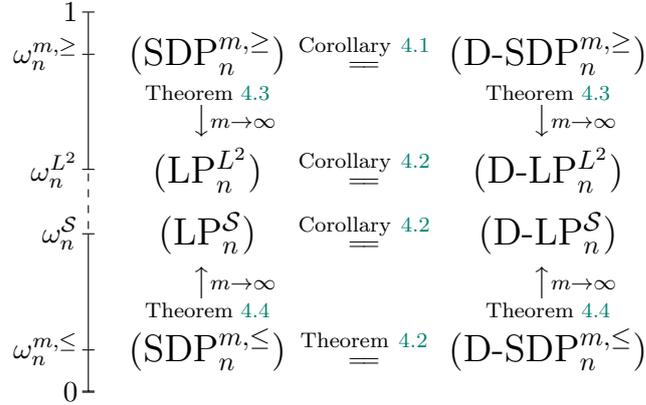


Figure 4.5: Hierarchy of semidefinite relaxations converging to the linear program  $(\text{LP}_n^{L^2})$  and hierarchy of semidefinite restrictions converging to the linear program  $(\text{LP}_n^S)$ , together with their dual programs. The upper index  $m$  denotes the level of the relaxation or restriction. On the left are the associated optimal values. The equal sign denotes strong duality, i.e., equality of optimal values, and the arrows denote convergence of the corresponding sequences of optimal values. Note that the question of the closing the gap between the values of  $(\text{LP}_n^{L^2})$  and  $(\text{LP}_n^S)$  is left open.

### 4.3.1 Sketch of the proofs

This section aims at giving an overview of the rather technical proofs that follow. Our reasoning can be split into the following steps:

#### Obtaining an infinite-dimensional linear program and its dual

- We exploit the phase-space rotational invariance of Fock states in Lemma 4.4 to express the computation of threshold values Eq. (4.3) on states that are diagonal in the Fock basis.
- This provides a maximisation problem which can be rephrased as an infinite-dimensional linear program  $(\text{LP}_n^{L^2})$  (resp.  $(\text{LP}_n^S)$ ) on the space of square integrable functions  $L^2(\mathbb{R}_+)$  (resp. Schwartz functions  $\mathcal{S}(\mathbb{R}_+)$ ). We denote  $\omega_n^{L^2}$  its optimal value (resp.  $\omega_n^S$ ).
- By duality, we recast this as an optimisation problem on finite-signed measures which is given by the dual infinite-dimensional linear program  $(\text{D-LP}_n^{L^2})$  (resp.  $(\text{D-LP}_n^S)$ ).

#### Hierarchy of relaxations

See upper part of figure 4.5.

- We relax the program  $(\text{LP}_n^{L^2})$ : instead of optimising over positive functions, we optimise over functions that have a positive inner product with positive polynomials of fixed degree  $m$ . The intuition for this procedure is given by Theorem 4.1. This degree fixes a level

within a hierarchy of relaxations. These constraints can be cast as a positive semidefinite constraint with Lemma 4.7 and we thus obtain a hierarchy of semidefinite programs (see  $(\text{SDP}_n^{m,\geq})$ ).

- For each level  $m$ , we derive the dual program  $(\text{D-SDP}_n^{m,\geq})$  and show that strong duality holds in Theorem 4.1 by finding a strictly feasible solution of  $(\text{SDP}_n^{m,\geq})$ .
- We prove the convergence of the hierarchy of semidefinite relaxations towards  $(\text{D-LP}_n^{L^2})$ <sup>3</sup> in Theorem 4.3 by first showing that the feasible set of  $(\text{SDP}_n^{m,\geq})$  is compact. Then, we perform a diagonal extraction on a sequence of optimal solutions of  $(\text{SDP}_n^{m,\geq})$  and we finally show that this provides a feasible solution of  $(\text{LP}_n^{L^2})$  proving that  $\lim_{m \rightarrow +\infty} \omega_n^{m,\geq} = \omega_n^{L^2}$ .

### Convergence of hierarchy of restrictions

See bottom part of figure 4.5.

- We restrict the program  $(\text{LP}_n^{L^2})$  (or equivalently the program  $(\text{LP}_n^{L^2})$ ): instead of optimising over positive functions, we optimise over positive polynomials of fixed degree  $m$ . Again, this degree fixes a level within a hierarchy of restrictions. Using the fact that univariate positive polynomials are sum-of-squares which can be written as a semidefiniteness constraint, we obtain a hierarchy of semidefinite programs (see  $(\text{SDP}_n^{m,\leq})$ ).
- For each level  $m$ , we derive the dual program  $(\text{D-SDP}_n^{m,\leq})$  and show that strong duality holds in Theorem 4.2 by finding a strictly feasible solution of  $(\text{SDP}_n^{m,\leq})$ .
- We prove the convergence of the hierarchy towards  $(\text{D-LP}_n^S)$  in Theorem 4.4 by first showing that the feasible set of  $(\text{D-SDP}_n^{m,\leq})$  is compact. This is highly nontrivial since it requires exhibiting an analytical feasible solution of  $(\text{SDP}_n^{m,\leq})$  which is difficult in general and in the present case. Then, we perform a diagonal extraction on a sequence of optimal solutions of  $(\text{D-SDP}_n^{m,\leq})$ . A technicality arises as it does not necessarily provide a feasible solution of  $(\text{D-LP}_n^{L^2})$  which is why we introduce the linear program expressed over Schwartz functions. We then show that diagonal extraction indeed provides a feasible solution of  $(\text{D-LP}_n^S)$  proving that  $\lim_{m \rightarrow +\infty} \omega_n^{m,\leq} = \omega_n^S$ .

### 4.3.2 Function spaces

We now review some function spaces which appear in the following sections, together with a few notations.

The half line of nonnegative real numbers is denoted  $\mathbb{R}_+$ . The space of real square-integrable functions over  $\mathbb{R}_+$  is denoted  $L^2(\mathbb{R}_+)$  and is equipped with the usual inner product:

$$\langle f, g \rangle = \int_{\mathbb{R}_+} f(x)g(x)dx, \tag{4.16}$$

---

<sup>3</sup>By ‘convergence of the hierarchy of semidefinite relaxations towards  $(\text{D-LP}_n^{L^2})$ ’ we mean that the sequence obtained by taking the values of the programs in the hierarchy, and indexed by this hierarchy, converges to the value of  $(\text{D-LP}_n^{L^2})$

for  $f, g \in L^2(\mathbb{R}_+)$ . This space is isomorphic to the space of square-summable real sequences indexed by  $\mathbb{N}$  denoted  $l^2(\mathbb{N})$ , by considering the expansion in a countable basis. Such a basis is given, e.g., by the Laguerre functions [Sze59] (see Eq. (1.34)), modified here by a  $(-1)^k$  prefactor to correspond to Fock state Wigner functions. We recall here its expression

$$\mathcal{L}_k(x) := (-1)^k L_k(x) e^{-\frac{x}{2}}, \quad (4.17)$$

for all  $k \in \mathbb{N}$  and all  $x \in \mathbb{R}_+$ , where  $L_k(x) = \sum_{l=0}^k \frac{(-1)^l}{l!} \binom{k}{l} x^l$  is the  $k^{\text{th}}$  Laguerre polynomial. These functions form an orthonormal basis: for all  $p, q \in \mathbb{N}$ ,  $\langle \mathcal{L}_p, \mathcal{L}_q \rangle = \delta_{pq}$ .

The space  $L^2(\mathbb{R}_+)$  is also isomorphic to its dual space  $L^{2'}(\mathbb{R}_+)$ : via the Radon–Nikodym theorem [Nik30] elements of  $L^{2'}(\mathbb{R}_+)$  can be identified by the Lebesgue measure on  $\mathbb{R}_+$  times the corresponding function in  $L^2(\mathbb{R}_+)$ .

We write  $\mathcal{S}(\mathbb{R}_+)$  the space of Schwartz functions over  $\mathbb{R}_+$ , i.e., the space of  $C^\infty$  functions that go to 0 at infinity faster than any inverse polynomial, as do their derivatives.  $\mathcal{S}'(\mathbb{R}_+)$  is its dual space, the space of tempered distributions over  $\mathbb{R}_+$ .  $\mathcal{S}(\mathbb{R}_+) \subset L^2(\mathbb{R}_+)$  is dense in  $L^2(\mathbb{R}_+)$ . We denote the space of rapidly decreasing real sequences by  $\mathcal{S}(\mathbb{N})$  (sequences that go to 0 at infinity faster than any inverse polynomial), together with its dual space of slowly increasing real sequences  $\mathcal{S}'(\mathbb{N})$  (sequences that are upper bounded by a polynomial). The spaces  $\mathcal{S}(\mathbb{R}_+)$  and  $\mathcal{S}(\mathbb{N})$  are isomorphic: any Schwartz function over  $\mathbb{R}_+$  can be expanded uniquely in the basis of Laguerre functions with a rapidly decreasing sequence of coefficients. Similarly, the spaces  $\mathcal{S}'(\mathbb{R}_+)$  and  $\mathcal{S}'(\mathbb{N})$  are also isomorphic: any tempered distribution over  $\mathbb{R}_+$  can be written uniquely as a formal series of Laguerre functions with a slowly increasing sequence of coefficients [GT71]. We extend the definition of the duality  $\langle -, - \rangle$  in Eq. (4.16) to these spaces.

In order to denote nonnegative elements of these spaces, we will use the notations  $L_+^2(\mathbb{R}_+)$ ,  $L_+^{2'}(\mathbb{R}_+)$ ,  $\mathcal{S}_+(\mathbb{R}_+)$  and  $\mathcal{S}'_+(\mathbb{R}_+)$ . A distribution  $\mu$  in  $L_+^{2'}(\mathbb{R}_+)$  (resp. in  $\mathcal{S}'_+(\mathbb{R}_+)$ ) satisfies:  $\forall f \in L_+^2(\mathbb{R}_+)$  (resp.  $\forall f \in \mathcal{S}_+(\mathbb{R}_+)$ ),  $\langle \mu, f \rangle \geq 0$ .

For all  $m \in \mathbb{N}$ , we define the following space of truncated series of Laguerre functions over  $\mathbb{R}_+$ :

$$\mathcal{R}_m(\mathbb{R}_+) := \text{span}_{\mathbb{R}} \{ \mathcal{L}_k \}_{0 \leq k \leq m}, \quad (4.18)$$

which is equal to the set of real polynomials over  $\mathbb{R}_+$  of degree less or equal to  $m$  multiplied by the function  $x \mapsto e^{-\frac{x}{2}}$ . We denote by  $\mathcal{R}_{m,+}(\mathbb{R}_+)$  its subset of nonnegative elements.

For all  $\mathbf{s} = (s_k)_k \in \mathbb{R}^{\mathbb{N}}$ , we define the associated formal series of Laguerre functions:

$$f_{\mathbf{s}} := \sum_{k \geq 0} s_k \mathcal{L}_k, \quad (4.19)$$

with the (formal) relation:

$$s_k = \langle f_{\mathbf{s}}, \mathcal{L}_k \rangle, \quad (4.20)$$

for all  $k \in \mathbb{N}$ . We refer to  $\mathbf{s}$  as the sequence of Laguerre moments of  $f_{\mathbf{s}}$ . We extend this definition to finite sequences by completing these sequences with zeros. For  $m \in \mathbb{N}$ , we also define the

matrix  $A_s$  (thus omitting the dependence in  $m$  as it is always clear from the context) by

$$(A_s)_{0 \leq i, j \leq m} := \begin{cases} \sum_{k=0}^l s_k \binom{l}{k} l! & \text{if } i + j = 2l, \\ 0 & \text{otherwise.} \end{cases} \quad (4.21)$$

$A_s$  can be seen as the Laguerre moment matrix of the measure  $f_s$ . In what follows, we use standard techniques relating to the Stieltjes moment problem [RS75], which seeks conditions for a real sequence  $\nu = (\nu_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  to be the sequence of moments  $\int_{\mathbb{R}_+} x^k d\nu(x)$  of a nonnegative distribution  $\nu$  over  $\mathbb{R}_+$ . We adapt these techniques to the basis of Laguerre functions, rather than the canonical basis. We start by proving a change of basis result:

**Lemma 4.2.** *Let  $\mu, \nu \in \mathbb{R}^{\mathbb{N}}$ . For all  $m \in \mathbb{N}$ , the following conditions are equivalent:*

$$(i) \quad \forall k \in \llbracket 0, m \rrbracket, \quad \mu_k = \sum_{l=0}^k \nu_l \frac{(-1)^{k+l}}{l!} \binom{k}{l},$$

$$(ii) \quad \forall l \in \llbracket 0, m \rrbracket, \quad \nu_l = \sum_{k=0}^l \mu_k \binom{l}{k} l!.$$

As a direct consequence, we retrieve the formula:

$$x^l = \sum_{k=0}^l (-1)^k \binom{l}{k} l! L_k(x), \quad (4.22)$$

for all  $l \in \mathbb{N}$  and all  $x \in \mathbb{R}_+$ .

*Proof.* (i) $\Rightarrow$ (ii): suppose that

$$\forall k \in \llbracket 0, m \rrbracket, \quad \mu_k = \sum_{p=0}^k \nu_p \frac{(-1)^{k+p}}{p!} \binom{k}{p}. \quad (4.23)$$

$$\begin{aligned} \forall l \in \llbracket 0, m \rrbracket \quad \sum_{k=0}^l \mu_k \binom{l}{k} l! &= \sum_{k=0}^l \sum_{p=0}^k \nu_p \frac{(-1)^{k+p}}{p!} \binom{k}{p} \binom{l}{k} l! \\ &= \sum_{p=0}^l \nu_p \frac{l!}{p!} \binom{l}{p} \sum_{k=p}^l (-1)^{k-p} \binom{l-p}{k-p} \\ &= \sum_{p=0}^l \nu_p \frac{l!}{p!} \binom{l}{p} \sum_{q=0}^{l-p} (-1)^q \binom{l-p}{q} \\ &= \nu_l, \end{aligned} \quad (4.24)$$

where we used Eq. (4.23) in the first line and the binomial theorem in the last line which imposes  $l = p$ .

(ii) $\Rightarrow$ (i): suppose that

$$\forall l \in \llbracket 0, m \rrbracket, \quad \nu_l = \sum_{p=0}^l \mu_p \binom{l}{p} l!. \quad (4.25)$$

$$\begin{aligned}
 \forall k \in \llbracket 0, m \rrbracket, \quad \sum_{l=0}^k \nu_l \frac{(-1)^{k+l}}{l!} \binom{k}{l} &= \sum_{l=0}^k \sum_{p=0}^l \mu_p (-1)^{k+l} \binom{l}{p} \binom{k}{l} \\
 &= \sum_{p=0}^k \mu_p (-1)^{k+p} \binom{k}{p} \sum_{l=p}^k (-1)^{l-p} \binom{k-p}{l-p} \\
 &= \sum_{p=0}^k \mu_p (-1)^{k+p} \binom{k}{p} \sum_{q=0}^{k-p} (-1)^q \binom{k-p}{q} \\
 &= \mu_k,
 \end{aligned} \tag{4.26}$$

where we used Eq. (4.25) in the first line and the binomial theorem in the last line which imposes  $k = p$ . □

We prove a similar result that we will use later for proving that a pair of semidefinite programs are indeed dual programs.

**Lemma 4.3.** *Let  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{m+1}$ . The following propositions are equivalent:*

$$\begin{aligned}
 (i) \quad \forall k \in \llbracket 0, m \rrbracket, \quad u_k &= \sum_{l=0}^m v_l \binom{l}{k} l!, \\
 (ii) \quad \forall l \in \llbracket 0, m \rrbracket, \quad v_l &= \sum_{k=l}^m \frac{(-1)^{l+k}}{l!} \binom{k}{l} u_k.
 \end{aligned}$$

*Proof.* The proof is similar to that of Lemma 4.2. Note that we could start the first sum at  $l = k$  since  $\binom{l}{k} = 0$  for  $l < k$  but for convenience we start it at  $l = 0$ .

(i)  $\Rightarrow$  (ii): suppose that

$$\forall k \in \llbracket 0, m \rrbracket, \quad u_k = \sum_{p=0}^m v_p \binom{p}{k} p!. \tag{4.27}$$

$$\begin{aligned}
 \forall l \in \llbracket 0, m \rrbracket \frac{(-1)^l}{l!} \sum_{k=l}^m (-1)^k \binom{k}{l} u_k &= \frac{(-1)^l}{l!} \sum_{k=l}^m (-1)^k \binom{k}{l} \sum_{p=0}^m v_p \binom{p}{k} p! \\
 &= \sum_{p=0}^m v_p \frac{p!}{l!} \sum_{k=l}^m (-1)^{k+l} \binom{p}{k} \binom{k}{l} \\
 &= \sum_{p=l}^m v_p \frac{p!}{l!} \sum_{k=l}^p (-1)^{k+l} \binom{p}{k} \binom{k}{l} \\
 &= \sum_{p=l}^m v_p \frac{p!}{l!} \sum_{k=l}^p (-1)^{k+l} \frac{p!k!}{k!(p-k)!l!(k-l)!} \\
 &= \sum_{p=l}^m v_p \frac{p!}{l!} \binom{p}{l} \sum_{q=0}^{p-l} (-1)^q \binom{p-l}{q} \\
 &= v_l,
 \end{aligned} \tag{4.28}$$

where we used Eq. (4.27) in the first line, the fact that  $\binom{p}{k} = 0$  if  $k > p$  in the third line,  $q := k - l$  in the fifth line, and the binomial theorem in the last line which imposes  $p = l$ .

(ii) $\Rightarrow$ (i): suppose that

$$\forall l \in \llbracket 0, m \rrbracket, \quad v_l = \frac{(-1)^l}{l!} \sum_{p=l}^m (-1)^p \binom{p}{l} u_p. \quad (4.29)$$

$$\begin{aligned} \forall k \in \llbracket 0, m \rrbracket \sum_{l=0}^m v_l \binom{l}{k} l! &= \sum_{l=0}^m \frac{(-1)^l}{l!} \sum_{p=l}^m (-1)^p \binom{p}{l} u_p \binom{l}{k} l! \\ &= \sum_{p=0}^m u_p (-1)^p \sum_{l=0}^p (-1)^l \binom{p}{l} \binom{l}{k} \\ &= \sum_{p=k}^m u_p (-1)^p \sum_{l=k}^p (-1)^l \binom{p}{l} \binom{l}{k} \\ &= \sum_{p=k}^m u_p (-1)^p \sum_{l=k}^p (-1)^l \frac{p! l!}{l!(p-l)!k!(l-k)!} \\ &= \sum_{p=k}^m u_p (-1)^{p-k} \binom{p}{k} \sum_{q=0}^{p-k} (-1)^q \binom{p-k}{q} \\ &= u_k, \end{aligned} \quad (4.30)$$

where we used Eq. (4.29) in the first line, the fact that  $\binom{l}{k} = 0$  if  $k > l$  in the third line,  $q := l - k$  in the fifth line, and the binomial theorem in the last line which imposes  $p = k$ .  $\square$

Now we turn our attention to a modified Stieltjes moment condition.

**Theorem 4.1.** *Let  $\boldsymbol{\mu} = (\mu_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ . The sequence  $\boldsymbol{\mu}$  is the sequence of Laguerre moments  $\int_{\mathbb{R}_+} \mathcal{L}_k(x) d\mu(x)$  of a nonnegative distribution  $\mu$  supported on  $\mathbb{R}_+$  if and only if*

$$\forall m \in \mathbb{N}, \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_{\boldsymbol{\mu}}, g \rangle \geq 0. \quad (4.31)$$

This result is based on the well-known Riesz–Haviland theorem [Rie23, Hav36] (see Theorem 1.3). The Riesz functional  $L_{\boldsymbol{\nu}}$  for a sequence  $\boldsymbol{\nu} = (\nu_l)_{l \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  is defined in Definition 1.14.

*Proof.* Let  $\boldsymbol{\mu} = (\mu_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ , and suppose that the sequence  $\boldsymbol{\mu}$  is the sequence of Laguerre moments  $\int_{\mathbb{R}_+} \mathcal{L}_k(x) d\mu(x)$  of a nonnegative distribution  $\mu$  supported on  $\mathbb{R}_+$ . Let  $m \geq 0$  and let  $g = \sum_{k=0}^m g_k \mathcal{L}_k \in \mathcal{R}_{m,+}(\mathbb{R}_+)$ . The distribution  $\mu$  is nonnegative, so by definition  $\langle \mu, g \rangle \geq 0$ . Moreover,

$$\begin{aligned} \langle f_{\boldsymbol{\mu}}, g \rangle &= \sum_{k=0}^m \mu_k g_k \\ &= \int_{\mathbb{R}_+} \sum_{k=0}^m g_k \mathcal{L}_k d\mu \\ &= \langle \mu, g \rangle. \end{aligned} \quad (4.32)$$

Hence, for all  $m \in \mathbb{N}$  and all  $g \in \mathcal{R}_{m,+}(\mathbb{R}_+)$ ,  $\langle f_{\boldsymbol{\mu}}, g \rangle \geq 0$ .

Conversely, let  $\boldsymbol{\mu} = (\mu_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ , and suppose that for all  $m \in \mathbb{N}$  and all  $g \in \mathcal{R}_{m,+}(\mathbb{R}_+)$ ,  $\langle f_{\boldsymbol{\mu}}, g \rangle \geq 0$ . We define the sequence  $\boldsymbol{\nu} = (\nu_l)_{l \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  by

$$\nu_l := \sum_{k=0}^l \mu_k \binom{l}{k} l!, \quad (4.33)$$

for all  $l \in \mathbb{N}$ .

Let  $m \in \mathbb{N}$  and let  $P(x) = \sum_{l=0}^m p_l x^l$  be a nonnegative polynomial over  $\mathbb{R}_+$ . By Eq. (4.22), for all  $x \in \mathbb{R}_+$ ,

$$\begin{aligned} P(x) &= \sum_{l=0}^m p_l \sum_{k=0}^l (-1)^k \binom{l}{k} l! L_k(x) \\ &= \sum_{k=0}^m (-1)^k L_k(x) \left( \sum_{l=k}^m p_l \binom{l}{k} l! \right). \end{aligned} \quad (4.34)$$

Let  $g_P(x) := P(x)e^{-\frac{x}{2}}$ , for  $x \in \mathbb{R}_+$ . We have  $g_P \in \mathcal{R}_{m,+}(\mathbb{R}_+)$ , so  $\langle f_{\boldsymbol{\mu}}, g_P \rangle \geq 0$ . Moreover, with Eq. (4.34)

$$\begin{aligned} \langle f_{\boldsymbol{\mu}}, g_P \rangle &= \sum_{k=0}^m \mu_k \left( \sum_{l=k}^m p_l \binom{l}{k} l! \right) \\ &= \sum_{l=0}^m \left( \sum_{k=0}^l \mu_k \binom{l}{k} l! \right) p_l \\ &= L_{\boldsymbol{\nu}}(P) \end{aligned} \quad (4.35)$$

where we used Eq. (4.33) and the definition of the Riesz functional from Definition 1.14 in the last line. In particular,  $L_{\boldsymbol{\nu}}(P) \geq 0$ , and this holds for all nonnegative polynomials  $P$  over  $\mathbb{R}_+$ . By the Riesz–Haviland theorem (Theorem 1.3), this implies that  $\boldsymbol{\nu}$  is the sequence of moments of a nonnegative distribution  $\nu$  supported on  $\mathbb{R}_+$ .

Furthermore, we have that for all  $k \in \mathbb{N}$ :

$$\begin{aligned} \mu_k &= \sum_{l=0}^k \nu_l \frac{(-1)^{k+l}}{l!} \binom{k}{l} \\ &= \sum_{l=0}^k \frac{(-1)^{k+l}}{l!} \binom{k}{l} \int_{\mathbb{R}_+} x^l d\nu(x) \\ &= \int_{\mathbb{R}_+} (-1)^k \sum_{l=0}^k \frac{(-1)^l}{l!} \binom{k}{l} x^l d\nu(x) \\ &= \int_{\mathbb{R}_+} (-1)^k L_k(x) d\nu(x) \\ &= \int_{\mathbb{R}_+} \mathcal{L}_k(x) e^{\frac{x}{2}} d\nu(x) \end{aligned} \quad (4.36)$$

where we used Lemma 4.2 in the first line. Hence,  $\boldsymbol{\mu}$  is the sequence of Laguerre moments of the distribution  $\mu(x) := e^{\frac{x}{2}} \nu(x)$  supported on  $\mathbb{R}_+$ , which is nonnegative since  $\nu$  is nonnegative.  $\square$

### 4.3.3 Computing the threshold value as a linear program

In this section, we phrase the computation of the witness threshold value introduced in Eq. (4.3) as an infinite-dimensional linear program, in the case where one entry of the vector  $\mathbf{a}$  is equal to 1 and all the other entries are 0, the generalisation being straightforward by linearity.

Formally, we fix hereafter  $n \in \mathbb{N}^*$  and we look for the witnesses threshold value  $\omega_n$ <sup>4</sup> defined as

$$\omega_n := \sup_{\substack{\rho \in \mathcal{D}(\mathcal{H}) \\ W_\rho \geq 0}} \langle n | \rho | n \rangle. \quad (4.37)$$

This is the maximal values such that for all states  $\rho \in \mathcal{D}(\mathcal{H})$ :

$$\langle n | \rho | n \rangle > \omega_n \quad \Rightarrow \quad \exists \alpha \in \mathbb{C}, W_\rho(\alpha) < 0. \quad (4.38)$$

Let  $\mathcal{C}(\mathcal{H})$  be the set of states that are invariant under phase-space rotations:

$$\mathcal{C}(\mathcal{H}) := \{\sigma \in \mathcal{D}(\mathcal{H}) : \forall \varphi \in [0, 2\pi], e^{i\varphi \hat{n}} \sigma e^{-i\varphi \hat{n}} = \sigma\}, \quad (4.39)$$

where  $\hat{n} = \hat{a}^\dagger \hat{a}$  is the number operator. The witnesses corresponding to the fidelity with a single Fock state (and linear combination of Fock states) feature a rotational symmetry in phase space, which we exploit in the following lemma.

**Lemma 4.4.** *The threshold value in Eq. (4.37) can be expressed as*

$$\omega_n = \sup_{\substack{\sigma \in \mathcal{C}(\mathcal{H}) \\ W_\sigma \geq 0}} \langle n | \sigma | n \rangle. \quad (4.40)$$

*Proof.* Let  $\rho \in \mathcal{D}(\mathcal{H})$ . We start by applying a random dephasing to the state  $\rho$ :

$$\sigma = \int_0^{2\pi} \frac{d\varphi}{2\pi} e^{i\varphi \hat{n}} \rho e^{-i\varphi \hat{n}} \in \mathcal{C}(\mathcal{H}). \quad (4.41)$$

The random dephasing does not change the fidelity with any Fock state because of the rotational symmetry in phase space of the latter (see Eq. (1.33)), that is

$$\forall n \in \mathbb{N}, \langle n | \sigma | n \rangle = \langle n | \rho | n \rangle. \quad (4.42)$$

Moreover, it can only decrease the maximum negativity of the Wigner function. Indeed for all  $\alpha \in \mathbb{C}$ ,

$$\begin{aligned} W_\sigma(\alpha) &= \int_0^{2\pi} \frac{d\varphi}{2\pi} W_{e^{i\varphi \hat{n}} \rho e^{-i\varphi \hat{n}}}(\alpha) \\ &= \int_0^{2\pi} \frac{d\varphi}{2\pi} W_\rho(\alpha e^{i\varphi}) \\ &\geq \min_{\varphi \in [0, 2\pi]} W_\rho(\alpha e^{i\varphi}) \\ &\geq \min_{\beta \in \mathbb{C}} W_\rho(\beta), \end{aligned} \quad (4.43)$$

<sup>4</sup>Here we write generically the threshold value as  $\omega_n$  while we use the more precise notation  $\omega_n^{L^2}$  (resp.  $\omega_n^S$ ) to refer to its computation in the space of square integrable functions (resp. Schwartz functions).

and taking the minimum over all  $\alpha \in \mathbb{C}$  then gives

$$\min_{\alpha \in \mathbb{C}} W_\sigma(\alpha) \geq \min_{\beta \in \mathbb{C}} W_\rho(\beta). \quad (4.44)$$

In particular, applying a random dephasing to a Wigner positive state yields a Wigner positive mixtures of Fock states, which is invariant under phase-space rotations. Hence, we can restrict without loss of generality to states that are invariant under phase-space rotations when looking for the maximum fidelity of Wigner positive states with a given Fock state  $|n\rangle$ .  $\square$

Lemma 4.4 ensures that the supremum in Eq. (4.37) can be computed over states that have a rotational symmetry in phase space. Such states  $\sigma$  can be expanded diagonally in the Fock basis:

$$\sigma = \sum_{k=0}^{\infty} F_k |k\rangle\langle k|, \quad (4.45)$$

with the normalisation condition  $\sum_k F_k = 1$  and  $0 \leq F_k \leq 1$  for all  $k \in \mathbb{N}$  assuring that  $\sigma$  is a positive semidefinite operator. By linearity of the Wigner function:

$$\forall \alpha \in \mathbb{C}, W_\sigma(\alpha) = \sum_k F_k W_k(\alpha), \quad (4.46)$$

where  $W_k$  is the Wigner function of the  $k^{\text{th}}$  Fock state [KŽ04]:

$$\forall \alpha \in \mathbb{C}, W_k(\alpha) = \frac{2}{\pi} \mathcal{L}_k(4|\alpha|^2), \quad (4.47)$$

with  $\mathcal{L}_k$  the  $k^{\text{th}}$  Laguerre function, defined in Eq. (4.17). As noted before, Fock states are invariant under phase-space rotations: their Wigner function only depends on the amplitude of the phase-space point considered. We fix  $x = 4|\alpha|^2 \in \mathbb{R}^+$  hereafter.

### Computation of the threshold value over square-integrable functions.

We first consider the computation of the threshold value in Eq. (4.40) over  $L^2(\mathbb{R}_+)$  functions and we will denote the corresponding threshold value by  $\omega_n^{L^2}$ . With Lemma 4.4, the computation of  $\omega_n^{L^2}$  can thus be expressed as the following infinite-dimensional linear program:

$$(\text{LP}_n^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } (F_k)_{k \in \mathbb{N}} \in \ell^2(\mathbb{N}) \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_k F_k = 1 \\ \forall k \in \mathbb{N}, F_k \geq 0 \\ \forall x \in \mathbb{R}_+, \sum_k F_k \mathcal{L}_k(x) \geq 0. \end{array} \right.$$

The first constraint ensures unit trace of the corresponding state  $\sigma$ , the second one ensures that its fidelity with each Fock state is nonnegative and thus that the state is a proper positive semidefinite operator, and the last one ensures that its Wigner function  $W_\sigma$  is nonnegative. Note that  $\omega_n > 0$  for all  $n \in \mathbb{N}^*$ , by considering a mixture of  $|0\rangle$  and  $|n\rangle$  with the vacuum component close enough to 1. For convenience when considering the dual formulation, we use the superscript  $L$  instead of  $\ell$  in the label of the program. Even though it is formulated as an optimisation over sequences from  $\ell^2$ , we can equivalently see it as an optimisation over  $L^2(\mathbb{R}_+)$  functions as  $L^2(\mathbb{R}_+)$  and  $\ell^2$  are isomorphic.

The dual linear program reads:

$$(D\text{-LP}_n^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } \mu \in L^{2'}(\mathbb{R}_+) \\ \text{minimising } y \\ \text{subject to} \\ \forall k \neq n \in \mathbb{N}, y \geq \int_{\mathbb{R}_+} \mathcal{L}_k d\mu \\ y \geq 1 + \int_{\mathbb{R}_+} \mathcal{L}_n d\mu \\ \forall f \in L_+^2(\mathbb{R}_+), \langle \mu, f \rangle \geq 0. \end{array} \right.$$

### Retrieving the canonical form of infinite-dimensional LP

Here we follow Subsection 1.4.1 to retrieve the standard form of infinite-dimensional linear program as presented in [Bar02, IV–(6.1)] and see why programs  $(\text{LP}_n^{L^2})$  and  $(D\text{-LP}_n^{L^2})$  are indeed dual programs.

Recall that via expansion on a basis of  $L^2(\mathbb{R}_+)$ ,  $\ell^2(\mathbb{N})$  and  $L^2(\mathbb{R}_+)$  are isomorphic and that the spaces  $L^2(\mathbb{R}_+)$  and  $L^{2'}(\mathbb{R}_+)$  are isomorphic by the Radon–Nikodym theorem. Let us introduce the spaces:

- $E_1 = \ell^2(\mathbb{N}) \times L^2(\mathbb{R}_+)$ .
- $F_1 = \ell^2(\mathbb{N}) \times L^{2'}(\mathbb{R}_+)$  the dual space of  $E_1$ .
- $E_2 = \mathbb{R} \times L^2(\mathbb{R}_+)$ .
- $F_2 = \mathbb{R} \times L^{2'}(\mathbb{R}_+)$  the dual space of  $E_2$ .

We also define the dualities  $\langle -, - \rangle_1 : E_1 \times F_1 \rightarrow \mathbb{R}$  and  $\langle -, - \rangle_2 : E_2 \times F_2 \rightarrow \mathbb{R}$  as follows:

$$\forall e_1 = ((u_k), f) \in E_1, \forall f_1 = ((v_k), \mu) \in F_1, \quad \langle e_1, f_1 \rangle_1 := \sum_k u_k v_k + \int_{\mathbb{R}_+} f d\mu, \quad (4.48)$$

$$\forall e_2 = (x, f) \in E_2, \forall f_2 = (y, \mu) \in F_2, \quad \langle e_2, f_2 \rangle_2 := xy + \int_{\mathbb{R}_+} f d\mu. \quad (4.49)$$

Let  $A: E_1 \rightarrow E_2$  be the following linear transformation:

$$\forall e_1 = ((u_k), f) \in E_1, \quad A(e_1) := \left( \sum_k u_k, x \in \mathbb{R}_+ \mapsto f(x) - \sum_k u_k \mathcal{L}_k(x) \right), \quad (4.50)$$

and  $A^* : F_2 \rightarrow F_1$  be defined as:

$$\forall f_2 = (y, \mu) \in F_2, \quad A^*(f_2) := \left( \left( y - \int_{\mathbb{R}_+} \mathcal{L}_k d\mu \right)_{k \in \mathbb{N}}, \mu \right). \quad (4.51)$$

We can easily verify that  $A^*$  is the dual transformation of  $A$ , i.e.,  $\forall e_1 \in E_1, \forall f_2 \in F_2$  we have  $\langle A(e_1), f_2 \rangle_2 = \langle e_1, A^*(f_2) \rangle_1$ .

Recall that  $L_+^2(\mathbb{R}_+)$  is the cone of nonnegative functions in  $L^2(\mathbb{R}_+)$  and  $\ell_+^2$  the cone of sequences in  $\ell^2$  with nonnegative coefficients. We will optimise in the convex cones  $K_1 = \ell_+^2 \times L_+^2(\mathbb{R}_+) \subset E_1$  and  $K_2 = \{0\}$ . The dual cones are then respectively:  $K_1^* = \{f_1 \in F_1 : \forall e_1 \in K_1, \langle e_1, f_1 \rangle \geq 0\}$  and  $K_2^* = F_2$ .

We can now rewrite the problem  $(\text{LP}_n^{L^2})$  as a standard linear program in convex cones. We choose the vector function in the objective to be  $c_n = ((\delta_{kn})_k, \mathbf{0}) \in F_1$  and we also set  $b = (1, \mathbf{0}) \in E_2$  for the constraints. The standard form of  $(\text{LP}_n^{L^2})$  in the sense of [Bar02] can be written as follows:

$$(\text{LP}_n^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } e_1 \in E_1 \\ \text{maximising } \langle e_1, c_n \rangle_1 \\ \text{subject to:} \\ \quad A(e_1) = b \\ \quad e_1 \geq_{K_1} 0. \end{array} \right.$$

This is expressed with an equality constraint with a slack variable rather than an inequality constraint (see the canonical program in 1.4.1). Both formulations are equivalent. The standard form of the dual can be expressed as follows:

$$(\text{D-LP}_n^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } f_2 \in F_2 \\ \text{minimising } \langle b, f_2 \rangle_2 \\ \text{subject to:} \\ \quad A^*(f_2) \geq_{K_1^*} c_n. \end{array} \right.$$

which can be expanded as program  $(\text{D-LP}_n^{L^2})$ . Note the same derivation holds if we optimise over Schwartz functions by considering  $E_1 = \mathcal{S}(\mathbb{N}) \times \mathcal{S}(\mathbb{R}_+)$ ,  $E_2 = \mathcal{S}'(\mathbb{N}) \times \mathcal{S}'(\mathbb{R}_+)$ ,  $F_1 = \mathbb{R} \times \mathcal{S}(\mathbb{R}_+)$  and  $F_2 = \mathbb{R} \times \mathcal{S}'(\mathbb{R}_+)$ .

### Strong duality

This will follow immediately from the proof of convergence we provide later. As a sanity check, we use standard tools for proving strong duality as presented in 1.4.1 with Theorem 1.1.

There is no duality gap between programs  $(\text{LP}_n^{L^2})$  and  $(\text{D-LP}_n^{L^2})$  if there is a primal feasible plan and if the convex cone

$$\begin{aligned} \mathcal{K} &= \left\{ \left( A(e_1), \langle e_1, c \rangle_1 \right) : e_1 \in K_1 \right\} \\ &= \left\{ \left( \sum_k u_k, x \in \mathbb{R}_+ \mapsto f(x) - \sum_k u_k \mathcal{L}_k(x), F_n \right) : ((u_k), f) \in K_1 \right\} \end{aligned} \quad (4.52)$$

is closed in  $E_2 \oplus \mathbb{R}$  (for the weak topology).

*Proof.* The null sequence and the null function provides a feasible plan for the primal problem.

Next, we consider a sequence  $(e_{1j})_j = (((u_k^j)_k)_j, (f_j)_j) \in K_1^{\mathbb{N}} = \ell^2(\mathbb{N}) \times L_+^2(\mathbb{R}_+)^{\mathbb{N}}$  and we want to show that the accumulation point  $(b, g, a) = \lim_{j \rightarrow \infty} (A(e_{1j}), \langle e_{1j}, c \rangle_1)$  belongs to  $\mathcal{K}$  where  $a, b \in \mathbb{R}$  and  $g \in L^2(\mathbb{R}_+)$ .

For all  $j \in \mathbb{N}$ ,  $(u_k^j)_k \in \ell^2$  and for all  $k \in \mathbb{N}$ ,  $u_k^j$  is bounded. Thus, for all  $k \in \mathbb{N}$ , the sequence  $(u_k^j)_j$  is bounded and via diagonal extraction there exists  $\phi : \mathbb{N} \rightarrow \mathbb{N}$  strictly increasing such that  $(u_k^j)_{\phi(j)}$  converges. We denote  $\tilde{u}_k$  its limit. Since  $\ell^2(\mathbb{N})$  is closed, the sequence  $(\tilde{u}_k)_k$  belongs to  $\ell^2(\mathbb{N})$  and we have  $b = \sum_k \tilde{u}_k$  and  $a = \tilde{u}_n$ .

Now  $f_j - \sum_k u_k^j \mathcal{L}_k \rightarrow g$  so that  $f_j \rightarrow g + \sum_k \tilde{u}_k \mathcal{L}_k \in L_+^2(\mathbb{R}_+)$  since  $L_+^2(\mathbb{R}_+)$  is closed. Thus, for  $\tilde{e}_1 = ((\tilde{u}_k)_k, g + \sum_k \tilde{u}_k \mathcal{L}_k) \in K_1$ ,  $(b, g, a) = (A(\tilde{e}_1), \langle \tilde{e}_1, c \rangle_1)$  and  $(b, g, a) \in \mathcal{K}$ .  $\square$

Note that the same proof does not hold for the problems expressed over Schwartz functions as  $\mathcal{S}(\mathbb{N})$  and  $\mathcal{S}_+(\mathbb{R}_+)$  are not closed. We will rely on the proof of convergence of the lower bounding hierarchy of semidefinite programs to prove strong duality in this case.

### Computation of the threshold over Schwartz functions

We also consider a restriction of  $(\text{LP}_n^{L^2})$  by optimising over Schwartz functions rather than square integrable functions as  $\mathcal{S}(\mathbb{R}_+) \subset L^2(\mathbb{R}_+)$ . The primal program becomes:

$$(\text{LP}_n^{\mathcal{S}}) \quad \left\{ \begin{array}{l} \text{Find } (F_k)_{k \in \mathbb{N}} \in \mathcal{S}(\mathbb{N}) \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_k F_k = 1 \\ \forall k \in \mathbb{N}, F_k \geq 0 \\ \forall x \in \mathbb{R}_+, \sum_k F_k \mathcal{L}_k(x) \geq 0, \end{array} \right.$$

<sup>5</sup>Because we are dealing with a sequence of sequences, we use the upper index to refer to the embracing sequence.

We denote its value by  $\omega_n^S$ . Because it is a restriction of  $(\text{LP}_n^{L^2})$ , we have that  $\omega_n^{L^2} \geq \omega_n^S$ . Its dual linear program can be expressed as:

$$(\text{D-LP}_n^S) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } \mu \in \mathcal{S}'(\mathbb{R}_+) \\ \text{minimising } y \\ \text{subject to} \\ \forall k \neq n \in \mathbb{N}, y \geq \int_{\mathbb{R}_+} \mathcal{L}_k d\mu \\ y \geq 1 + \int_{\mathbb{R}_+} \mathcal{L}_n d\mu \\ \forall f \in \mathcal{S}_+(\mathbb{R}_+), \langle \mu, f \rangle \geq 0. \end{array} \right.$$

Note there is a rather strong structure with Wigner functions. In [HR21] it is shown that the Wigner function only need to decay towards infinity faster than any polynomial to be a Schwartz function (i.e. it also implies that all its derivatives decay towards infinity faster than any polynomial).

### Analytical solutions

Even without strong duality (it is not proven in the case of Schwartz functions), weak duality of linear programming already ensures that the optimal value  $\omega_n^{L^2}$  of  $(\text{LP}_n^{L^2})$  (resp.  $\omega_n^S$  of  $(\text{LP}_n^S)$ ) is upper bounded by the optimal value of  $(\text{D-LP}_n^{L^2})$  (resp.  $(\text{D-LP}_n^S)$ ). Hence, a possible way of solving the optimisation  $(\text{LP}_n^{L^2})$  is to exhibit a feasible solution for  $(\text{LP}_n^{L^2})$  and a feasible solution for  $(\text{D-LP}_n^{L^2})$  that have the same value.

For  $n = 1$ , choosing  $(F_k)_{k \in \mathbb{N}} = (\frac{1}{2}, \frac{1}{2}, 0, 0, \dots)$  gives a feasible solution for  $(\text{LP}_1^{L^2})$  (resp.  $(\text{LP}_1^S)$ ) with the value  $\frac{1}{2}$ , while choosing  $(y, \mu) = (\frac{1}{2}, \frac{1}{2}\delta(x))$ , where  $\delta$  is the Dirac delta function<sup>6</sup> over  $\mathbb{R}_+$ , gives a feasible solution for  $(\text{D-LP}_1^{L^2})$  (resp.  $(\text{D-LP}_1^S)$ ) with the value  $\frac{1}{2}$ . This shows that  $\omega_1^{L^2} = \omega_1^S = \frac{1}{2}$ .

Similarly, for  $n = 2$ , choosing  $(F_k)_{k \in \mathbb{N}} = (\frac{1}{2}, 0, \frac{1}{2}, 0, 0, \dots)$  gives a feasible solution for  $(\text{LP}_2^{L^2})$  (resp.  $\text{LP}_2^S$ ) with the value  $\frac{1}{2}$ , while choosing  $(y, \mu) = (\frac{1}{2}, \frac{e}{2}\delta(x-2))$  gives a feasible solution for  $(\text{D-LP}_2)$  (resp.  $(\text{D-LP}_2^S)$ ), up to a conjecture<sup>7</sup>, with the value  $\frac{1}{2}$ . This shows that  $\omega_2^{L^2} = \omega_2^S = \frac{1}{2}$ .

While this approach is sensible for small values of  $n$ , finding optimal analytical solutions for higher values of  $n$  seems highly nontrivial. Moreover, the infinite number of variables prevents us from performing the optimisation  $(\text{LP}_n^{L^2})$  numerically. A natural workaround is to find finite-dimensional relaxations or restrictions of the original problem—thus providing upper and lower bounds for the optimal value  $\omega_n^{L^2}$ , respectively. This is the approach we follow in the next section.

<sup>6</sup>Technically  $\delta \notin L^2(\mathbb{R}_+)$ , but the result holds by considering a sequence of functions converging to a Dirac delta.

<sup>7</sup>We checked numerically the corresponding constraints  $|L_k(2)| \leq 1$  for  $k$  up to  $10^3$  and, considering asymptotic behaviors, we conjecture that these hold for all  $k \geq 0$ .

### 4.3.4 Hierarchies of semidefinite programs

Semidefinite programming is a convex optimisation technique in the cone of positive semidefinite matrices. For more details, see a brief review in Subsection 1.4.2.

#### 4.3.4.1 Preliminaries

In this section, we introduce preliminary technical lemmas.

We recall the following standard result, which comes from the fact that any univariate polynomial nonnegative over  $\mathbb{R}$  can be written as a sum-of-squares:

**Lemma 4.5** ([Hil88]). *Let  $p \in \mathbb{N}$  and let  $P$  be a univariate polynomial of degree  $2p$ . Let  $\mathbf{v}(x) = (1, x, \dots, x^p)$  be the vector of monomials. Then,  $P$  is nonnegative over  $\mathbb{R}$  if and only if there exists a real  $(p+1) \times (p+1)$  positive semidefinite matrix  $Q$  such that for all  $x \in \mathbb{R}$ ,*

$$P(x) = \mathbf{v}(x)^T Q \mathbf{v}(x). \quad (4.53)$$

This is the univariate case of Proposition 1.1. From this lemma we deduce the following characterisation of nonnegative polynomials over  $\mathbb{R}_+$ :

**Lemma 4.6.** *nonnegative polynomials on  $\mathbb{R}_+$  can be written as sums of polynomials of the form  $\sum_{l=0}^p x^l \sum_{i+j=2l} y_i y_j$ , where  $p \in \mathbb{N}$  and  $y_i \in \mathbb{R}$ , for all  $0 \leq i \leq p$ .*

*Proof.* Let  $P$  be a univariate polynomial of degree  $p$  which is nonnegative on  $\mathbb{R}_+$ . Writing  $\mathbf{v}(x) = (1, x, \dots, x^p)$ , the polynomial  $x \mapsto P(x^2)$  of degree  $2p$  is nonnegative on  $\mathbb{R}$ , so by Lemma 4.5 there exists a real positive semidefinite matrix  $Q = (Q_{ij})_{0 \leq i, j \leq p}$  such that for all  $x \in \mathbb{R}$ :

$$\begin{aligned} P(x^2) &= \mathbf{v}(x)^T Q \mathbf{v}(x) \\ &= \sum_{k=0}^{2p} x^k \sum_{i+j=k} Q_{ij} \\ &= \sum_{l=0}^p x^{2l} \sum_{i+j=2l} Q_{ij}, \end{aligned} \quad (4.54)$$

where the last line comes from the fact that  $x \mapsto P(x^2)$  has no monomial of odd degree. Hence, for all  $x \in \mathbb{R}_+$ ,

$$P(x) = \sum_{l=0}^p x^l \sum_{i+j=2l} Q_{ij}. \quad (4.55)$$

$Q$  is a real  $(p+1) \times (p+1)$  positive semidefinite matrix, so via Cholesky decomposition

$$Q = \sum_{k=0}^p \mathbf{y}^{(k)} \mathbf{y}^{(k)T}, \quad (4.56)$$

where  $\mathbf{y}^{(k)} \in \mathbb{R}^{p+1}$  for all  $k \in \llbracket 0, p \rrbracket$ . We finally obtain, for all  $x \in \mathbb{R}_+$ ,

$$\begin{aligned} P(x) &= \sum_{l=0}^p x^l \sum_{i+j=2l} \sum_{k=0}^p \left( \mathbf{y}^{(k)} \mathbf{y}^{(k)T} \right)_{ij} \\ &= \sum_{k=0}^p \left( \sum_{l=0}^p x^l \sum_{i+j=2l} \mathbf{y}_i^{(k)} \mathbf{y}_j^{(k)} \right). \end{aligned} \tag{4.57}$$

□

Note that the characterisation in Lemma 4.6 differs from that of Stieltjes [RS75], which expresses nonnegative polynomials over  $\mathbb{R}_+$  as  $x \mapsto A_1(x) + xA_2(x)$ , where  $A_1$  and  $A_2$  are sums of squares. This slightly slows down the numerical resolution, which does not matter given the size of the programs considered. At the same time, this allows us to obtain more compact expressions for the semidefinite programs.

We use the characterisation of Lemma 4.6 to obtain the following crucial result: for  $\mathbf{s} \in \mathbb{R}^{\mathbb{N}}$ , the fact that the series  $f_{\mathbf{s}}$  defined in Eq. (4.19) has nonnegative scalar product with nonnegative truncated Laguerre series up to degree  $m$  can be expressed as a positive semidefinite constraint involving the matrix  $A_{\mathbf{s}}$  defined in Eq. (4.21). Formally:

**Lemma 4.7.** *Let  $m \geq n$  and let  $\mathbf{s} \in \mathbb{R}^{\mathbb{N}}$ . The following propositions are equivalent:*

- (i)  $\forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_{\mathbf{s}}, g \rangle \geq 0,$
- (ii)  $A_{\mathbf{s}} \succeq 0.$

*Proof.* By Lemma 4.6, any nonnegative polynomial over  $\mathbb{R}_+$  of degree less or equal to  $m$  can be expressed as a sum of polynomials of the form  $\sum_{l=0}^m x^l \sum_{i+j=2l} y_i y_j$ . We set the row vector  $Y = (y_0, \dots, y_m) \in \mathbb{R}^{m+1}$ . Hence, any nonnegative truncated Laguerre series (the elements of  $\mathcal{R}_{m,+}(\mathbb{R}_+)$ ) can be expressed as a sum of terms of the form  $e^{-\frac{x}{2}} \sum_{l=0}^m x^l \sum_{i+j=2l} y_i y_j$ . By linearity, it is sufficient to check that the scalar products with one of these terms are nonnegative.

Recall that for all  $k \in \mathbb{N}$  and  $\mathbf{s} = (s_k)_k \in \mathbb{R}^{\mathbb{N}}$  (see Eq. (4.20)) we have

$$s_k = \int_{\mathbb{R}_+} \mathcal{L}_k(x) f_{\mathbf{s}}(x) dx. \tag{4.58}$$

Thus,

$$\begin{aligned}
 A_s \succeq 0 &\Leftrightarrow \forall Y \in \mathbb{R}^{m+1}, Y^T A_s Y \geq 0 \\
 &\Leftrightarrow \forall Y \in \mathbb{R}^{m+1}, \sum_{i,j=0}^m y_i y_j (A_s)_{ij} \geq 0 \\
 &\Leftrightarrow \forall Y \in \mathbb{R}^{m+1}, \sum_{l=0}^m \sum_{i+j=2l}^m y_i y_j \sum_{k=0}^l s_k \binom{l}{k} l! \geq 0 \\
 &\Leftrightarrow \forall Y \in \mathbb{R}^{m+1}, \int_{\mathbb{R}_+} \sum_{l=0}^m \sum_{i+j=2l}^m y_i y_j \sum_{k=0}^l \binom{l}{k} l! \mathcal{L}_k(x) f_s(x) dx \geq 0 \\
 &\Leftrightarrow \forall Y \in \mathbb{R}^{m+1}, \int_{\mathbb{R}_+} \sum_{l=0}^m \sum_{i+j=2l}^m y_i y_j \sum_{k=0}^l (-1)^k \binom{l}{k} l! L_k(x) e^{-\frac{x}{2}} f_s(x) dx \geq 0 \\
 &\Leftrightarrow \forall Y \in \mathbb{R}^{m+1}, \int_{\mathbb{R}_+} \left( e^{-\frac{x}{2}} \sum_{l=0}^m x^l \sum_{i+j=2l}^m y_i y_j \right) f_s(x) dx \geq 0 \\
 &\Leftrightarrow \forall Y \in \mathbb{R}^{m+1}, \left\langle f_s, x \mapsto e^{-\frac{x}{2}} \sum_{l=0}^m x^l \sum_{i+j=2l}^m y_i y_j \right\rangle \geq 0 \\
 &\Leftrightarrow \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_s, g \rangle \geq 0,
 \end{aligned} \tag{4.59}$$

where we used Eq. (4.21) in the third line, Eq. (4.58) in the fourth line and Eq. (4.22) in the sixth line.  $\square$

Using these results, we derive hierarchies of semidefinite relaxations and restrictions for the infinite-dimensional linear program  $(LP_n^{L^2})$  in the following sections.

#### 4.3.4.2 Semidefinite restrictions for computing the threshold value

A trivial way to obtain a restriction of  $(LP_n^{L^2})$  (or equivalently of  $(LP_n^S)$ ) is to impose  $F_l = 0$  for  $l > m$ , for some  $m \geq n$ . What is less trivial is that this yields a finite-dimensional semidefinite program. Indeed, the constraint (4.76) becomes

$$\forall x \in \mathbb{R}_+, \sum_{k=0}^m F_k \mathcal{L}_k(x) \geq 0, \tag{4.60}$$

or equivalently:

$$\forall x \in \mathbb{R}, \sum_{k=0}^m (-1)^k F_k L_k(x^2) \geq 0, \tag{4.61}$$

where we used Eq. (4.17). Then, by Theorem 4.1, instead of imposing Eq. (4.61), one may equivalently require that:

$$\forall m \in \mathbb{N}, \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_F, g \rangle \geq 0, \tag{4.62}$$

The restriction at rank  $m \geq n$  can be written as:

$$(\text{LP}_n^{\text{restr},m}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{F} = (F_k)_{0 \leq k \leq m} \in \mathbb{R}^{m+1} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{k=0}^m F_k = 1 \\ \forall k \leq m, F_k \geq 0 \\ \forall p \in \mathbb{N}, \forall g \in \mathcal{R}_{p,+}(\mathbb{R}_+), \langle f_{\mathbf{F}}, g \rangle \geq 0. \end{array} \right.$$

We now express this program more straightforwardly as a semidefinite program. By Lemma 4.5, writing  $\mathbf{v}(x) = (1, x, \dots, x^m)$ , Eq. (4.61) is equivalent to the existence of a positive semidefinite matrix  $Q = (Q_{ij})_{0 \leq i, j \leq m}$  such that for all  $x \in \mathbb{R}$ ,

$$\sum_{k=0}^m (-1)^k F_k L_k(x^2) = \mathbf{v}(x)^T Q \mathbf{v}(x) = \sum_{l=0}^m \left( \sum_{i+j=l} Q_{ij} \right) x^l. \quad (4.63)$$

This is in turn equivalent to the linear constraints:

$$\left\{ \begin{array}{l} \forall l \in \llbracket 1, m \rrbracket, \sum_{i+j=2l-1} Q_{ij} = 0, \\ \forall l \leq m, \sum_{i+j=2l} Q_{ij} = \frac{(-1)^l}{l!} \sum_{k=l}^m (-1)^k \binom{k}{l} F_k, \end{array} \right. \quad (4.64)$$

by identifying the coefficients in front of each monomial in Eq. (4.63). Hence, the restriction of  $(\text{LP}_n^{L^2})$  obtained by imposing  $F_l = 0$  for  $l > m$ , for a fixed  $m \geq n$ , is a semidefinite program given by:

$$(\text{SDP}_n^{m, \leq}) \quad \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{m+1} \text{ and } \mathbf{F} \in \mathbb{R}^{m+1} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{k=0}^m F_k = 1 \\ \forall k \leq m, F_k \geq 0 \\ \forall l \in \llbracket 1, m \rrbracket, \sum_{i+j=2l-1} Q_{ij} = 0 \\ \forall l \leq m, \sum_{i+j=2l} Q_{ij} = \sum_{k=l}^m \frac{(-1)^{k+l}}{l!} \binom{k}{l} F_k \\ Q \succeq 0. \end{array} \right.$$

Let us denote its optimal value by  $\omega_n^{m, \leq}$ . Each choice of  $m$  leads to a different semidefinite restriction of  $(\text{LP}_n^{L^2})$  and  $(\text{LP}_n^S)$ , whose optimal value gets closer to  $\omega_n^S$  as  $m$  increases (since

a feasible solution at rank  $m$  necessary provides a feasible solution for the program at rank  $m + 1$ ). The sequence  $(\omega_n^{m,\leq})_{m \geq n}$  is thus an increasing sequence and for all  $m \geq n$ , we have  $\omega_n^{m,\leq} \leq \omega_n^S \leq \omega_n^{L^2}$ .

For each  $m \geq n$ , the program  $(\text{SDP}_n^{m,\leq})$  has a dual semidefinite program which is given by:

$$(\text{D-SDP}_n^{m,\leq}) \quad \left\{ \begin{array}{l} \text{Find } A \in \text{Sym}_{m+1}, \boldsymbol{\mu} \in \mathbb{R}^{m+1} \text{ and } y \in \mathbb{R} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_n \\ \forall k \leq m, y \geq \mu_k \\ \forall l \leq m, \forall i + j = 2l, A_{ij} = \sum_{k=0}^l \mu_k \binom{l}{k} l! \\ A \succeq 0. \end{array} \right.$$

### Retrieving the canonical form of SDP

Here we write those SDP in the standard forms (as detailed in Subsection 1.4.2) to see why they are indeed dual programs. In the programs (SDP) and (D-SDP) we can exchange *minimisation* and *maximisation* if we also change the sign of the semidefinite constraint for the dual program. We recall them below with this slight modification. For  $M, N \in \mathbb{N}$ ,  $\mathbf{b} = (b_1, \dots, b_M) \in \mathbb{R}^M$ ,  $C \in \text{Sym}_N$ , and  $B^{(i)} \in \text{Sym}_N$  for all  $i \in \llbracket 1, M \rrbracket$ :

$$(\text{SDP}) \quad \left\{ \begin{array}{l} \text{Find } X \in \text{Sym}_N \\ \text{maximising } \text{Tr}(C^T X) \\ \text{subject to:} \\ \forall i \in \llbracket 1, M \rrbracket, \quad \text{Tr}(B^{(i)} X) = b_i \\ X \succeq 0. \end{array} \right.$$

$$(\text{D-SDP}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} \in \mathbb{R}^M \\ \text{minimising } \mathbf{b}^T \mathbf{y} \\ \text{subject to:} \\ \sum_{i=1}^M y_i B^{(i)} \succeq C. \end{array} \right.$$

To put  $(\text{SDP}_n^{m,\leq})$  in the standard form (SDP) we set  $N = 2 \times (m+1)$  and  $M = 1 + m + (m+1)$ . For all  $r \in \mathbb{N}^*$  and all  $i, j \in \llbracket 1, r \rrbracket$ , let  $E_r^{(i,j)}$  be the  $r \times r$  matrix whose  $(i, j)$  entry is 1 and all

other entries are 0. We set

$$\begin{aligned}
 X &= \text{Diag}_{k=0,\dots,m}(F_k) \oplus Q \in \text{Sym}_N, \\
 C &= E_N^{(n,n)} = E_{m+1}^{(n,n)} \oplus \mathbf{0}_{m+1} \in \text{Sym}_N, \\
 \mathbf{b} &= (1, 0, 0, \dots, 0) \in \mathbb{R}^M, \\
 B'^{(0)} &= \mathbf{1}_{m+1} \oplus \mathbf{0}_{m+1} \in \text{Sym}_N, \\
 \forall l \in \llbracket 1, m \rrbracket, \quad B'^{(l)} &= \mathbf{0}_{m+1} \oplus \left( \sum_{i+j=2l-1} E_{m+1}^{(i,j)} \right), \\
 \forall l \in \llbracket 0, m \rrbracket, \quad B^{(l)} &= \text{Diag}_{k=0,\dots,m} \left( -\frac{(-1)^{k+l}}{l!} \binom{k}{l} \right) \oplus \left( \sum_{i+j=2l} E_{m+1}^{(i,j)} \right),
 \end{aligned} \tag{4.65}$$

with the convention  $\binom{k}{l} = 0$  when  $l > k$ . The matrix  $B'^{(0)}$  corresponds to the constraint  $\sum_{k=0}^m F_k = 1$  and we denote the corresponding dual variable  $y \in \mathbb{R}$ . Similarly, the matrices  $B'^{(l)}$  correspond to the  $m$  constraints  $\sum_{i+j=2l-1} Q_{ij} = 0$ , and we denote the corresponding dual variables  $\nu'_l \in \mathbb{R}$ . Finally, the matrices  $B^{(l)}$  correspond to the  $m+1$  constraints  $\sum_{i+j=2l} Q_{ij} = \frac{(-1)^l}{l!} \sum_{k=l}^m (-1)^k \binom{k}{l} F_k$ , and we denote the corresponding dual variables  $\nu_l \in \mathbb{R}$ .

The standard form of the dual program (D-SDP $_n^{m,\leq}$ ) thus reads:

$$\text{(D-SDP}_n^{m,\leq}) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } \boldsymbol{\nu}, \boldsymbol{\nu}' \in \mathbb{R}^{m+1} \times \mathbb{R}^m \\ \text{minimising } y \\ \text{subject to:} \\ \text{Diag}_{k=0,\dots,m} \left[ y - \sum_{l=0}^k \nu_l \frac{(-1)^{k+l}}{l!} \binom{k}{l} \right] \\ \oplus \left( \sum_{l=0}^m \sum_{i+j=2l} \nu_l E_{m+1}^{(i,j)} + \sum_{l=1}^m \sum_{i+j=2l-1} \nu'_l E_{m+1}^{(i,j)} \right) \succeq E_{m+1}^{(n,n)} \oplus \mathbf{0}_{m+1}. \end{array} \right.$$

Due to the block-diagonal structure of the matrices involved, the positive semidefinite constraint above is equivalent to the following constraints:

$$\left\{ \begin{array}{l} y \geq 1 + \sum_{l=0}^n \nu_l \frac{(-1)^{n+l}}{l!} \binom{n}{l}, \\ \forall k \in \llbracket 0, m \rrbracket \setminus \{n\}, \quad y \geq \sum_{l=0}^k \nu_l \frac{(-1)^{k+l}}{l!} \binom{k}{l}, \\ \left( \sum_{l=0}^m \sum_{i+j=2l} \nu_l E_{m+1}^{(i,j)} + \sum_{l=1}^m \sum_{i+j=2l-1} \nu'_l E_{m+1}^{(i,j)} \right) \succeq 0. \end{array} \right. \tag{4.66}$$

Let us define  $A = (A_{ij})_{0 \leq i,j \leq m}$  by

$$A := \sum_{l=0}^m \sum_{i+j=2l} \nu_l E_{m+1}^{(i,j)} + \sum_{l=1}^m \sum_{i+j=2l-1} \nu'_l E_{m+1}^{(i,j)}, \tag{4.67}$$

or equivalently

$$A_{ij} = \begin{cases} \nu_l & \text{when } i + j = 2l, \\ \nu'_l & \text{when } i + j = 2l - 1, \end{cases} \quad (4.68)$$

for all  $i, j \in \llbracket 0, m \rrbracket$ . For  $k \in \llbracket 0, m \rrbracket$ , we also define

$$\mu_k := \sum_{l=0}^k \nu_l \frac{(-1)^{k+l}}{l!} \binom{k}{l} \in \mathbb{R}. \quad (4.69)$$

By Lemma 4.2, the following conditions are equivalent:

$$\begin{aligned} \text{(i)} \quad & \forall k \in \llbracket 0, m \rrbracket, \quad \mu_k = \sum_{l=0}^k \nu_l \frac{(-1)^{k+l}}{l!} \binom{k}{l}, \\ \text{(ii)} \quad & \forall l \in \llbracket 0, m \rrbracket, \quad \nu_l = \sum_{k=0}^l \mu_k \binom{l}{k} l!. \end{aligned}$$

With Eq. (4.68) we thus have

$$A_{ij} = \sum_{k=0}^l \mu_k \binom{l}{k} l! \quad \text{when } i + j = 2l, \quad (4.70)$$

and we obtain the following expected expression for (D-SDP $_n^{m,\leq}$ ):

$$\text{(D-SDP}_n^{m,\leq}) \quad \left\{ \begin{array}{l} \text{Find } y, \boldsymbol{\mu} \in \mathbb{R} \times \mathbb{R}^{m+1} \text{ and } A \in \text{Sym}_{m+1} \\ \text{minimising } y \\ \text{subject to:} \\ y \geq 1 + \mu_n \\ \forall k \leq m, \quad y \geq \mu_k \\ \forall l \leq m, \forall i + j = 2l, \quad A_{ij} = \sum_{k=0}^l \mu_k \binom{l}{k} l! \\ A \succeq 0. \end{array} \right.$$

Note that the constraint  $y \geq 1 + \mu_n$  implies the constraint  $y \geq \mu_n$ .

### Strong duality

We show that strong duality holds between the primal and the dual versions of this semidefinite program. In particular, numerical computations with either of these programs will yield the same value.

**Theorem 4.2.** *Strong duality holds between the programs (SDP $_n^{m,\leq}$ ) and (D-SDP $_n^{m,\leq}$ ).*

*Proof.* We make use of Slater's condition for (finite-dimensional) semidefinite programs: strict feasibility of  $(\text{SDP}_n^{m,\leq})$  implies strong duality between  $(\text{SDP}_n^{m,\leq})$  and  $(\text{D-SDP}_n^{m,\leq})$ .

In order to obtain a strictly feasible solution, we define  $Q \in \text{Sym}_{m+1}$  and  $\mathbf{F} = (F_0, \dots, F_m) \in \mathbb{R}^{m+1}$  by:

$$Q := \frac{1}{2^{m+1} - 1} \text{Diag}_{k=0, \dots, m} \left( \frac{1}{k!} \right) \quad (4.71)$$

$$F_k := \frac{1}{2^{m+1} - 1} \binom{m+1}{k+1} \quad (4.72)$$

Then  $Q \succ 0$  and  $F_k > 0$  for all  $k \in \llbracket 0, m \rrbracket$ . Moreover, we have

$$\begin{aligned} \sum_{k=0}^m F_k &= \frac{1}{2^{m+1} - 1} \sum_{k=0}^m \binom{m+1}{k+1} \\ &= \frac{1}{2^{m+1} - 1} \left( \sum_{k=0}^{m+1} \binom{m+1}{k} - 1 \right) \\ &= 1. \end{aligned} \quad (4.73)$$

We also have  $\sum_{i+j=2l-1} Q_{ij} = 0$  for all  $l \in \llbracket 1, m \rrbracket$  since  $Q$  is diagonal. Furthermore, for all  $l \leq m$ ,

$$\sum_{i+j=2l} Q_{ij} = Q_{ll} = \frac{1}{2^{m+1} - 1} \frac{1}{l!}, \quad (4.74)$$

and indeed

$$\begin{aligned} &\frac{(-1)^l}{l!} \sum_{k=l}^m (-1)^k \binom{k}{l} F_k \\ &= \frac{1}{2^{m+1} - 1} \frac{1}{l!} \sum_{k=l}^m (-1)^{k-l} \binom{k}{m} \binom{m+1}{k+1} \\ &= \frac{1}{2^{m+1} - 1} \frac{1}{l!} \binom{m}{l} \sum_{q=0}^{m-l} (-1)^q \frac{m+1}{q+l+1} \binom{m-l}{q} \\ &= \frac{1}{2^{m+1} - 1} \frac{1}{l!}, \end{aligned} \quad (4.75)$$

where we used [Gou72, (1.41)] in the last line. Therefore,  $(Q, \mathbf{F})$  is a strictly feasible solution of  $(\text{SDP}_n^{m,\leq})$ , which implies strong duality.  $\square$

#### 4.3.4.3 Semidefinite relaxations for computing the threshold value

One way to obtain a relaxation of  $(\text{LP}_n^{L^2})$  is to relax the constraint:

$$\forall x \in \mathbb{R}_+, f_{\mathbf{F}}(x) = \sum_k F_k \mathcal{L}_k(x) \geq 0. \quad (4.76)$$

Instead, one may impose the weaker constraint:

$$\forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_{\mathbf{F}}, g \rangle \geq 0, \quad (4.77)$$

for some fixed  $m \geq n$ . This relaxation is motivated by Theorem 4.1 as it indicates that we will indeed have convergence. We will prove this formally later. Moreover the constraint in Eq. (4.77) only concerns the coefficients  $F_k$  for  $0 \leq k \leq m$  and the other variables  $F_l$  for  $l > m$  are only constrained by  $\sum_k F_k = 1$  and  $F_l \geq 0$ . Since  $m \geq n$  and the objective is to maximise  $F_n$  we can set all of the coefficients  $F_l$  for  $l > m$  to 0 without loss of generality. The relaxation at rank  $m \geq n$  can thus be expressed as:

$$(\text{LP}_n^{\text{relax},m}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{F} = (F_k)_{0 \leq k \leq m} \in \mathbb{R}^{m+1} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{k=0}^m F_k = 1 \\ \forall k \leq m, F_k \geq 0 \\ \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_{\mathbf{F}}, g \rangle \geq 0. \end{array} \right.$$

It is interesting to relate this program to  $(\text{LP}_n^{\text{restr},m})$ . At first glance they seem identical. The only difference appears on the last constraint. In  $(\text{LP}_n^{\text{relax},m})$  we require that  $f_{\mathbf{F}}$  has a nonnegative inner product with all nonnegative polynomials over  $\mathbb{R}_+$  of degree less or equal than  $m$  multiplied by the function  $x \mapsto e^{-\frac{x}{2}}$ . In  $(\text{LP}_n^{\text{restr},m})$  we impose the more restrictive constraint that  $f_{\mathbf{F}}$  has a nonnegative inner product with all nonnegative square integrable functions  $g$ . Of course because  $f_{\mathbf{F}}$  has an expansion on the Laguerre basis of fixed degree  $m$ , only the first  $m + 1$  coefficients of  $g$  will contribute. This amounts to taking  $g \in \mathcal{R}_m(\mathbb{R}_+)$  such that  $g$  can be completed into a nonnegative  $L^2(\mathbb{R}_+)$  function.

By Lemma 4.7, the constraint in Eq. (4.77) may in turn be expressed as a positive semidefinite constraint on the  $(m + 1) \times (m + 1)$  matrix  $A_{\mathbf{F}}$  defined in Eq. (4.21). Each choice of  $m$  thus leads to a different semidefinite program, whose optimal value gets closer to  $\omega_n$  as  $m$  increases (since the constraint (4.77) gets stronger when  $m$  increases). This gives a hierarchy of finite-dimensional semidefinite relaxations for  $(\text{LP}_n^{L^2})$ . The semidefinite relaxation of order  $m$  is given by:

$$(\text{SDP}_n^{m,\geq}) \quad \left\{ \begin{array}{l} \text{Find } A \in \text{Sym}_{m+1} \text{ and } \mathbf{F} \in \mathbb{R}^{m+1} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{k=0}^m F_k = 1 \\ \forall k \leq m, F_k \geq 0 \\ \forall l \leq m, \forall i + j = 2l, A_{ij} = \sum_{k=0}^l F_k \binom{l}{k} l! \\ \forall l \in \llbracket 1, m \rrbracket, \forall i + j = 2l - 1, A_{ij} = 0 \\ A \succeq 0. \end{array} \right.$$

Let us denote its optimal value by  $\omega_n^{m,\geq}$ . The sequence  $(\omega_n^{m,\geq})_{m \geq n}$  is a decreasing sequence and for all  $m \geq n$ , we have  $\omega_n^{L^2} \leq \omega_n^{m,\geq}$ .

For each  $m \geq n$ , the program  $(\text{SDP}_n^{m,\geq})$  has a dual semidefinite program which is given by:

$$(\text{D-SDP}_n^{m,\geq}) \quad \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{m+1}, \boldsymbol{\mu} \in \mathbb{R}^{m+1} \text{ and } y \in \mathbb{R} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_n \\ \forall k \leq m, y \geq \mu_k \\ \forall l \leq m, \sum_{i+j=2l} Q_{ij} = \sum_{k=l}^m \frac{(-1)^{k+l}}{l!} \binom{k}{l} \mu_k \\ Q \succeq 0. \end{array} \right.$$

### Retrieving the canonical form of SDP

To put  $(\text{SDP}_n^{m,\geq})$  in the standard form (SDP) we set  $N = 2 \times (m + 1)$  and  $M = 1 + (m + 1)^2$ . For all  $r \in \mathbb{N}^*$  and all  $i, j \in \llbracket 1, r \rrbracket$ , recall that  $E_r^{(i,j)}$  denote the  $r \times r$  matrix whose  $(i, j)$  entry is 1 and all other entries are 0. We set

$$\begin{aligned} X &= \text{Diag}_{k=0,\dots,m}(F_k) \oplus A \in \text{Sym}_N, \\ C &= E_N^{(n,n)} = E_{m+1}^{(n,n)} \oplus \mathbf{0}_{m+1} \in \text{Sym}_N, \\ \mathbf{b} &= (1, 0, 0, \dots, 0) \in \mathbb{R}^M, \\ B^{(0)} &= \mathbf{1}_{m+1} \oplus \mathbf{0}_{m+1} \in \text{Sym}_N, \\ \forall i, j \in \llbracket 0, m \rrbracket, \quad B^{(i,j)} &= \begin{cases} \text{Diag}_{k=0,\dots,m} \left( -\binom{l}{k} l! \right) \oplus \left( \frac{1}{2} E_{m+1}^{(i,j)} + \frac{1}{2} E_{m+1}^{(j,i)} \right) & \text{when } i + j = 2l, \\ \mathbf{0}_{m+1} \oplus \left( \frac{1}{2} E_{m+1}^{(i,j)} + \frac{1}{2} E_{m+1}^{(j,i)} \right) & \text{otherwise,} \end{cases} \end{aligned} \tag{4.78}$$

with the convention  $\binom{l}{k} = 0$  when  $k > l$ . The matrix  $B^{(0)}$  corresponds to the constraint  $\sum_{k=0}^m F_k = 1$ , and we denote the corresponding dual variable  $y \in \mathbb{R}$ . Similarly, the matrices  $B^{(i,j)}$  correspond to the  $(m + 1)^2$  constraints defining the symmetric matrix  $A$ , and we denote the corresponding dual variables  $Q_{ij} \in \mathbb{R}$ , with  $Q_{ij} = Q_{ji}$  for all  $i, j \in \llbracket 0, m \rrbracket$ . We write  $Q = (Q_{ij})_{0 \leq i, j \leq m}$ . The standard form of the dual program  $(\text{D-SDP}_n^{m,\geq})$  thus reads:

$$(\text{D-SDP}_n^{m,\geq}) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } Q \in \text{Sym}_{m+1} \\ \text{minimising } y \\ \text{subject to:} \\ \text{Diag}_{k=0,\dots,m} \left[ y - \sum_{l=0}^m \sum_{i+j=2l} Q_{ij} \binom{l}{k} l! \right] \oplus \frac{1}{2} Q \succeq E_{m+1}^{(n,n)} \oplus \mathbf{0}_{m+1}. \end{array} \right.$$

Due to the block-diagonal structure of the matrices involved, the positive semidefinite constraint above is equivalent to the following constraints:

$$\left\{ \begin{array}{l} y \geq 1 + \sum_{l=0}^m \sum_{i+j=2l} Q_{ij} \binom{l}{n} l!, \\ \forall k \in \llbracket 0, m \rrbracket \setminus \{n\}, \quad y \geq \sum_{l=0}^m \sum_{i+j=2l} Q_{ij} \binom{l}{k} l!, \\ Q \succeq 0. \end{array} \right. \quad (4.79)$$

For  $k \in \llbracket 0, m \rrbracket$ , we define

$$\mu_k := \sum_{l=0}^m \sum_{i+j=2l} Q_{ij} \binom{l}{k} l! \in \mathbb{R}. \quad (4.80)$$

We obtain the program:

$$(D\text{-SDP}_n^{m,\geq}) \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R}, \boldsymbol{\mu} \in \mathbb{R}^{m+1} \text{ and } Q \in \text{Sym}_{m+1} \\ \text{minimising } y \\ \text{subject to:} \\ y \geq 1 + \mu_n \\ \forall k \in \llbracket 0, m \rrbracket \setminus \{n\}, \quad y \geq \mu_k \\ \forall k \in \llbracket 0, m \rrbracket, \quad \mu_k = \sum_{l=0}^m \sum_{i+j=2l} Q_{ij} \binom{l}{k} l! \\ Q \succeq 0. \end{array} \right.$$

Combining Lemma 4.3 for  $u_k = \mu_k$  and  $v_l = \sum_{i+j=2l} Q_{ij}$  for all  $k, l \in \llbracket 0, m \rrbracket$  with the previous expression of  $(D\text{-SDP}_n^{m,\geq})$  we finally obtain:

$$(D\text{-SDP}_n^{m,\geq}) \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R}, \boldsymbol{\mu} \in \mathbb{R}^{m+1} \text{ and } Q \in \text{Sym}_{m+1} \\ \text{minimising } y \\ \text{subject to:} \\ y \geq 1 + \mu_n \\ \forall k \in \llbracket 0, m \rrbracket \setminus \{n\}, \quad y \geq \mu_k \\ \forall l \in \llbracket 0, m \rrbracket, \quad \sum_{i+j=2l} Q_{ij} = \frac{(-1)^l}{l!} \sum_{k=l}^m (-1)^k \binom{k}{l} \mu_k \\ Q \succeq 0, \end{array} \right.$$

which is the required form.

### Strong duality

This proof is a direct consequence of the proof of Theorem 4.2.

**Corollary 4.1.** *Strong duality holds between the programs  $(\text{SDP}_n^{m,\geq})$  and  $(\text{D-SDP}_n^{m,\geq})$ .*

*Proof.* The program  $(\text{SDP}_n^{m,\geq})$  is a relaxation of  $(\text{LP}_n^{L^2})$  and  $(\text{SDP}_n^{m,\leq})$  is a restriction of  $(\text{LP}_n^{L^2})$ , so  $(\text{SDP}_n^{m,\geq})$  is a relaxation of  $(\text{SDP}_n^{m,\leq})$ . Hence, the strictly feasible solution of  $(\text{SDP}_n^{m,\leq})$  derived in the proof of Theorem 4.2 yields a strictly feasible solution  $(A, \mathbf{F})$  for  $(\text{SDP}_n^{m,\geq})$ : we set  $\mathbf{F} = (F_0, \dots, F_m) \in \mathbb{R}^{m+1}$ , where for all  $k \in \llbracket 0, m \rrbracket$ ,  $F_k := \frac{1}{2^{m+1}-1} \binom{m+1}{k+1}$  and  $A = A_{\mathbf{F}} \in \text{Sym}_{m+1}$ , where  $A_{\mathbf{F}}$  is defined in Eq. (4.21).

With Slater's condition, this shows again that strong duality holds between the programs  $(\text{SDP}_n^{m,\geq})$  and  $(\text{D-SDP}_n^{m,\geq})$ .  $\square$

### 4.3.5 Convergence of the hierarchies of semidefinite programs

From the previous sections, for  $m \geq n$  the optimal values  $\omega_n^{m,\geq}$  and  $\omega_n^{m,\leq}$  of  $(\text{SDP}_n^{m,\geq})$  and  $(\text{SDP}_n^{m,\leq})$  form decreasing and increasing sequences, respectively, which satisfy

$$0 \leq \omega_n^{m,\leq} \leq \omega_n^{\mathcal{S}} \leq \omega_n^{L^2} \leq \omega_n^{m,\geq} \leq 1. \quad (4.81)$$

Recall that  $\omega_n^{L^2}$  is the optimal value of  $(\text{LP}_n^{L^2})$  while  $\omega_n^{\mathcal{S}}$  is the optimal value of  $(\text{LP}_n^{\mathcal{S}})$ . These sequences thus both converge. The remaining question is whether  $(\omega_n^{m,\geq})_m$  converges to  $\omega_n^{L^2}$  and  $(\omega_n^{m,\leq})_m$  converges to  $\omega_n^{\mathcal{S}}$ . In this section, we show that this is indeed the case. The problem of proving that  $\omega_n^{L^2} = \omega_n^{\mathcal{S}}$  is still open and we discuss this point later.

#### 4.3.5.1 Convergence of the sequence of upper bounds

**Theorem 4.3.** *The decreasing sequence of optimal values of  $(\text{SDP}_n^{m,\geq})$  converges to the optimal value of  $(\text{LP}_n^{L^2})$  that is*

$$\lim_{m \rightarrow +\infty} \omega_n^{m,\geq} = \omega_n^{L^2}. \quad (4.82)$$

In order to prove this theorem, we extract a limit from a sequence of optimal solutions of  $(\text{SDP}_n^{m,\geq})$ , for  $m \geq n$ , and we show using Theorem 4.1 that it provides a feasible solution of  $(\text{LP}_n^{L^2})$ .

*Proof.* For all  $m \geq n$ , the feasible set of  $(\text{SDP}_n^{m,\geq})$  is non-empty (consider, e.g.,  $\mathbf{F} = (1, 0, 0, \dots, 0) \in \mathbb{R}^{m+1}$ ). Moreover, due to the constraints  $\sum_{k=0}^m F_k = 1$  and  $F_k \geq 0$  for all  $k \leq m$ , the feasible set of  $(\text{SDP}_n^{m,\geq})$  is compact. Hence, the program  $(\text{SDP}_n^{m,\geq})$  has feasible optimal solutions, for all  $m \geq n$ , by diagonal extraction.

The matrix  $A$  in  $(\text{SDP}_n^{m,\geq})$  is entirely fixed by the choice of  $\mathbf{F}$ . Let  $(\mathbf{F}^m)_{m \geq n}$  be a sequence of optimal solutions of  $(\text{SDP}_n^{m,\geq})$ , for  $m \geq n$ . For each  $m \geq n$ , we have by optimality that  $F_n^m = \omega_n^{m,\geq}$ , and the sequence  $(F_n^m)_{m \geq n}$  converges. We complete each tuple  $\mathbf{F}^m = (F_0^m, F_1^m, \dots, F_m^m) \in \mathbb{R}^{m+1}$  with zeros to obtain a sequence in  $\mathbb{R}^{\mathbb{N}}$ , which we still denote  $\mathbf{F}^m = (F_0^m, F_1^m, \dots, F_m^m, 0, 0, \dots) \in \mathbb{R}^{\mathbb{N}}$  for simplicity.

Performing a diagonal extraction  $\phi$  on the sequence of optimal solutions  $(\mathbf{F}^m)_{m \geq n}$ , we obtain a sequence of sequences  $(\mathbf{F}^{\phi(m)})_{m \geq n}$  such that each sequence  $(F_k^{\phi(m)})_{m \geq n}$  converges when  $m \rightarrow +\infty$ , for all  $k \in \mathbb{N}$ . Let  $F_k$  denote its limit, for each  $k \in \mathbb{N}$ . We write  $\mathbf{F} = (F_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  the sequence of limits.

For all  $m \geq n$ ,  $F_k^{\phi(m)} \geq 0$  for all  $k \in \mathbb{N}$  and  $\sum_k F_k^{\phi(m)} = 1$ , so taking  $m \rightarrow +\infty$  we obtain  $F_k \geq 0$  for all  $k \in \mathbb{N}$ , and  $\sum_k F_k \leq 1$ . Moreover,

$$F_n = \lim_{m \rightarrow +\infty} F_n^m = \lim_{m \rightarrow +\infty} \omega_n^{m, \geq}. \quad (4.83)$$

For all  $m \geq n$ , we have  $\omega_n^{m, \geq} \geq \omega_n^{L^2}$ , so  $F_n \geq \omega_n^{L^2} > 0$ . In particular,  $\sum_k F_k > 0$ , so without loss of generality we may assume that  $\sum_k F_k = 1$  (otherwise we can always replace  $F_k$  by  $\frac{F_k}{\sum_l F_l} > F_k$ ).

Let  $f_{\mathbf{F}} = \sum_k F_k \mathcal{L}_k \in L^2(\mathbb{R}_+)$ . By construction we have (see program  $(\text{LP}_n^{\text{relax}, m})$ ):

$$\forall m \geq n, \forall g \in \mathcal{R}_{m, +}(\mathbb{R}_+), \langle f_{\mathbf{F}}, g \rangle \geq 0. \quad (4.84)$$

Hence, by Theorem 4.1,  $\mathbf{F}$  is the sequence of Laguerre moments of a nonnegative distribution over  $\mathbb{R}_+$  (the Lebesgue measure times the function  $f_{\mathbf{F}}$ ). In particular,

$$\forall x \in \mathbb{R}_+, f_{\mathbf{F}}(x) = \sum_k F_k \mathcal{L}_k(x) \geq 0. \quad (4.85)$$

With the constraints  $F_k \geq 0$  for all  $k \in \mathbb{N}$ , and  $\sum_k F_k = 1$ , this implies that  $\mathbf{F}$  is a feasible solution of  $(\text{LP}_n^{L^2})$ , and in particular  $F_n \leq \omega_n^{L^2}$ , since  $(\text{LP}_n^{L^2})$  is a maximisation problem. Since we already had  $F_n \geq \omega_n^{L^2}$  we obtain with Eq. (4.83):

$$\lim_{m \rightarrow +\infty} \omega_n^{m, \geq} = \omega_n^{L^2}, \quad (4.86)$$

which concludes the proof.  $\square$

This immediately implies strong duality for programs  $(\text{LP}_n^{L^2})$  and  $(\text{D-LP}_n^{L^2})$  because of the following remarks:

- (i) we have weak duality between those programs so that the optimal value  $\omega_n'^{L^2}$  of  $(\text{D-LP}_n^{L^2})$  upper bounds the optimal value of  $(\text{LP}_n^{L^2})$   $\omega_n^{L^2}$  i.e.  $\omega_n^{L^2} \leq \omega_n'^{L^2}$ ;
- (ii) we have strong duality between  $(\text{SDP}_n^{m, \geq})$  and  $(\text{D-SDP}_n^{m, \geq})$  by Corollary 4.1;
- (iii)  $(\text{SDP}_n^{m, \geq})$  is a relaxation of  $(\text{LP}_n^{L^2})$  so that  $\forall m, \omega_n^{m, \geq} \geq \omega_n^{L^2}$ ;
- (iv)  $(\text{D-SDP}_n^{m, \geq})$  is a restriction of  $(\text{D-LP}_n^{L^2})$  so that  $\forall m, \omega_n^{m, \geq} \geq \omega_n'^{L^2} \geq \omega_n^{L^2}$ ;
- (v) we showed that the optimal value of the hierarchy  $(\text{SDP}_n^{m, \geq})$  converges to  $\omega_n^{L^2}$  i.e.  $\lim_{m \rightarrow +\infty} \omega_n^{m, \geq} = \omega_n^{L^2}$ .

Thus necessarily  $\omega_n'^{L^2} = \omega_n^{L^2}$ .

### 4.3.5.2 Convergence of the sequence of lower bounds

**Theorem 4.4.** *The increasing sequence of optimal values of  $(\text{SDP}_n^{m,\leq})$  converges to the optimal value of  $(\text{LP}_n^{L^2})$ :*

$$\lim_{m \rightarrow +\infty} \omega_n^{m,\leq} = \omega_n^S. \quad (4.87)$$

As a sanity check, we will try to prove the harder result that it converges in order to  $\omega_n^{L^2}$  to see where might be the gap. The proof is similar to that of Theorem 4.3 using the dual programs: we attempt to construct a feasible optimal solution of  $(\text{D-LP}_n^{L^2})$  by extracting a limit from a sequence of optimal solutions of  $(\text{D-SDP}_n^{m,\leq})$ , for  $m \geq n$  and then conclude using the strong duality between  $(\text{SDP}_n^{m,\leq})$  and  $(\text{D-SDP}_n^{m,\leq})$ , which we proved in Theorem 4.2. However, it turns out that  $(\text{D-LP}_n^{L^2})$  may not have feasible optimal solutions in  $L^{2'}(\mathbb{R}_+)$  (as anticipated with the analytical optimal solutions for  $n = 1$  and  $n = 2$  from Section 4.3.3).

To deal with this issue, we have extended the formulation of  $(\text{D-LP}_n^{L^2})$  to a larger space where it has feasible optimal solutions, namely the space of tempered distributions  $\mathcal{S}'(\mathbb{R}_+)$  (see  $(\text{D-LP}_n^S)$ ).

Note that the semidefinite restrictions  $(\text{SDP}_n^{m,\leq})$  of  $(\text{LP}_n^{L^2})$  are also restrictions of  $(\text{LP}_n^S)$  for all  $m \geq n$ , while  $(\text{LP}_n^S)$  is itself a restriction of  $(\text{LP}_n^{L^2})$ , and  $(\text{D-LP}_n^S)$  is a relaxation of  $(\text{D-LP}_n^{L^2})$ . We denote by  $\omega_n'^S$  the optimal value of  $(\text{D-LP}_n^S)$ . Recall that the optimal value of  $(\text{LP}_n^S)$  is denoted  $\omega_n^S$  and that of  $(\text{LP}_n^{L^2})$  and  $(\text{D-LP}_n^{L^2})$  are denoted  $\omega_n^{L^2}$  (by strong duality). By weak duality of linear programming we thus have

$$\omega_n^{m,\leq} \leq \omega_n^S \leq \omega_n'^S \leq \omega_n^{L^2}, \quad (4.88)$$

for all  $m \geq n$ .

First we use a reformulation of  $(\text{D-SDP}_n^{m,\leq})$  over Schwartz space. For all  $m \geq n$ :

$$(\text{D-SDP}_n^{m,\leq}) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } \boldsymbol{\mu} \in \mathcal{S}'(\mathbb{N}) \\ \text{minimising } y \\ \text{subject to} \\ \quad \forall k \neq n \in \mathbb{N}, y \geq \mu_k \\ \quad y \geq 1 + \mu_n \\ \quad \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_{\boldsymbol{\mu}}, g \rangle \geq 0. \end{array} \right.$$

This amounts to develop  $\boldsymbol{\mu} \in \mathcal{S}'(\mathbb{R}_+)$  in  $(\text{D-LP}_n^S)$  in the Laguerre basis with a sequence  $\boldsymbol{\mu} \in \mathcal{S}'(\mathbb{N})$  and then relax the positivity constraint as before. Note that we only consider a sequence  $\boldsymbol{\mu} \in \mathbb{R}^{m+1}$  in  $(\text{D-SDP}_n^{m,\leq})$  and that by completing it with 0 we obtain automatically a sequence from  $\mathcal{S}'(\mathbb{N})$ .

Before proving Theorem 4.4, we provide a nontrivial analytical solution to the primal program  $(\text{SDP}_n^{m,\leq})$ . Let us define  $\mathbf{F}^m = (F_k^n)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  by

- if  $n$  is even:

$$F_k^n := \begin{cases} \frac{1}{2^n} \binom{k}{\frac{k}{2}} \binom{n-k}{\frac{n-k}{2}} & \text{when } k \leq n, k \text{ even,} \\ 0 & \text{otherwise,} \end{cases} \quad (4.89)$$

- if  $n$  is odd:

$$F_k^n := \begin{cases} \frac{1}{2^n} \frac{\binom{n}{\lfloor \frac{n}{2} \rfloor} \binom{\lfloor \frac{n}{2} \rfloor}{\lfloor \frac{k}{2} \rfloor}^2}{\binom{n}{k}}, & \text{when } k \leq n, \\ 0 & \text{otherwise.} \end{cases} \quad (4.90)$$

In both cases,

$$F_n^n = \frac{1}{2^n} \binom{n}{\lfloor \frac{n}{2} \rfloor}, \quad (4.91)$$

and we have that:

$$F_n^n \geq \frac{1}{n+1} \quad (4.92)$$

where we used  $\binom{n}{j} \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$  for all  $j \in \llbracket 0, n \rrbracket$ , summed over  $j$  from 0 to  $n$ . We extrapolated these analytical expressions from numerical values: running  $(\text{SDP}_n^{m, \leq})$  for several values of  $n$  and  $m$  allowed us to deduce these sequences (we acknowledge here the great help from [oeis.org](http://oeis.org)).

The proof of feasibility consists in checking that the constraints of  $(\text{SDP}_n^{m, \leq})$  are satisfied by  $F^n$ . To do so, we make use of Zeilberger's algorithm [Zei91], a powerful algorithm for proving binomial identities. The proof of optimality for  $m = n$  is obtained by deriving an analytical feasible solution of  $(\text{D-SDP}_n^{n, \leq})$  with the same optimal value. Note that the optimality of this solution does not play a role in the proof of convergence. A word of warning as this proof is rather long. We start by showing two results, corresponding to  $n$  even and  $n$  odd, respectively, where we make use of Zeilberger's algorithm, a powerful algorithm for proving binomial identities [Zei91]. Given a holonomic function, this algorithm outputs a recurrence relation that it satisfies, thus reducing the proof of identity between binomial expressions to the verification that the initialisation is correct. A Mathematica notebook is available for the implementation of Zeilberger's algorithm [CE].

**Lemma 4.8.** *For  $n \in \mathbb{N}$  even:*

$$\sum_{k=0}^n (-1)^k F_k^n L_k(x) = \sum_{l=0}^n x^l \sum_{i+j=2n-2l} p_{n-i} p_{n-j}, \quad (4.93)$$

where:

$$p_{n-k} := \begin{cases} \sqrt{\frac{1}{2^n n!} \binom{n}{\frac{n}{2}}}, & \text{when } k = 0, \\ (-1)^{\frac{k}{2}} 2^{\frac{k}{2}} \left(\frac{k}{2}\right)! \left(\frac{n}{2}\right)^2 p_n, & \text{when } 0 < k \leq n, k \text{ even,} \\ 0, & \text{otherwise.} \end{cases} \quad (4.94)$$

The coefficients  $p_{n-k}$  (and  $q_{n-k}$  later on) were found by hand when looking for an analytical sum-of-squares decomposition.

*Proof.* To prove the polynomial equality (4.93), we start by equating the coefficients in  $x^l$  for  $l \in \llbracket 0, n \rrbracket$  which gives:

$$\begin{aligned} \frac{(-1)^l}{l!} \sum_{k=l}^n (-1)^k \binom{k}{l} F_k^n &= \sum_{i+j=2n-2l} p_{n-i} p_{n-j} \\ &= \sum_{i=0}^{2n-2l} p_{n-i} p_{n-(2n-2l-i)} \\ &= \sum_{i=0}^{n-l} p_{n-2i} p_{n-(2n-2l-2i)}, \end{aligned} \quad (4.95)$$

where we used  $p_{n-i} = 0$  for  $n-i < 0$  in the second line. These are equalities between holonomic functions of parameters  $n$  and  $l$  that are trivial for  $l > n$ .

- For  $l \leq n$  even, because  $F_k^n = 0$  for  $k$  odd, Eq. (4.95) becomes:

$$\sum_{k=\frac{l}{2}}^{\frac{n}{2}} \frac{1}{l!} \binom{2k}{l} F_{2k}^n = \sum_{i=0}^{n-l} p_{n-2i} p_{n-(2n-2l-2i)}, \quad (4.96)$$

that is, taking into account the parity of  $l = 2s$  and  $n = 2t$ ,

$$\sum_{k=s}^t \frac{1}{(2s)!} \binom{2k}{2s} F_{2k}^{2t} = \sum_{i=0}^{2t-2s} p_{2t-2i} p_{2t-(4t-4s-2i)}. \quad (4.97)$$

Inserting the expressions from Eq. (4.89) and Eq. (4.94), we thus have to check the identity:

$$\sum_{k=s}^t \frac{1}{2^{2t}(2s)!} \binom{2k}{2s} \binom{2k}{k} \binom{2t-2k}{t-k} = \sum_{i=0}^{2t-2s} \frac{i!(2t-2s-i)! (2t)!}{2^{2s}(2t)!} \binom{2t}{t} \binom{t}{i}^2 \binom{t}{2t-2s-i}^2 \quad (4.98)$$

for all  $t \in \mathbb{N}$  and all  $s \leq t$  (with the convention  $\binom{k}{j} = 0$  for  $j > k$ ). We ran Zeilberger's algorithm to show that the right-hand side and the left-hand side of Eq. (4.97) both satisfy the following recurrence relation, for all  $s, t \in \mathbb{N}$ :

$$\begin{aligned} 2(t+1)^2 S(s, t) + (-2s^2 - 4t^2 + 4st + 5s - 11t - 8) S(s, t+1) \\ + (s-t-2)(2s-2t-3) S(s, t+2) = 0. \end{aligned} \quad (4.99)$$

It remains to check that the initialisation is correct. For all  $s \in \mathbb{N}$ , this recurrence relation in  $t$  is of order 2. Since the identities in Eq. (4.98) are trivial when  $l > n$ , i.e.,  $s > t$ , we thus only need to check Eq. (4.98) for  $(s, t) = (0, 0)$ ,  $(s, t) = (0, 1)$ , and  $(s, t) = (1, 1)$ , which is straightforward: we obtain the values 1, 1 and  $\frac{1}{4}$  respectively, for both sides of Eq. (4.98).

- For  $l \leq n$  odd, Eq. (4.95) becomes:

$$- \sum_{k=\frac{l+1}{2}}^{\frac{n}{2}} \frac{1}{l!} \binom{2k}{l} F_{2k}^n = \sum_{i=0}^{n-l} p_{n-2i} p_{n-(2n-2l-2i)}, \quad (4.100)$$

that is, taking into account the parity of  $l = 2s + 1$  and  $n = 2t$ :

$$- \sum_{k=s+1}^t \frac{1}{(2s+1)!} \binom{2k}{2s+1} F_{2k}^{2t} = \sum_{i=0}^{2t-2s-1} p_{2t-2i} p_{2t-(4t-4s-2i-2)}. \quad (4.101)$$

Inserting the expressions from Eq. (4.89) and Eq. (4.94), we thus have to check the identity:

$$\begin{aligned} & \sum_{k=s+1}^t \frac{1}{2^{2t}(2s+1)!} \binom{2k}{2s+1} \binom{2k}{k} \binom{2t-2k}{t-k} \\ &= \sum_{i=0}^{2t-2s-1} \frac{i!(2t-2s-i-1)!}{2^{2s+1}(2t)!} \binom{2t}{t} \binom{t}{i}^2 \binom{t}{2t-2s-i-1}^2 \end{aligned} \quad (4.102)$$

for all  $t \in \mathbb{N}$  and all  $s \leq t$  (with the convention  $\binom{k}{j} = 0$  for  $j > k$ ). Likewise, we ran Zeilberger's algorithm to show that the right-hand side and the left-hand side of Eq. (4.101) both satisfy the following recurrence relation, for all  $s, t \in \mathbb{N}$ :

$$\begin{aligned} & 2(t+1)^2 S(s, t) + (-2s^2 - 4t^2 + 4st + 3s - 9t - 6) S(s, t+1) \\ & + (s-t-1)(2s-2t-3) S(s, t+2) = 0. \end{aligned} \quad (4.103)$$

It remains to check that the initialisation is correct. For all  $s \in \mathbb{N}$ , this recurrence relation in  $t$  is of order 2. Since the identities in Eq. (4.102) are trivial when  $l > n$ , i.e.,  $s \geq t$ , we thus only need to check Eq. (4.102) for  $(s, t) = (0, 1)$ , which is straightforward: we obtain the value 1 for both sides of Eq. (4.102). □

Now we turn to the case where  $n \in \mathbb{N}$  is odd.

**Lemma 4.9.** *For  $n \in \mathbb{N}$  odd:*

$$\sum_{k=0}^n (-1)^k F_k^n L_k(x) = \sum_{l=0}^n x^l \sum_{i+j=2l} q_i q_j \quad (4.104)$$

where:

$$q_{n-k} := \begin{cases} \sqrt{\frac{1}{2^n n!} \binom{n}{\lfloor \frac{n}{2} \rfloor}}, & \text{when } k = 0, \\ (-1)^{\frac{k}{2}} 2^{\frac{k}{2}} \left(\frac{k}{2}\right)! \frac{n+1}{n-k+1} \left(\frac{\lfloor \frac{n}{2} \rfloor}{\frac{k}{2}}\right)^2 q_n, & \text{when } 0 < k < n, k \text{ even,} \\ 0. & \text{otherwise.} \end{cases} \quad (4.105)$$

*Proof.* Unlike the case where  $n$  is even,  $F^n$  is non-zero for  $k \leq n$  odd, and the expression of  $F_k^n$  depends on the parity of  $k$  (which is not linear in  $k$ ). Thus we cannot use directly Eq. (4.90) in Zeilberger's algorithm as we did in the previous lemma, hence the development below in order to obtain expressions that the algorithm can take as inputs. We fix  $n$  odd and  $l \leq n$ .

We start by equating coefficients in  $x^l$  in Eq. (4.104):

$$\begin{aligned} \frac{(-1)^l}{l!} \sum_{k=l}^n (-1)^k \binom{k}{l} F_k^n &= \sum_{i+j=2n-2l} q_{n-i} q_{n-j} \\ &= \sum_{i=0}^{n-l} q_{n-2i} q_{n-(2n-2l-2i)}. \end{aligned} \quad (4.106)$$

- For  $l$  even, writing  $l = 2s$  and  $n = 2t + 1$ , the left-hand side of Eq. (4.106) becomes:

$$\begin{aligned} &\frac{1}{(2s)!} \sum_{k=2s}^{2t+1} (-1)^k \binom{k}{2s} F_k^{2t+1} \\ &= \frac{1}{(2s)!} \sum_{k=0}^{2t+1-2s} (-1)^k \binom{k+2s}{2s} F_{k+2s}^{2t+1} \\ &= \frac{1}{(2s)!} \sum_{k=0}^{t-s} \left( \binom{2k+2s}{2s} F_{2k+2s}^{2t+1} - \binom{2k+2s+1}{2s} F_{2k+2s+1}^{2t+1} \right) \\ &= \frac{\binom{2t+1}{t}}{2^{2t+1} (2s)!} \sum_{k=0}^{t-s} \left( \frac{\binom{2k+2s}{2s} \binom{t}{k+s}^2}{\binom{2t+1}{2k+2s}} - \frac{\binom{2k+2s+1}{2s} \binom{t}{k+s}^2}{\binom{2t+1}{2k+2s+1}} \right) \\ &= q_{2t+1}^2 \frac{(2t+1)!}{(2s)!} \sum_{k=0}^{t-s} \frac{\binom{t}{k+s}^2 \binom{2k+2s}{2s}}{\binom{2t+1}{2k+2s}} \left( 1 - \frac{(2k+2s+1)^2}{(2k+1)(2t-2s-2k+1)} \right). \end{aligned} \quad (4.107)$$

With Eq. (4.105) and Eq. (4.106), we thus have to check the identity:

$$\begin{aligned} &\frac{(2t+1)!}{(2s)!} \sum_{k=0}^{t-s} \frac{\binom{t}{k+s}^2 \binom{2k+2s}{2s}}{\binom{2t+1}{2k+2s}} \left( 1 - \frac{(2k+2s+1)^2}{(2k+1)(2t-2s-2k+1)} \right) \\ &= \sum_{i=0}^{2t+1-2s} \frac{q_{2t+1-2i} q_{2t+1-(4t+2-4s-2i)}}{q_{2t+1}^2} \\ &= - \sum_{i=0}^{2t+1-2s} 2^{2t-2s+1} \frac{(2t+2)^2 i! (2t+1-2s-i)!}{(2t-2i+2)(4s+2i-2t)} \binom{t}{i}^2 \binom{t}{2t-2s-i+1} \end{aligned} \quad (4.108)$$

for all  $t \in \mathbb{N}$  and all  $s \leq t$  (with the convention  $\binom{k}{j} = 0$  for  $j > k$ ). Zeilberger's algorithm certifies that the right-hand side and the left-hand side of Eq. (4.108) both satisfy for all  $s, t \in \mathbb{N}$ :

$$\begin{aligned} &-32(t+2)^3(t+1)^2(t+3)S(s, t) \\ &\quad + 4(t+3)(t+2)(2s^2 + 4t^2 - 4st - 7s + 15t + 14)S(s, t+1) \\ &\quad + (-2s + 2t + 5)(s - t - 2)S(s, t+2) = 0. \end{aligned} \quad (4.109)$$

It remains to check that the initialisation is correct. For all  $s \in \mathbb{N}$ , this recurrence relation in  $t$  is of order 2. Since the identities in Eq. (4.108) are trivial when  $l > n$ , i.e.,  $s > t$ , we thus only need to check Eq. (4.108) for  $(s, t) = (0, 0)$ ,  $(s, t) = (0, 1)$ , and  $(s, t) = (1, 1)$ . We obtain the values 0, 0 and  $-8$  respectively, for both sides of Eq. (4.108).

- For  $l$  odd, writing  $l = 2s + 1$  and  $n = 2t + 1$ , the left-hand side of Eq. (4.106) becomes:

$$\begin{aligned}
 & \frac{-1}{(2s+1)!} \sum_{k=2s+1}^{2t+1} (-1)^k \binom{k}{2s+1} F_k^{2t+1} \\
 &= \frac{-1}{(2s+1)!} \sum_{k=0}^{2t-2s} (-1)^{k+2s+1} \binom{k+2s+1}{2s+1} F_{k+2s+1}^{2t+1} \\
 &= \frac{1}{(2s+1)!} \sum_{k=0}^{t-s} \left( -\binom{2k+2s}{2s+1} F_{2k+2s}^{2t+1} + \binom{2k+2s+1}{2s+1} F_{2k+2s+1}^{2t+1} \right) \quad (4.110) \\
 &= q_{2t+1}^2 \frac{(2t+1)!}{(2s+1)!} \sum_{k=0}^{t-s} \left( -\frac{\binom{2k+2s}{2s+1} \binom{t}{k+s}^2}{\binom{2t+1}{2k+2s}} + \frac{\binom{2k+2s+1}{2s+1} \binom{t}{k+s}^2}{\binom{2t+1}{2k+2s+1}} \right) \\
 &= q_{2t+1}^2 \frac{(2t+1)!}{(2s+1)!} \sum_{k=0}^{t-s} \frac{\binom{2k+2s}{2s+1} \binom{t}{k+s}^2}{\binom{2t+1}{2k+2s}} \left( -1 + \frac{(2k+2s+1)^2}{2k(2t-2k-2s+1)} \right),
 \end{aligned}$$

where we used that  $\binom{2k+2s}{2s+1} = 0$  for  $k = 0$  in the third line. Note that when factorising we introduced an indeterminate form in the last line that Zeilberger's algorithm can resolve. This is necessary since the algorithm cannot deal with differences of binomial terms. With Eq. (4.105) and Eq. (4.106), we thus have to check the identity:

$$\begin{aligned}
 & \frac{(2t+1)!}{(2s+1)!} \sum_{k=0}^{t-s} \frac{\binom{2k+2s}{2s+1} \binom{t}{k+s}^2}{\binom{2t+1}{2k+2s}} \left( -1 + \frac{(2k+2s+1)^2}{2k(2t-2k-2s+1)} \right) \\
 &= \sum_{i=0}^{2t-2s} \frac{q_{2t+1-2i} q_{2t+1-(4t-4s-2i)}}{q_{2t+1}^2} \quad (4.111) \\
 &= \sum_{i=0}^{2t-2s} 2^{2t-2s} \frac{(2t+2)^2 i! (2t-2s-i)!}{(2t-2i+2)(4s+2i-2t+2)} \binom{t}{i}^2 \binom{t}{2t-2s-i}^2,
 \end{aligned}$$

for all  $t \in \mathbb{N}$  and all  $s \leq t$  (with the convention  $\binom{k}{j} = 0$  for  $j > k$ ). Zeilberger's algorithm certifies that both the right-hand side and the left-hand side of Eq. (4.111) satisfy for all  $s \leq t$ :

$$\begin{aligned}
 & -32(t+2)^3(t+1)^2(t+3)S(s,t) \\
 & \quad + 4(t+3)(t+2)(2s^2+4t^2-4st-5s+13t+11)S(s,t+1) \quad (4.112) \\
 & \quad + (-2s+2t+3)(s-t-2)S(s,t+2) = 0.
 \end{aligned}$$

It remains to check that the initialisation is correct. For all  $s \in \mathbb{N}$ , this recurrence relation in  $t$  is of order 2. Since the identities in Eq. (4.111) are trivial when  $l > n$ , i.e.,  $s > t$ , we thus only need to check Eq. (4.111) for  $(s, t) = (0, 0)$ ,  $(s, t) = (0, 1)$ , and  $(s, t) = (1, 1)$ . We obtain the values 1, 16 and 1 respectively, for both sides of Eq. (4.111).

□

Having derived these identities, we prove the lemma of interest before proving the theorem on the convergence of the lower bounding hierarchy:

**Lemma 4.10.** *For all  $m \geq n$ ,  $\mathbf{F}^n$  is a feasible solution of  $(\text{SDP}_n^{m,\leq})$ . Moreover, it is optimal when  $m = n$ .*

*Proof.* The proof has three parts. In the first we focus on  $n$  even, and in the second on  $n$  odd. In the last part, we exhibit a feasible solution of  $(\text{D-SDP}_n^{m,\leq})$  for  $m = n$  with the same optimal value  $F_n^n$ .

•  **$n$  even,  $m \geq n$ :** we check that  $\mathbf{F}^n$  defined in Eq. (4.89) satisfies all the constraints of  $(\text{SDP}_n^{m,\leq})$ .

- For all  $k \in \mathbb{N}$ ,  $F_k^n \geq 0$ .
- We have

$$\begin{aligned} \sum_{k=0}^{\infty} F_k^n &= \sum_{\substack{k=0 \\ \text{even}}}^n \frac{1}{2^n} \binom{k}{\frac{k}{2}} \binom{n-k}{\frac{n-k}{2}} \\ &= \sum_{k=0}^{\frac{n}{2}} \frac{1}{2^n} \binom{2k}{k} \binom{n-2k}{\frac{n}{2}-k} \\ &= 1, \end{aligned} \tag{4.113}$$

where the last equality follows from [Gou72, (3.90)].

- We have to show that  $x \mapsto \sum_{k=0}^n (-1)^k F_k^n L_k(x^2)$  is a positive function on  $\mathbb{R}$ . Due to Lemma 4.5, we aim to find a sum-of-squares decomposition for this polynomial. Guided by numerical results, we look for a polynomial  $P(x) := \sum_{i=0}^n p_i x^i$  such that:

$$\begin{aligned} \sum_{k=0}^n (-1)^k F_k^n L_k(x^2) &= P^2(x^2) \\ &= \left( \sum_{i=0}^n p_i x^{2i} \right)^2 \\ &= \sum_{l=0}^n \left( \sum_{\substack{i+j=l \\ 0 \leq i,j \leq n}} p_i p_j \right) x^{2l}, \end{aligned} \tag{4.114}$$

and the sought coefficients are given by Lemma 4.8, which concludes the first part of the proof.

•  **$n$  odd,  $m \geq n$ :** similarly, we check that  $\mathbf{F}^n$  defined in Eq. (4.90) satisfies all the constraints of  $(\text{SDP}_n^{m,\leq})$ .

- For all  $k \in \mathbb{N}$ ,  $F_k^n \geq 0$ .

- Writing  $n = 2t + 1$ , from Eq. (4.90) we have

$$\begin{aligned} \forall s \leq t, F_{2s}^{2t+1} &= \frac{1}{2^{2t+2}} \left(1 - \frac{s}{t+1}\right) \binom{2s}{s} \binom{2t+2-2s}{t+1-s} \\ \forall s \in \llbracket 1, t+1 \rrbracket, F_{2s-1}^{2t+1} &= \frac{1}{2^{2t+2}} \frac{s}{t+1} \binom{2s}{s} \binom{2t+2-2s}{t+1-s}. \end{aligned} \quad (4.115)$$

In particular, for all  $s \leq t$  we have

$$F_{2s-1}^{2t+1} + F_{2s}^{2t+1} = F_{2s}^{2t+2}, \quad (4.116)$$

where  $F_{2s}^{2t+2}$  is defined in Eq. (4.89). Since  $F_{2s+1}^{2t+2} = 0$  for all  $s \leq t$ , and  $F_k^n = 0$  for all  $k > n$ , we have:

$$\begin{aligned} \sum_{k=0}^{\infty} F_k^n &= \sum_{k=0}^{\infty} F_k^{n+1} \\ &= 1, \end{aligned} \quad (4.117)$$

where we used Eq. (4.113).

- We have to show that  $x \mapsto \sum_{k=0}^n (-1)^k F_k^n L_k(x^2)$  is a positive function on  $\mathbb{R}$ . Due to Lemma 4.5, we aim to find a sum-of-squares decomposition for this polynomial. Guided by numerical results, we look for a polynomial  $Q(x) := \sum_{i=0}^n q_i x^i$  such that:

$$\begin{aligned} \sum_{k=0}^n (-1)^k F_k^n L_k(x^2) &= Q^2(x^2) \\ &= \left( \sum_{i=0}^n q_i x^{2i} \right)^2 \\ &= \sum_{l=0}^n \left( \sum_{\substack{i+j=l \\ 0 \leq i, j \leq n}} q_i q_j \right) x^{2l}, \end{aligned} \quad (4.118)$$

and the sought coefficients are given by Lemma 4.9, which concludes the second part of the proof.

We thus obtained a feasible solution of  $(\text{LP}_n^{L^2})$  for all  $n \in \mathbb{N}^*$ , which is feasible for  $(\text{SDP}_n^{m, \leq})$  for all  $m \geq n$ .

- **Optimality for  $m = n$ :** we now find an analytical solution of the dual  $(\text{D-SDP}_n^{n, \leq})$  with the same optimal value as the primal  $(\text{SDP}_n^{n, \leq})$ , by finding a Cholesky decomposition for the matrix appearing in the dual program  $(\text{D-SDP}_n^{m, \leq})$  for  $m = n$ . The coefficients of the Cholesky

decomposition are given by the triangular matrix  $L$  with coefficients:

$$\begin{aligned}
 \forall j \leq i, \quad l_{2i,2j} &= 2^i i! \binom{i}{j} \\
 \forall j \leq i, \quad l_{2i+1,2j+1} &= 2^{i+1/2} \frac{(i+1)!}{\sqrt{j+1}} \binom{i}{j} \\
 l_{nn} &= 0 \\
 l_{ij} &= 0 \quad \text{otherwise.}
 \end{aligned} \tag{4.119}$$

Once again, these analytical expressions were extrapolated from numerical values. Then,  $A = LL^T$  is a positive semidefinite matrix given by:

$$A_{ij} = \sum_{k=0}^{\min(i,j)} l_{ik} l_{jk}. \tag{4.120}$$

Now  $l_{ij}$  is non-zero only when  $i$  and  $j$  have the same parity, so for all  $k \in \llbracket 0, \min(i, j) \rrbracket$ ,  $i$  and  $j$  must have the same parity than  $k$  for  $l_{ik} l_{jk}$  to be non-zero.

- Suppose  $i = 2i'$ ,  $j = 2j'$ ,  $i' \leq j'$  and  $(i', j') \neq (n, n)$ . Furthermore let  $l = \frac{i+j}{2}$ .

$$\begin{aligned}
 A_{ij} &= \sum_{\substack{k=0 \\ k \text{ even}}}^i l_{2i',k} l_{2j',k} \\
 &= \sum_{k=0}^{i'} 2^{i'} i'! \binom{i'}{k} 2^{j'} j'! \binom{j'}{k} \\
 &= 2^l i'! (l - i')! \sum_{k=0}^{i'} \binom{i'}{k} \binom{l - i'}{k} \\
 &= 2^l l!.
 \end{aligned} \tag{4.121}$$

- Suppose  $i = 2i' + 1$ ,  $j = 2j' + 1$ ,  $i' \leq j'$  and  $(i', j') \neq (n, n)$ . Furthermore let  $l = \frac{i+j}{2}$ .

$$\begin{aligned}
 A_{ij} &= \sum_{\substack{k=0 \\ k \text{ odd}}}^i l_{2i'+1,k} l_{2j'+1,k} \\
 &= \sum_{k=0}^{i'} 2^{i'+1/2} \frac{(i'+1)!}{\sqrt{k+1}} \binom{i'}{k} 2^{j'+1/2} \frac{(j'+1)!}{\sqrt{k+1}} \binom{j'}{k} \\
 &= 2^l i'! (l - i')! \sum_{k=0}^{i'} \binom{i'+1}{k+1} \binom{l - i' - 1}{k} \\
 &= 2^l l!.
 \end{aligned} \tag{4.122}$$

- Suppose  $n = 2t$ :

$$\begin{aligned}
 A_{nn} &= \sum_{k=0}^t l_{2t,2k}^2 \\
 &= 2^{2t} (t!)^2 \sum_{k=0}^{t-1} \binom{t}{k}^2 \\
 &= 2^n (t!)^2 \left( \binom{2t}{t} - 1 \right) \\
 &= 2^n n! \left( 1 - \binom{n}{\lfloor \frac{n}{2} \rfloor}^{-1} \right).
 \end{aligned} \tag{4.123}$$

- Suppose  $n = 2t + 1$ :

$$\begin{aligned}
 A_{nn} &= \sum_{\substack{k=0 \\ k \text{ odd}}}^t l_{2t+1,k}^2 \\
 &= 2^{2t+1} (t+1)!^2 \sum_{k=0}^{t-1} \frac{1}{k+1} \binom{t}{k}^2 \\
 &= 2^n t! (t+1)! \sum_{k=0}^{t-1} \binom{t+1}{k+1} \binom{t}{k} \\
 &= 2^n t! (t+1)! \left( \binom{2t+1}{t} - 1 \right) \\
 &= 2^n n! \left( 1 - \binom{n}{\lfloor \frac{n}{2} \rfloor}^{-1} \right).
 \end{aligned} \tag{4.124}$$

In both cases,  $A$  is indeed constructed as:

$$(A_{\mu})_{i,j} = \begin{cases} \sum_{k=0}^l \mu_k \binom{l}{k} l! & \text{when } i+j = 2l, \\ 0 & \text{otherwise,} \end{cases} \tag{4.125}$$

for  $\mu = (F_n^n, F_n^n, \dots, F_n^n, 1 - F_n^n)$  with  $F_n^n = \frac{1}{2^n} \binom{n}{\lfloor \frac{n}{2} \rfloor}$ , and this provides a feasible solution of (D-SDP $_n^{m,\leq}$ ) for  $m = n$ , with value  $F_n^n$ . This shows the optimality of  $\mathbf{F}^n$  for (SDP $_n^{n,\leq}$ ) (and the fact that strong duality holds between the programs (SDP $_n^{n,\leq}$ ) and (D-SDP $_n^{n,\leq}$ ), which we already knew from Theorem 4.2).

□

For (D-SDP $_n^{m,\leq}$ ), we see numerically that the optimal solution is the same then for (D-SDP $_n^{n,\leq}$ ), for a few values of  $m$  greater than  $n$ . However, this is no longer the case for higher values, for example when  $n = 3$  and  $m = 10$ .

We also obtain the following analytical lower bound for the optimal value of (LP $_n^{L^2}$ ):

$$\omega_n^{L^2} \geq \omega_n^{n,\leq} = \frac{1}{2^n} \binom{n}{\lfloor \frac{n}{2} \rfloor} \underset{n \rightarrow +\infty}{\sim} \sqrt{\frac{2}{\pi n}}, \tag{4.126}$$

which is superseded by numerical bounds when  $n \geq 3$  (see Table 4.1).

We now combine the reformulation of program (D-SDP $_n^{m,\leq}$ ) and Lemma 4.10 to prove Theorem 4.4.

*Proof.* The feasible set of (D-SDP $_n^{m,\leq}$ ) is non-empty, by considering the null sequence, which achieves value 1. Without loss of generality, we add the constraint  $y \leq 1$  in (D-LP $_n^S$ ) and (D-SDP $_n^{m,\leq}$ ).

Let  $m \geq n$  and let  $(y, \boldsymbol{\mu}) \in \mathbb{R} \times S'(\mathbb{N})$  be a feasible solution of (D-SDP $_n^{m,\leq}$ ). The constraint  $\langle f_{\boldsymbol{\mu}}, x \mapsto e^{-\frac{x}{2}} \rangle \geq 0$  implies  $\mu_0 \geq 0$  and thus  $y \geq 0$ . Without loss of generality, we may set  $\mu_k = 0$  for  $k > m$ , since these coefficients are only constrained by  $\mu_k \leq y \leq 1$ . We also have  $\mu_k \leq 1$  for all  $k \in \mathbb{N}$ .

By Lemma 4.10,  $\mathbf{F}^l$  is a feasible solution of (SDP $_l^{m,\leq}$ ) for all  $l \leq m$ , so in particular  $f_{\mathbf{F}^l} = \sum_{k=0}^l F_k^l \mathcal{L}_k \in \mathcal{R}_{m,+}(\mathbb{R}_+)$ . Hence,  $\boldsymbol{\mu}$  must satisfy the constraint  $\langle f_{\boldsymbol{\mu}}, f_{\mathbf{F}^l} \rangle \geq 0$ , which gives

$$\sum_{k=0}^l \mu_k F_k^l \geq 0, \quad (4.127)$$

for all  $l \leq m$ . Thus we have, for all  $l \in \llbracket 1, m \rrbracket$ ,

$$\begin{aligned} \mu_l &\geq -\frac{1}{F_l^l} \sum_{k=0}^{l-1} \mu_k F_k^l \\ &\geq -\frac{1}{F_l^l} \sum_{k=0}^{l-1} F_k^l \\ &= 1 - \frac{1}{F_l^l} \\ &\geq -l, \end{aligned} \quad (4.128)$$

where we used  $F_l^l > 0$  in the first line,  $\mu_k \leq 1$  and  $F_k^l \geq 0$  in the second line,  $\sum_{k=0}^l F_k^l = 1$  in the third line and Eq. (4.92) in the last line. With  $\mu_k \leq 1$ , we obtain  $|\mu_k| \leq k$  for  $k \in \mathbb{N}^*$ , and thus  $|\mu_k| \leq k + 1$ <sup>8</sup> for all  $k \in \mathbb{N}$ . Hence, the feasible set of (D-SDP $_n^{m,\leq}$ ) is compact and the program (D-SDP $_n^{m,\leq}$ ) has feasible optimal solutions for all  $m \geq n$ , by diagonal extraction.

Let  $(y^m, \boldsymbol{\mu}^m)_{m \geq n}$  be a sequence of optimal solutions of (D-SDP $_n^{m,\leq}$ ), for  $m \geq n$ . By Theorem 4.2, we have strong duality between the programs (SDP $_n^{m,\leq}$ ) and (D-SDP $_n^{m,\leq}$ ), so the optimal value of (D-SDP $_n^{m,\leq}$ ) is given by  $\omega_n^{m,\leq}$ , for all  $m \geq n$ . By optimality  $y^m = \omega_n^{m,\leq}$ , for all  $m \geq n$ , and the sequence  $(y^m)_{m \geq n}$  converges.

Performing a diagonal extraction  $\phi$  on the sequence  $(\boldsymbol{\mu}^m)_{m \geq n}$ , we obtain a sequence of sequences  $(\boldsymbol{\mu}^{\phi(m)})_{m \geq n}$  such that each sequence  $(\mu_k^{\phi(m)})_{m \geq n}$  converges when  $m \rightarrow +\infty$ , for all  $k \in \mathbb{N}$ . Let  $\mu_k$  denote its limit, for each  $k \in \mathbb{N}$ . We write  $\boldsymbol{\mu} = (\mu_k)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  the sequence of limits. We also write

$$y := \lim_{m \rightarrow +\infty} y^m = \lim_{m \rightarrow +\infty} \omega_n^{m,\leq}. \quad (4.129)$$

<sup>8</sup>Critically this bounds does not depend on the level  $m$  of the hierarchy.

For all  $m \geq n$ , we have  $\omega_n^{\phi(m), \leq} \geq \mu_k^{\phi(m)}$  for all  $k \in \mathbb{N}$  and  $\omega_n^{\phi(m), \leq} \geq 1 + \mu_n^{\phi(m)}$ , so taking  $m \rightarrow +\infty$  we obtain  $y \geq \mu_k$  for all  $k \in \mathbb{N}$  and  $y \geq 1 + \mu_n$ . By Eq. (4.88), we have  $\omega_n^{m, \leq} \leq \omega_n^S$  for all  $m \geq n$ , so  $y \leq \omega_n^S$ .

Moreover,  $|\mu_k^{\phi(m)}| \leq k + 1$  for all  $k \in \mathbb{N}$ , so taking  $m \rightarrow +\infty$  we obtain  $|\mu_k| \leq k + 1$  for all  $k \in \mathbb{N}$ , which implies that  $\boldsymbol{\mu} \in \mathcal{S}'(\mathbb{N})$  [GT71]. Let  $f_{\boldsymbol{\mu}} = \sum_k \mu_k \mathcal{L}_k \in \mathcal{S}'(\mathbb{R}_+)$ . We have

$$\mu_k = \langle f_{\boldsymbol{\mu}}, \mathcal{L}_k \rangle. \quad (4.130)$$

By construction we also have:

$$\forall m \geq n, \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+), \langle f_{\boldsymbol{\mu}}, g \rangle \geq 0. \quad (4.131)$$

By Theorem 4.1, this implies that the distribution  $\mu := f_{\boldsymbol{\mu}}(x) \in \mathcal{S}'(\mathbb{R}_+)$  is nonnegative, i.e.,

$$\forall f \in \mathcal{S}_+(\mathbb{R}_+), \langle f_{\boldsymbol{\mu}}, f \rangle \geq 0. \quad (4.132)$$

With the constraints  $y \geq \mu_k$  for all  $k \in \mathbb{N}$  and  $y \geq 1 + \mu_n$ , we have that  $(y, \mu)$  is a feasible solution of (D-LP $_n^S$ ), and in particular  $y \geq \omega_n'^S$ , since (D-LP $_n^S$ ) is a minimisation problem. Since  $y \leq \omega_n^S$  we obtain with Eq. (4.88) and Eq. (4.129):

$$\lim_{m \rightarrow +\infty} \omega_n^{m, \leq} = \omega_n^S = \omega_n'^S. \quad (4.133)$$

□

As a direct corollary of the proofs of Theorem 4.3 and Theorem 4.4 (in the same spirit than the remark after the proof of Theorem 4.3), we obtain the following strong duality result:

**Corollary 4.2.** *Strong duality holds between the programs (LP $_n^S$ ) and (D-LP $_n^S$ ) and between the programs (LP $_n^{L^2}$ ) and (D-LP $_n^{L^2}$ ).*

We thus have shown the convergence of the semidefinite hierarchies of upper and lower bounds:  $(\omega_n^{m, \geq})_{m \geq n}$  towards the optimal value of (LP $_n^{L^2}$ ) and  $(\omega_n^{m, \leq})_{m \geq n}$  towards the optimal value of (LP $_n^S$ ). By linearity, these results generalise straightforwardly to the case of witnesses corresponding to linear combinations of fidelities with displaced Fock states:

$$\lim_{m \rightarrow +\infty} \omega_a^{m, \leq} = \omega_a^S, \quad (4.134)$$

$$\lim_{m \rightarrow +\infty} \omega_a^{m, \geq} = \omega_a^{L^2}, \quad (4.135)$$

for all  $n \in \mathbb{N}^*$  and all  $\mathbf{a} \in [0, 1]^n$  where, as an example,  $\omega_{\mathbf{a}}^{L^2}$  is the value of the following program:

$$(\text{LP}_{\mathbf{a}}^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } (F_k)_{k \in \mathbb{N}} \in \ell^2(\mathbb{N}) \\ \text{maximising } \sum_{k=1}^n a_k F_k \\ \text{subject to} \\ \sum_k F_k = 1 \\ \forall k \in \mathbb{N}, \quad F_k \geq 0 \\ \forall x \in \mathbb{R}_+, \quad \sum_k F_k \mathcal{L}(x) \geq 0, \end{array} \right.$$

and its dual program reads:

$$(\text{D-LP}_{\mathbf{a}}^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } \mu \in L^2(\mathbb{R}_+) \\ \text{minimising } y \\ \text{subject to} \\ \forall k \leq n \in \mathbb{N}, \quad y \geq a_k + \int_{\mathbb{R}_+} \mathcal{L}_k d\mu \\ \forall k > n \in \mathbb{N}, \quad y \geq \int_{\mathbb{R}_+} \mathcal{L}_k d\mu \\ \forall f \in L^2_+(\mathbb{R}_+), \quad \langle \mu, f \rangle = \int_{\mathbb{R}_+} f d\mu \geq 0. \end{array} \right.$$

The relaxation is given by  $(\text{SDP}_{\mathbf{a}}^{m, \geq})$  and the restriction is given by  $(\text{SDP}_{\mathbf{a}}^{m, \leq})$ .

### 4.3.5.3 Discussion on the possible gap between $\omega_n^{L^2}$ and $\omega_n^S$

In practice, we have two hierarchies providing numerical lower bounds and upper bounds on the threshold value. A natural question that now arises is the following: is there a gap between the optimal values of  $(\text{LP}_n^{L^2})$  and  $(\text{LP}_n^S)$ ? This is left as an open question but we provide several leads.

One potential direction would be to prove there is convergence at a finite rank in the hierarchy (see [Nie13]). This would automatically implies the equivalence of the formulation in the space of square integrable functions and in the space of Schwartz functions.

Another direction would be to prove that the upper hierarchy actually converges towards the optimal value or  $(\text{LP}_n^S)$  rather than  $(\text{LP}_n^{L^2})$ . However the limit  $(F_k)_k$  of optimal feasible solutions for  $(\text{SDP}_n^{m, \geq})$  is only constrained by  $\sum_k F_k = 1$  which gives that  $(F_k)_k$  belongs to  $\ell^2$  but not to  $\mathcal{S}(\mathbb{N})$  in general.

Another lead would be to prove the other direction i.e. that the upper lower hierarchy converges towards the optimal value of  $(\text{LP}_n^{L^2})$  rather than  $(\text{LP}_n^S)$ . But as we saw in the proof above, the analytical feasible solution only ensures that for all  $k \in \mathbb{N}$ ,  $|\mu_k| \leq k + 1$  and so that

$(\mu)_k \in \mathcal{S}'(\mathbb{N})$ . To prove a better bound on  $\mu_k$  so that  $(\mu)_k \in \ell^2(\mathbb{N})$ , we would need to exhibit another analytical solution which seems very difficult.

Finally, we also considered proving that any feasible solution of  $(\text{LP}_n^{L^2})$  (resp.  $(\text{D-LP}_n^{\mathcal{S}})$ ) can be approximated by a feasible solution of  $(\text{LP}_n^{\mathcal{S}})$  (resp.  $(\text{D-LP}_n^{L^2})$ ). However the constraints present in the programs prevent us from using standard density results.

## 4.4 Witnessing multimode Wigner negativity

In this section we discuss the generalisation of our Wigner negativity witnesses to the more challenging multimode setting. Hereafter,  $M$  denotes the number of modes.

### 4.4.1 Multimode notations

We use the multi-index notations presented in Eq. 1. A multivariate polynomial of degree  $p \in \mathbb{N}$  may then be written in the compact form  $P(\mathbf{x}) = \sum_{|\mathbf{l}| \leq p} p_{\mathbf{l}} \mathbf{x}^{\mathbf{l}}$ , where the sum is over all the tuples  $\mathbf{l} \in \mathbb{N}^M$  such that  $|\mathbf{l}| \leq m$ , also known as the weak compositions of the integer  $m$ . There are  $s(m)$  such tuples. In what follows, we will also consider sums over all the tuples  $\mathbf{l} \in \mathbb{N}^M$  such that  $\mathbf{l} \leq \mathbf{k}$ , for  $\mathbf{k} = (k_1, \dots, k_M) \in \mathbb{N}^M$ . There are  $\pi_{\mathbf{k}}$  such tuples. In particular, for all  $\mathbf{x} \in \mathbb{R}_+^M$  and all  $\mathbf{k} \in \mathbb{N}^M$ ,

$$L_{\mathbf{k}}(\mathbf{x}) = \sum_{\mathbf{l} \leq \mathbf{k}} \frac{(-1)^{|\mathbf{l}|}}{\mathbf{l}!} \binom{\mathbf{k}}{\mathbf{l}} \mathbf{x}^{\mathbf{l}} \quad \text{and} \quad \mathbf{x}^{\mathbf{k}} = \sum_{\mathbf{l} \leq \mathbf{k}} (-1)^{|\mathbf{l}|} \binom{\mathbf{k}}{\mathbf{l}} \mathbf{k}! L_{\mathbf{l}}(\mathbf{x}). \quad (4.136)$$

Note that there are several ways of considering relaxations or restrictions of fixed degree for the multimode case. The ‘triangle’ approach is to fix a degree  $m \in \mathbb{N}$  and consider monomials of degrees  $\mathbf{l} \in \mathbb{N}^M$  such that  $|\mathbf{l}| \leq m$ . Another way—the ‘rectangle’ approach—is to fix a degree  $\mathbf{m} \in \mathbb{N}^M$  and consider monomials of degrees  $\mathbf{l} \in \mathbb{N}^M$  such that  $\mathbf{l} \leq \mathbf{m}$ . The former is usually used in literature (see [Las09]) while we will need the latter for proving the convergence of the lower bounding hierarchy. These two procedures are interleaved and the convergence of one will provide the convergence of the other.

We now extend a few definitions from the single-mode case. For  $\mathbf{s} = (s_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^M} \in \mathbb{R}^{\mathbb{N}^M}$ , we define the associated formal series of multivariate Laguerre functions:

$$f_{\mathbf{s}} := \sum_{\mathbf{k}} s_{\mathbf{k}} \mathcal{L}_{\mathbf{k}}, \quad (4.137)$$

where  $\mathbf{s}$  is the so-called sequence of Laguerre moments of  $f_{\mathbf{s}}$ . For  $m \in \mathbb{N}$ , we define the associated  $s(m) \times s(m)$  moment matrix  $A_{\mathbf{s}}$  by

$$(A_{\mathbf{s}})_{\mathbf{i}, \mathbf{j}} = \begin{cases} \sum_{\mathbf{k} \leq \mathbf{l}} s_{\mathbf{k}} \binom{\mathbf{l}}{\mathbf{k}} \mathbf{l}! & \text{when } \mathbf{i} + \mathbf{j} = 2\mathbf{l}, \\ 0 & \text{otherwise,} \end{cases} \quad (4.138)$$

where  $\mathbf{i}, \mathbf{j} \in \mathbb{N}^M$  with  $|\mathbf{i}| \leq m$  and  $|\mathbf{j}| \leq m$ .

The multimode Laguerre functions  $(\mathcal{L}_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^M}$  form an orthonormal basis of the space  $L^2(\mathbb{R}_+^M)$  of real square-integrable functions over  $\mathbb{R}_+^M$  equipped with the usual scalar product:

$$\langle f, g \rangle = \int_{\mathbb{R}_+^M} f(\mathbf{x})g(\mathbf{x})d\mathbf{x}, \quad (4.139)$$

for  $f, g \in L^2(\mathbb{R}_+^M)$ . We denote by  $L_+^2(\mathbb{R}_+^M)$  its subset of nonnegative elements. The space  $L^2(\mathbb{R}_+^M)$  is isomorphic to its dual space  $L^{2'}(\mathbb{R}_+^M)$ : elements of  $L^{2'}(\mathbb{R}_+^M)$  can be identified by the Lebesgue measure on  $\mathbb{R}_+^M$  times the corresponding function in  $L^2(\mathbb{R}_+^M)$ . The space  $L^2(\mathbb{R}_+^M)$  is also isomorphic to  $\ell^2(\mathbb{N}^M)$  via expansion on the Laguerre basis.

Moreover, the elements of the space  $\mathcal{S}(\mathbb{R}_+^M)$  of Schwartz functions over  $\mathbb{R}_+^M$ , i.e., the space of  $C^\infty$  functions that go to 0 at infinity faster than any inverse polynomial, as do their derivatives, can be written as series of Laguerre functions with a sequence indexed by  $\mathbb{N}^M$  of rapidly decreasing coefficients  $\mathcal{S}(\mathbb{N}^M)$  (which go to 0 at infinity faster than any inverse  $M$ -variate polynomial). Its dual space  $\mathcal{S}'(\mathbb{R}_+^M)$  of tempered distributions over  $\mathbb{R}_+^M$  is characterised as the space of formal series of Laguerre functions with a slowly increasing sequence of coefficients  $\mathcal{S}'(\mathbb{N}^M)$  (sequences that are upper bounded by an  $M$ -variate polynomial) [GT71]. We also extend the definition of the duality  $\langle -, - \rangle$  in Eq. (4.139) to these spaces.

For all  $m \in \mathbb{N}$ , the set of series of Laguerre functions over  $\mathbb{R}_+^M$  truncated at  $m$  is denoted  $\mathcal{R}_m(\mathbb{R}_+^M)$ . This is the set of  $M$ -variate polynomials  $P(\mathbf{x}) = \sum_{|\mathbf{k}| \leq m} p_{\mathbf{k}} \mathbf{x}^{\mathbf{k}}$  of degree at most  $m$  multiplied by the function  $\mathbf{x} \mapsto e^{-\frac{\mathbf{x}}{2}}$ . Let  $\mathcal{R}_{m,+}(\mathbb{R}_+^M)$  denotes its subset of nonnegative elements where the polynomial  $P$  is such that  $\mathbf{x} \mapsto P(\mathbf{x}^2)$  has a sum-of-squares decomposition.

Similarly, for  $\mathbf{m} \in \mathbb{N}^M$ , the set of truncated series of Laguerre functions over  $\mathbb{R}_+^M$  with monomials smaller than  $\mathbf{m}$  is denoted  $\mathcal{R}_{\mathbf{m}}(\mathbb{R}_+^M)$  (with a bold  $\mathbf{m}$  subscript). This is the set of  $M$ -variate polynomials  $P(\mathbf{x}) = \sum_{\mathbf{k} \leq \mathbf{m}} p_{\mathbf{k}} \mathbf{x}^{\mathbf{k}}$ , multiplied by the function  $\mathbf{x} \mapsto e^{-\frac{\mathbf{x}}{2}}$ . Let  $\mathcal{R}_{\mathbf{m},+}(\mathbb{R}_+^M)$  denotes its subset of nonnegative elements where the polynomial  $P$  is such that  $\mathbf{x} \mapsto P(\mathbf{x}^2)$  has a sum-of-squares decomposition.

#### 4.4.2 Multimode Wigner negativity witnesses

The single-mode Wigner negativity witnesses defined in Eq. (4.1) are naturally generalised to

$$\hat{\Omega}_{\mathbf{a}, \boldsymbol{\alpha}} := \sum_{1 \leq \mathbf{k} \leq \mathbf{n}} a_{\mathbf{k}} \hat{D}(\boldsymbol{\alpha}) |\mathbf{k}\rangle \langle \mathbf{k}| \hat{D}^\dagger(\boldsymbol{\alpha}), \quad (4.140)$$

for  $\mathbf{n} = (n_1, \dots, n_M) \in \mathbb{N}^M \setminus \{\mathbf{0}\}$ ,  $\mathbf{a} = (a_{\mathbf{k}})_{1 \leq \mathbf{k} \leq \mathbf{n}} \in [0, 1]^{n_1 \cdots n_M}$ , with  $\max_{\mathbf{k}} a_{\mathbf{k}} = 1$ , and  $\boldsymbol{\alpha} \in \mathbb{C}^M$ . Similar to the single-mode case, these POVM elements are weighted sums of multimode displaced Fock states projectors, and their expectation value for a quantum state  $\rho \in \mathcal{D}(\mathcal{H}^{\otimes M})$  is given by

$$\text{Tr}(\hat{\Omega}_{\mathbf{a}, \boldsymbol{\alpha}} \rho) = \sum_{1 \leq \mathbf{k} \leq \mathbf{n}} a_{\mathbf{k}} F\left(\hat{D}^\dagger(\boldsymbol{\alpha}) \rho \hat{D}(\boldsymbol{\alpha}), |\mathbf{k}\rangle\right) \quad (4.141)$$

where  $F$  is the fidelity. However, unlike in the single-mode case however, estimating this quantity with homodyne or heterodyne detection by direct fidelity estimation is no longer

efficient when the number of modes becomes large. Instead, one may use *robust lower bounds on the multimode fidelity* from [CGKM20] which can be obtained efficiently with homodyne or heterodyne detection. A lower bound on the estimated experimental multimode fidelity will allow to detect Wigner negativity if it is larger than an upper bound on the threshold value associated to a given witness.

These lower bounds are obtained as follows: given a target multimode Fock state  $|\mathbf{n}\rangle = |n_1\rangle \otimes \cdots \otimes |n_M\rangle$  and multiple copies of an  $M$ -mode experimental state  $\rho$ , measure all single-mode subsystems of  $\rho$  and perform fidelity estimation with each corresponding single-mode target Fock state. That is, the samples obtained from the detection of the  $i^{\text{th}}$  mode of  $\rho$  are used for single-mode fidelity estimation with the Fock state  $|n_i\rangle$ . Let  $F_1, \dots, F_M$  be the single-mode fidelity estimates obtained and let  $\tilde{F}(\rho, |\mathbf{n}\rangle) := 1 - \sum_{i=1}^M (1 - F_i)$ . Then [CGKM20],

$$1 - M(1 - F(\rho, |\mathbf{n}\rangle)) \leq \tilde{F}(\rho, |\mathbf{n}\rangle) \leq F(\rho, |\mathbf{n}\rangle). \quad (4.142)$$

In particular,  $\tilde{F}$  provides a good estimate of the multimode fidelity  $F$  whenever  $F$  is not too small. The same procedure is followed in the case of target *displaced* Fock states, with classical translations of the samples in order to account for the displacement parameters.

To each witness  $\hat{\Omega}_{\mathbf{a}, \alpha}$  is associated its threshold value:

$$\omega_{\mathbf{a}}^9 := \sup_{\substack{\rho \in \mathcal{D}(\mathcal{H}^{\otimes M}) \\ W_{\rho} \geq 0}} \text{Tr}(\hat{\Omega}_{\mathbf{a}, \alpha} \rho). \quad (4.143)$$

With Eq. (4.142), if the value of the bound  $\tilde{F}$  obtained experimentally is greater than  $\omega_{\mathbf{a}}$  (or an upper bound on it) then the multimode state  $\rho$  has a negative Wigner function.

With the same arguments as in the single-mode case, the multimode Wigner negativity witnesses in Eq. (4.140) form a complete family and preserve the interpretation from Lemma 4.1: the violation of the threshold value provides a lower bound on the distance to the set of multimode states with nonnegative Wigner function. However, the limited robustness of the bound  $\tilde{F}$  may affect the performance of the witnesses in practical scenarios, in particular for witnesses that are sums of different projectors. Still, we show in Section 4.4.5 the applicability of the method with a genuinely multimode example.

We first generalise the single-mode semidefinite programming approach for approximating the threshold values to the multimode case.

### 4.4.3 Infinite-dimensional linear programs

By linearity, we restrict our analysis to the case of Wigner negativity witnesses that are projectors onto a single multimode Fock state  $|\mathbf{n}\rangle$ , for  $\mathbf{n} \in \mathbb{N}^M \setminus \{\mathbf{0}\}$ . We thus consider the

<sup>9</sup>Again we write the threshold value as  $\omega_{\mathbf{a}}$  generically and we will use the more precise notation  $\omega_{\mathbf{a}}^{L^2}$  (resp.  $\omega_{\mathbf{a}}^S$ ) to refer to the optimisation over square integrable functions (resp. Schwartz functions) on  $\mathbb{R}^M$ .

computation of

$$\omega_{\mathbf{n}} = \sup_{\substack{\rho \in \mathcal{D}(\mathcal{H}^{\otimes M}) \\ W_{\rho} \geq 0}} \langle \mathbf{n} | \rho | \mathbf{n} \rangle. \quad (4.144)$$

A similar reasoning to the single-mode case shows that the computation of the corresponding threshold value in Eq. (4.143) may be rephrased as the following infinite-dimensional linear program:

$$(\text{LP}_{\mathbf{n}}^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } (F_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^M} \in \ell^2(\mathbb{N}^M) \\ \text{maximising } F_{\mathbf{n}} \\ \text{subject to} \\ \sum_{\mathbf{k}} F_{\mathbf{k}} = 1 \\ \forall \mathbf{k} \in \mathbb{N}^M, F_{\mathbf{k}} \geq 0 \\ \forall \mathbf{x} \in \mathbb{R}_+^M, \sum_{\mathbf{k}} F_{\mathbf{k}} \mathcal{L}_{\mathbf{k}}(\mathbf{x}) \geq 0, \end{array} \right.$$

where the optimisation is over square-summable real sequences indexed by elements of  $\mathbb{N}^M$ —or equivalently square integrable functions over  $\mathbb{R}_+^M$ . Its dual linear program reads

$$(\text{D-LP}_{\mathbf{n}}^{L^2}) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } \mu \in L^2(\mathbb{R}_+^M) \\ \text{minimising } y \\ \text{subject to} \\ \forall \mathbf{k} \neq \mathbf{n} \in \mathbb{N}^M, y \geq \int_{\mathbb{R}_+^M} \mathcal{L}_{\mathbf{k}} d\mu \\ y \geq 1 + \int_{\mathbb{R}_+^M} \mathcal{L}_{\mathbf{n}} d\mu \\ \forall f \in L^2(\mathbb{R}_+^M), \langle \mu, f \rangle \geq 0. \end{array} \right.$$

We denote their optimal value  $\omega_{\mathbf{n}}^{L^2}$ —as in the single mode case, we have strong duality which can be proven using the same proof with standard infinite-dimensional optimisation techniques as in Subsection 4.3.3. This will also be obtained from the convergence of the upper bounding hierarchy of semidefinite programs. We also introduce the programs over Schwartz functions:

$$(\text{LP}_{\mathbf{n}}^{\mathcal{S}}) \quad \left\{ \begin{array}{l} \text{Find } (F_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^M} \in \mathcal{S}(\mathbb{N}^M) \\ \text{maximising } F_{\mathbf{n}} \\ \text{subject to} \\ \sum_{\mathbf{k}} F_{\mathbf{k}} = 1 \\ \forall \mathbf{k} \in \mathbb{N}^M, F_{\mathbf{k}} \geq 0 \\ \forall \mathbf{x} \in \mathbb{R}_+^M, \sum_{\mathbf{k}} F_{\mathbf{k}} \mathcal{L}_{\mathbf{k}}(\mathbf{x}) \geq 0, \end{array} \right.$$

and

$$(D-LP_n^S) \quad \left\{ \begin{array}{l} \text{Find } y \in \mathbb{R} \text{ and } \mu \in \mathcal{S}'(\mathbb{R}_+^M) \\ \text{minimising } y \\ \text{subject to} \\ \forall \mathbf{k} \neq \mathbf{n} \in \mathbb{N}^M, y \geq \int_{\mathbb{R}_+^M} \mathcal{L}_{\mathbf{k}} d\mu \\ y \geq 1 + \int_{\mathbb{R}_+^M} \mathcal{L}_{\mathbf{n}} d\mu \\ \forall f \in L_+^2(\mathbb{R}_+^M), \langle \mu, f \rangle \geq 0. \end{array} \right.$$

We denote their optimal value  $\omega_n^S$ .

We will then derive two hierarchies of semidefinite programs, lower bounding and upper bounding the threshold value. While our proof of convergence of the single-mode hierarchy of upper bounds towards  $\omega_n^{L^2}$  transfers easily to the multimode setting, the proof of convergence of the hierarchy of lower bounds towards  $\omega_n^S$  requires the analytical expression of feasible solutions for each level of the hierarchy. We show how to construct such solutions in the multimode case using products of single-mode feasible solutions—this requires introducing an equivalent hierarchy of restrictions, where constraints are expressed on polynomials of  $M$  variables with the degree in each individual variable being less or equal to  $m$ , rather than on polynomials of degree  $m$  (that is, constraints of the form  $\mathbf{k} \leq m\mathbf{1}$  where  $m\mathbf{1} = (m, \dots, m) \in \mathbb{N}^M$  rather than  $|\mathbf{k}| \leq m$ ). Along the way, we also prove strong duality of the programs involved. This is the subject of the next subsections.

#### 4.4.3.1 Approximating the threshold value

Hereafter, we state without proofs the technical results used to derive the semidefinite programs and their dual programs. They are straightforward generalisation from their single-mode version by using multi-index notations and will be given without proofs.

As mentioned there are two natural ways to obtain relaxations and restrictions by replacing constraints on nonnegative functions by constraints on nonnegative polynomials: either by considering polynomials  $P(\mathbf{x}) = \sum_{|\mathbf{k}| \leq m} p_{\mathbf{k}} \mathbf{x}^{\mathbf{k}}$  of degree at most  $m$  for  $m \in \mathbb{N}$  ('triangle' procedure), or by considering polynomials  $P(\mathbf{x}) = \sum_{\mathbf{k} \leq \mathbf{m}} p_{\mathbf{k}} \mathbf{x}^{\mathbf{k}}$  with monomials smaller than  $\mathbf{m}$  for  $\mathbf{m} \in \mathbb{N}^M$  ('rectangle' procedure). Note that when  $M = 1$  these two are equal.

Crucially, all the results below containing conditions of the form  $|\mathbf{k}| \leq m$  for  $m \in \mathbb{N}$  are also valid when replaced by conditions of the form  $\mathbf{k} \leq \mathbf{m}$  for  $\mathbf{m} = (m_1, \dots, m_M) \in \mathbb{N}^M$ , with the same proofs, by replacing  $s(m)$  by  $\pi_{\mathbf{m}} = \prod_{i=1}^M (m_i + 1)$ .

**Lemma 4.11** (Equivalent of Lemma 4.5). *Let  $p \in \mathbb{N}$  and let  $P$  be a multivariate polynomial of degree  $2p$ . Let  $\mathbf{v}(\mathbf{x}) = (\mathbf{x}^{\mathbf{k}})_{|\mathbf{k}| \leq p}$  be the vector of monomials. Then,  $P$  has a sum-of-squares*

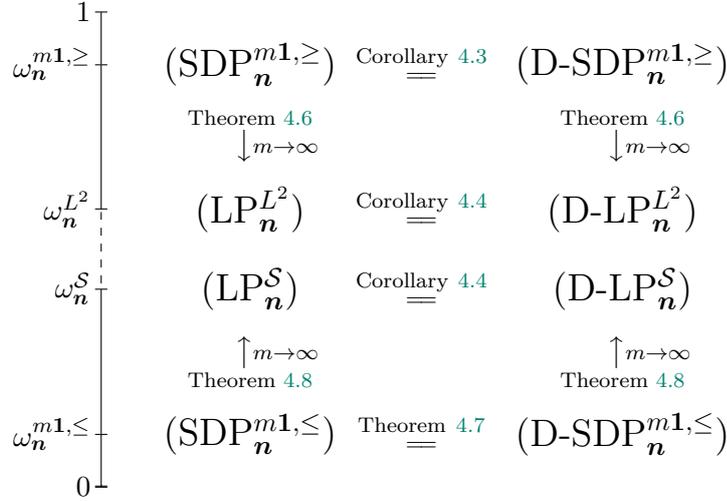


Figure 4.6: Multimode hierarchies of semidefinite relaxations and restrictions converging to the linear program  $(\text{LP}_n^{L^2})$ , together with their dual programs. The upper index  $m$  denotes the level of the relaxation or restriction. On the left are the associated optimal values. The equal sign denotes strong duality, i.e., equality of optimal values, and the arrows denote convergence of the corresponding sequences of optimal values. The hierarchies  $(\text{SDP}_n^{m, \ge})$  and  $(\text{SDP}_n^{m, \le})$  in the main text are different from the ones appearing in the figure, but equivalent by Lemma 4.14. The question of whether  $\omega_n^{L^2} = \omega_n^S$  is left open.

decomposition if and only if there exists a real  $s(p) \times s(p)$  positive semidefinite matrix  $Q$  such that for all  $x \in \mathbb{R}^M$ ,

$$P(x) = \mathbf{v}(x)^T Q \mathbf{v}(x). \quad (4.145)$$

**Lemma 4.12** (Generalisation of Lemma 4.6). *Let  $P$  be a nonnegative polynomial over  $\mathbb{R}_+^M$  such that  $x \mapsto P(x^2)$  has a sum-of-squares decomposition. Then,  $P$  can be written as a sum of polynomials of the form  $\sum_{|i| \leq p} x^i \sum_{i+j=2l} y_i y_j$  for  $p \in \mathbb{N}$  and  $y_i \in \mathbb{R}$  for all  $i \in \mathbb{N}^M$  such that  $|i| \leq p$ .*

**Lemma 4.13** (Generalisation of Lemma 4.7). *Let  $m \in \mathbb{N}$  and let  $\mathbf{s} = (s_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^M} \in \mathbb{R}^{\mathbb{N}^M}$ . The following propositions are equivalent:*

- (i)  $\forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+^M), \langle f_{\mathbf{s}}, g \rangle \geq 0,$
- (ii)  $A_{\mathbf{s}} \succeq 0.$

In the single-mode case, we obtained hierarchies of SDP relaxations and restrictions for  $(\text{LP}_n^{L^2})$  by replacing constraints involving nonnegative functions by constraints involving nonnegative polynomials  $P$  of fixed degree. We then exploited the existence of a sum-of-squares decomposition for nonnegative monovariate polynomials. In the multimode setting, the polynomials involved are multivariate, so that the set of nonnegative polynomials over  $\mathbb{R}$  of a given

degree may be strictly larger than the set of sum-of-square polynomials [Hil88]. Instead, we replace directly constraints involving nonnegative functions over  $\mathbb{R}_+$  by constraints involving nonnegative polynomials  $P$  of fixed degree such that  $\mathbf{x} \mapsto P(\mathbf{x}^2)$ <sup>10</sup> has a sum-of-squares decomposition, implying that the multimode semidefinite relaxations and restrictions are possibly looser than their single-mode counterparts.

Moreover, the dimension of the semidefinite programs increases exponentially with the level of the hierarchy  $m$ , as the number of  $M$ -variate monomials of degree less or equal to  $m$  is given by  $s(m)$ . This implies that the semidefinite programs remain tractable only for a constant number of levels.

In spite of these observations, and following similar steps to the single-mode case (see Subsection 4.3.4), we use Lemma 4.13 to obtain the following semidefinite *relaxations* providing upperbounds on the threshold value:

$$\left( \text{SDP}_{\mathbf{n}}^{m, \geq} \right) \left\{ \begin{array}{l} \text{Find } A = (A_{ij})_{|i|, |j| \leq m} \in \text{Sym}_{s(m)} \text{ and } \mathbf{F} = (F_{\mathbf{k}})_{|\mathbf{k}| \leq m} \in \mathbb{R}^{s(m)} \\ \text{maximising } F_{\mathbf{n}} \\ \text{subject to} \\ \sum_{|\mathbf{k}| \leq m} F_{\mathbf{k}} = 1 \\ \forall |\mathbf{k}| \leq m, F_{\mathbf{k}} \geq 0 \\ \forall |\mathbf{l}| \leq m, \forall \mathbf{i} + \mathbf{j} = 2\mathbf{l}, A_{ij} = \sum_{\mathbf{k} \leq \mathbf{l}} F_{\mathbf{k}} \binom{\mathbf{l}}{\mathbf{k}} \mathbf{l}! \\ \forall |\mathbf{r}| \leq 2m, \mathbf{r} \neq 2\mathbf{l}, \forall |\mathbf{l}| \leq m, \forall \mathbf{i} + \mathbf{j} = \mathbf{r}, A_{ij} = 0 \\ A \succeq 0, \end{array} \right.$$

for all  $m \geq |\mathbf{n}|$ . We denote its optimal value by  $\omega_{\mathbf{n}}^{m, \geq}$ . The corresponding dual programs are given by:

$$\left( \text{D-SDP}_{\mathbf{n}}^{m, \geq} \right) \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{s(m)}, y \in \mathbb{R} \text{ and } \boldsymbol{\mu} \in \mathbb{R}^{s(m)} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_{\mathbf{n}} \\ \forall |\mathbf{k}| \leq m, \mathbf{k} \neq \mathbf{n}, y \geq \mu_{\mathbf{k}} \\ \forall |\mathbf{l}| \leq m, \sum_{\mathbf{i} + \mathbf{j} = 2\mathbf{l}} Q_{ij} = \sum_{\mathbf{k} \geq \mathbf{l}} \frac{(-1)^{|\mathbf{k}| + |\mathbf{l}|}}{\mathbf{l}!} \binom{\mathbf{k}}{\mathbf{l}} \mu_{\mathbf{k}} \\ Q \succeq 0. \end{array} \right.$$

Similarly, using Lemma 4.11, the semidefinite restrictions providing lower bounds for the

<sup>10</sup>We write  $\mathbf{x}^2$  in short for  $\mathbf{x}^{2\mathbf{1}} = (x_1^2, \dots, x_M^2)$ .

threshold value are given by

$$(\text{SDP}_{\mathbf{n}}^{m, \leq}) \quad \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{s(m)} \text{ and } \mathbf{F} \in \mathbb{R}^{s(m)} \\ \text{maximising } F_{\mathbf{n}} \\ \text{subject to} \\ \sum_{|\mathbf{k}| \leq m} F_{\mathbf{k}} = 1 \\ \forall |\mathbf{k}| \leq m, F_{\mathbf{k}} \geq 0 \\ \forall |\mathbf{l}| \leq m, \sum_{i+j=2\mathbf{l}} Q_{ij} = \sum_{\mathbf{k} \geq \mathbf{l}} \frac{(-1)^{|\mathbf{k}|+|\mathbf{l}|}}{\mathbf{l}!} \binom{\mathbf{k}}{\mathbf{l}} F_{\mathbf{k}} \\ \forall |\mathbf{r}| \leq 2m, \mathbf{r} \neq 2\mathbf{l}, \forall |\mathbf{l}| \leq m, \sum_{i+j=\mathbf{r}} Q_{ij} = 0 \\ Q \succeq 0, \end{array} \right.$$

for all  $m \geq |\mathbf{n}|$ . We denote its optimal value by  $\omega_{\mathbf{n}}^{m, \leq}$ . The corresponding dual programs are given by:

$$(\text{D-SDP}_{\mathbf{n}}^{m, \leq}) \quad \left\{ \begin{array}{l} \text{Find } A \in \text{Sym}_{s(m)}, y \in \mathbb{R} \text{ and } \boldsymbol{\mu} \in \mathbb{R}^{s(m)} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_{\mathbf{n}} \\ \forall |\mathbf{k}| \leq m, \mathbf{k} \neq \mathbf{n}, \quad y \geq \mu_{\mathbf{k}} \\ \forall |\mathbf{l}| \leq m, \forall \mathbf{i} + \mathbf{j} = 2\mathbf{l}, \quad A_{ij} = \sum_{\mathbf{k} \leq \mathbf{l}} \mu_{\mathbf{k}} \binom{\mathbf{l}}{\mathbf{k}} \mathbf{l}! \\ A \succeq 0, \end{array} \right.$$

for all  $m \geq |\mathbf{n}|$ . Like in the single-mode case, note that without loss of generality the condition  $y \leq 1$ , and thus  $\mu_{\mathbf{k}} \leq 1$  for all  $\mathbf{k}$ , can be added to the optimisation, since setting  $A = 0$ ,  $y = 1$  and  $\boldsymbol{\mu} = 0$  gives a feasible solution with objective value 1.

These are the relaxations and restrictions of  $(\text{LP}_{\mathbf{n}}^{L^2})$  obtained by considering polynomials of degree less or equal to  $m$ , where the optimisation is over matrices and vectors indexed by elements of  $\mathbb{N}^m$  with sum of coefficients lower than  $m$ . Alternatively, we may also consider the relaxations and restrictions obtained by considering polynomials with monomials smaller than  $\mathbf{m} \in \mathbb{N}^M$ , where the optimisation is over matrices and vectors indexed by elements of  $\mathbb{N}^m$  lower or equal to  $\mathbf{m}$ . Recalling the notation  $\pi_{\mathbf{m}} = \prod_{i=1}^M (m_i + 1)$ , the corresponding semidefinite

relaxations are given by

$$(\text{SDP}_n^{m, \geq}) \left\{ \begin{array}{l} \text{Find } A \in \text{Sym}_{\pi_m} \text{ and } \mathbf{F} \in \mathbb{R}^{\pi_m} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{\mathbf{k} \leq m} F_{\mathbf{k}} = 1 \\ \forall \mathbf{k} \leq m, \quad F_{\mathbf{k}} \geq 0 \\ \forall l \leq m, \forall \mathbf{i} + \mathbf{j} = 2l, \quad A_{ij} = \sum_{\mathbf{k} \leq l} F_{\mathbf{k}} \binom{l}{\mathbf{k}} l! \\ \forall r \leq 2m, r \neq 2l, \forall l \leq m, \forall \mathbf{i} + \mathbf{j} = r, \quad A_{ij} = 0 \\ A \succeq 0. \end{array} \right.$$

for all  $m \geq n$ . We denote its optimal value by  $\omega_n^{m, \geq}$ . The corresponding dual programs are given by:

$$(\text{D-SDP}_n^{m, \geq}) \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{\pi_m}, y \in \mathbb{R} \text{ and } \boldsymbol{\mu} \in \mathbb{R}^{\pi_m} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_n \\ \forall \mathbf{k} \leq m, \mathbf{k} \neq \mathbf{n}, \quad y \geq \mu_{\mathbf{k}} \\ \forall l \leq m, \quad \sum_{\mathbf{i} + \mathbf{j} = 2l} Q_{ij} = \sum_{\mathbf{k} \geq l} \frac{(-1)^{|\mathbf{k}| + |\mathbf{l}|}}{l!} \binom{\mathbf{k}}{\mathbf{l}} \mu_{\mathbf{k}} \\ Q \succeq 0, \end{array} \right.$$

for all  $m \geq n$ . Similarly, the semidefinite *restrictions* are given by:

$$(\text{SDP}_n^{m, \leq}) \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{\pi_m} \text{ and } \mathbf{F} \in \mathbb{R}^{\pi_m} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{\mathbf{k} \leq m} F_{\mathbf{k}} = 1 \\ \forall \mathbf{k} \leq m, \quad F_{\mathbf{k}} \geq 0 \\ \forall l \leq m, \quad \sum_{\mathbf{i} + \mathbf{j} = 2l} Q_{ij} = \sum_{\mathbf{k} \geq l} \frac{(-1)^{|\mathbf{k}| + |\mathbf{l}|}}{l!} \binom{\mathbf{k}}{\mathbf{l}} F_{\mathbf{k}} \\ \forall r \leq 2m, r \neq 2l, \forall l \leq m, \quad \sum_{\mathbf{i} + \mathbf{j} = r} Q_{ij} = 0 \\ Q \succeq 0, \end{array} \right.$$

for all  $\mathbf{m} \geq \mathbf{n}$ . We denote its optimal value by  $\omega_n^{m, \leq}$ . The corresponding dual programs are given by:

$$(\text{D-SDP}_n^{m, \leq}) \left\{ \begin{array}{l} \text{Find } A \in \text{Sym}_{\pi_m}, y \in \mathbb{R} \text{ and } \boldsymbol{\mu} \in \mathbb{R}^{\pi_m} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_n \\ \forall \mathbf{k} \leq \mathbf{m}, \mathbf{k} \neq \mathbf{n}, \quad y \geq \mu_{\mathbf{k}} \\ \forall \mathbf{l} \leq \mathbf{m}, \forall \mathbf{i} + \mathbf{j} = 2\mathbf{l}, \quad A_{ij} = \sum_{\mathbf{k} \leq \mathbf{l}} \mu_{\mathbf{k}} \binom{\mathbf{l}}{\mathbf{k}} l! \\ A \succeq 0, \end{array} \right.$$

for all  $\mathbf{m} \geq \mathbf{n}$ .

The programs  $(\text{SDP}_n^{m, \geq})$  and  $(\text{SDP}_n^{m, \leq})$  respectively provide hierarchies of relaxations and restrictions of  $(\text{LP}_n^{L^2})$ , since the set of  $M$ -variate polynomials of degree  $m$  is included in the set of  $M$ -variate polynomials of degree  $m+1$ . On the other hand, there is no natural ordering in  $\mathbb{N}^M$  of the relaxations  $(\text{SDP}_n^{m, \geq})$  or the restrictions  $(\text{SDP}_n^{m, \leq})$  (consider for instance  $\mathbf{m} = (2, 1)$  and  $\mathbf{m}' = (1, 2)$ ). In order to obtain proper hierarchies of semidefinite programs, we thus consider the subset of these programs where the tuple  $\mathbf{m}$  is of the form  $m\mathbf{1} = (m, \dots, m) \in \mathbb{N}^M$ , for  $m \in \mathbb{N}$ . We have  $\pi_{m\mathbf{1}} = (m+1)^M$ , and the *relaxations* are then given by

$$(\text{SDP}_n^{m\mathbf{1}, \geq}) \left\{ \begin{array}{l} \text{Find } A \in \text{Sym}_{(m+1)^M} \text{ and } \mathbf{F} \in \mathbb{R}^{(m+1)^M} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{\mathbf{k} \leq m\mathbf{1}} F_{\mathbf{k}} = 1 \\ \forall \mathbf{k} \leq m\mathbf{1}, \quad F_{\mathbf{k}} \geq 0 \\ \forall \mathbf{l} \leq m\mathbf{1}, \forall \mathbf{i} + \mathbf{j} = 2\mathbf{l}, \quad A_{ij} = \sum_{\mathbf{k} \leq \mathbf{l}} F_{\mathbf{k}} \binom{\mathbf{l}}{\mathbf{k}} l! \\ \forall \mathbf{r} \leq 2m\mathbf{1}, \mathbf{r} \neq 2\mathbf{l}, \forall \mathbf{l} \leq m\mathbf{1}, \forall \mathbf{i} + \mathbf{j} = \mathbf{r}, \quad A_{ij} = 0 \\ A \succeq 0. \end{array} \right.$$

for  $m \geq \max_i n_i$ . We denote its optimal value by  $\omega_n^{m\mathbf{1}, \geq}$ . The corresponding dual programs are

given by:

$$\left( \text{D-SDP}_n^{m\mathbf{1}, \geq} \right) \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{(m+1)^M}, y \in \mathbb{R} \text{ and } \boldsymbol{\mu} \in \mathbb{R}^{(m+1)^M} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_n \\ \forall \mathbf{k} \leq m\mathbf{1}, \mathbf{k} \neq \mathbf{n}, \quad y \geq \mu_{\mathbf{k}} \\ \forall \mathbf{l} \leq m\mathbf{1}, \quad \sum_{i+j=2\mathbf{l}} Q_{ij} = \sum_{\mathbf{k} \geq \mathbf{l}} \frac{(-1)^{|\mathbf{k}|+|\mathbf{l}|}}{\mathbf{l}!} \binom{\mathbf{k}}{\mathbf{l}} \mu_{\mathbf{k}} \\ Q \succeq 0, \end{array} \right.$$

for  $m \geq \max_i n_i$ . Similarly, the *restrictions* are given by:

$$\left( \text{SDP}_n^{m\mathbf{1}, \leq} \right) \left\{ \begin{array}{l} \text{Find } Q \in \text{Sym}_{(m+1)^M} \text{ and } \mathbf{F} \in \mathbb{R}^{(m+1)^M} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{\mathbf{k} \leq m\mathbf{1}} F_{\mathbf{k}} = 1 \\ \forall \mathbf{k} \leq m\mathbf{1}, \quad F_{\mathbf{k}} \geq 0 \\ \forall \mathbf{l} \leq m\mathbf{1}, \quad \sum_{i+j=2\mathbf{l}} Q_{ij} = \sum_{\mathbf{k} \geq \mathbf{l}} \frac{(-1)^{|\mathbf{k}|+|\mathbf{l}|}}{\mathbf{l}!} \binom{\mathbf{k}}{\mathbf{l}} F_{\mathbf{k}} \\ \forall \mathbf{r} \leq 2m\mathbf{1}, \mathbf{r} \neq 2\mathbf{l}, \forall \mathbf{l} \leq m\mathbf{1}, \quad \sum_{i+j=\mathbf{r}} Q_{ij} = 0 \\ Q \succeq 0, \end{array} \right.$$

for  $m \geq \max_i n_i$ . We denote its optimal value by  $\omega_n^{m\mathbf{1}, \leq}$ . The corresponding dual programs are given by:

$$\left( \text{D-SDP}_n^{m\mathbf{1}, \leq} \right) \left\{ \begin{array}{l} \text{Find } A \in \text{Sym}_{(m+1)^M}, y \in \mathbb{R} \text{ and } \boldsymbol{\mu} \in \mathbb{R}^{(m+1)^M} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_n \\ \forall \mathbf{k} \leq m\mathbf{1}, \mathbf{k} \neq \mathbf{n}, \quad y \geq \mu_{\mathbf{k}} \\ \forall \mathbf{l} \leq m\mathbf{1}, \forall \mathbf{i} + \mathbf{j} = 2\mathbf{l}, \quad A_{ij} = \sum_{\mathbf{k} \leq \mathbf{l}} \mu_{\mathbf{k}} \binom{\mathbf{l}}{\mathbf{k}} \mathbf{l}! \\ A \succeq 0, \end{array} \right.$$

for  $m \geq \max_i n_i$ .

The programs  $(\text{SDP}_n^{m\mathbf{1},\geq})$  and  $(\text{SDP}_n^{m\mathbf{1},\leq})$  are respectively relaxations and restrictions of  $(\text{LP}_n^{L^2})$  obtained by considering polynomials of individual degree in each variable less or equal to  $m$ . These programs respectively provide hierarchies of relaxations and restrictions of  $(\text{LP}_n^{L^2})$ , since the set of  $M$ -variate polynomials with monomials lower than  $m\mathbf{1}$  is included in the set of  $M$ -variate polynomials with monomials lower than  $(m+1)\mathbf{1}$ .

Note that these hierarchies of programs obtained by setting  $\mathbf{m}$  of the form  $m\mathbf{1}$  capture the behaviour of all bounds that can be obtained from the more general family of programs indexed by  $\mathbf{m} = (m_1, \dots, m_M)$ , since  $\mathbf{m} \leq (\max_i m_i)\mathbf{1}$ , i.e., the bound obtained by considering the program indexed by  $(\max_i m_i)\mathbf{1}$  supersedes the bound obtained by considering the program indexed by  $\mathbf{m}$ . Formally, for all  $\mathbf{m} = (m_1, \dots, m_M) \in \mathbb{N}^M$ ,

$$\omega_n^{\mathbf{m},\geq} \geq \omega_n^{(\max_i m_i)\mathbf{1},\geq} \quad \text{and} \quad \omega_n^{\mathbf{m},\leq} \leq \omega_n^{(\max_i m_i)\mathbf{1},\leq}. \quad (4.146)$$

Finally, we show that both ways of defining the hierarchies—i.e. with  $m \in \mathbb{N}$  corresponding to programs  $(\text{SDP}_n^{m,\leq})$  and  $(\text{SDP}_n^{m,\geq})$  and with  $m\mathbf{1} \in \mathbb{N}^M$  corresponding to programs  $(\text{SDP}_n^{m\mathbf{1},\leq})$  and  $(\text{SDP}_n^{m\mathbf{1},\geq})$ —are equivalent:

**Lemma 4.14.** *For all  $m \in \mathbb{N}$ ,*

$$\omega_n^{m,\geq} \geq \omega_n^{m\mathbf{1},\geq} \geq \omega_n^{Mm,\geq}, \quad (4.147)$$

and

$$\omega_n^{m,\leq} \leq \omega_n^{m\mathbf{1},\leq} \leq \omega_n^{Mm,\leq}. \quad (4.148)$$

*Proof.* For all  $m \in \mathbb{N}$  we have (for  $M > 1$ ):

$$\{\mathbf{k} \in \mathbb{N}^M : |\mathbf{k}| \leq m\} \subset \{\mathbf{k} \in \mathbb{N}^M : \mathbf{k} \leq m\mathbf{1}\} \subset \{\mathbf{k} \in \mathbb{N}^M : |\mathbf{k}| \leq Mm\}. \quad (4.149)$$

Geometrically We thus obtain the corresponding inclusions between sets of  $M$ -variate polynomials: (i)  $M$ -variate polynomials of degree less or equal to  $m$  have all their monomials lower than  $m\mathbf{1}$ , and (ii) all  $M$ -variate polynomials with monomials lower than  $m\mathbf{1}$  have degree less or equal to  $Mm$ . Hence,

$$\mathcal{R}_{m,+}(\mathbb{R}_+^M) \stackrel{(i)}{\subset} \mathcal{R}_{m\mathbf{1},+}(\mathbb{R}_+^M) \stackrel{(ii)}{\subset} \mathcal{R}_{Mm,+}(\mathbb{R}_+^M). \quad (4.150)$$

As a consequence,  $(\text{SDP}_n^{m,\geq})$  is a relaxation of  $(\text{SDP}_n^{m\mathbf{1},\geq})$  which is itself a relaxation of  $(\text{SDP}_n^{Mm,\geq})$ , and  $(\text{SDP}_n^{m,\leq})$  is a restriction of  $(\text{SDP}_n^{m\mathbf{1},\leq})$  which is itself a restriction of  $(\text{SDP}_n^{Mm,\leq})$ .  $\square$

This result implies that the two versions of the hierarchies of relaxations are interleaved (Eq. (4.147)), and that the two versions of the hierarchies of restrictions are also interleaved (Eq. (4.148)). As such, for any bound obtained with one version of the hierarchy at some fixed level, a better bound can be obtained with the other version at some other level. While this means that the hierarchies are equivalent, note that in practice it may be simpler to solve numerically the version where the parameter space is smaller.

#### 4.4.4 Convergence of the multimode hierarchies

For  $\mathbf{n} = (n_1, \dots, n_M) \in \mathbb{N}^M$ , the sequences  $(\omega_{\mathbf{n}}^{m, \geq})_{m \geq |\mathbf{n}|}$  and  $(\omega_{\mathbf{n}}^{m1, \geq})_{m \geq \max_i n_i}$  (resp. the sequences  $(\omega_{\mathbf{n}}^{m, \leq})_{m \geq |\mathbf{n}|}$  and  $(\omega_{\mathbf{n}}^{m1, \leq})_{m \geq \max_i n_i}$ ) are decreasing (resp. increasing) sequences, lower bounded (resp. upper bounded) by  $\omega_{\mathbf{n}}^{L^2}$  (resp.  $\omega_{\mathbf{n}}^S$ ). Hence, these sequences are converging.

We show in what follows that  $(\omega_{\mathbf{n}}^{m1, \geq})_{m \geq \max_i n_i}$  (resp.  $(\omega_{\mathbf{n}}^{m1, \leq})_{m \geq \max_i n_i}$ ) converges to  $\omega_{\mathbf{n}}^{L^2}$  (resp.  $\omega_{\mathbf{n}}^S$ ). By Lemma 4.14, this implies that the sequence  $(\omega_{\mathbf{n}}^{Mm, \geq})_{m \geq \max_i n_i}$  (resp.  $(\omega_{\mathbf{n}}^{Mm, \leq})_{m \geq \max_i n_i}$ ) also converges to  $\omega_{\mathbf{n}}$ . Since this is a subsequence of the converging sequence  $(\omega_{\mathbf{n}}^{m, \geq})_{m \geq |\mathbf{n}|}$  (resp.  $(\omega_{\mathbf{n}}^{m, \leq})_{m \geq |\mathbf{n}|}$ ), it implies that the sequence  $(\omega_{\mathbf{n}}^{m, \geq})_{m \geq |\mathbf{n}|}$  (resp.  $(\omega_{\mathbf{n}}^{m, \leq})_{m \geq |\mathbf{n}|}$ ) also converges to  $\omega_{\mathbf{n}}^{L^2}$  (resp.  $\omega_{\mathbf{n}}^S$ ).

With similar proofs to the single-mode case using multi-index notations, we obtain the following result:

**Theorem 4.5** (Generalisation of Theorem 4.1). *Let  $\boldsymbol{\mu} = (\mu_{\mathbf{k}})_{\mathbf{k} \in \mathbb{N}^M} \in \mathbb{R}^{\mathbb{N}^M}$ . Then,  $\boldsymbol{\mu}$  is the sequence of Laguerre moments  $\int_{\mathbb{R}_+^M} \mathcal{L}_{\mathbf{k}}(\mathbf{x}) d\mu(\mathbf{x})$  of a nonnegative distribution  $\mu$  supported on  $\mathbb{R}_+^M$  if and only if  $\forall m \in \mathbb{N}, \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+^M), \langle f_{\boldsymbol{\mu}}, g \rangle \geq 0$ .*

The proof of this theorem is identical to the univariate case, with the use of Riesz–Haviland theorem over  $\mathbb{R}_+^M$  [Hav36] rather than  $\mathbb{R}_+$ .

With Eq. (4.150), the proof of convergence of the multimode hierarchy of upper bounds is then obtained directly from its single-mode counterpart using multi-index notations:

**Theorem 4.6** (Generalisation of Theorem 4.3). *The decreasing sequence of optimal values  $\omega_{\mathbf{n}}^{m1, \geq}$  of  $(\text{SDP}_{\mathbf{n}}^{m1, \geq})$  converges to the optimal value  $\omega_{\mathbf{n}}^{L^2}$  of  $(\text{LP}_{\mathbf{n}}^{L^2})$ :*

$$\lim_{m \rightarrow +\infty} \omega_{\mathbf{n}}^{m1, \geq} = \omega_{\mathbf{n}}^{L^2}. \quad (4.151)$$

With Lemma 4.14, we also obtain

$$\lim_{m \rightarrow +\infty} \omega_{\mathbf{n}}^{m1, \leq} = \omega_{\mathbf{n}}^S. \quad (4.152)$$

On the other hand, the proof of convergence of the single-mode hierarchy of lower bounds crucially exploits analytical feasible solutions of the programs  $(\text{SDP}_{\mathbf{n}}^{m, \leq})$  in order to obtain two results:

- Strong duality between programs  $(\text{SDP}_{\mathbf{n}}^{m, \leq})$  and  $(\text{D-SDP}_{\mathbf{n}}^{m, \leq})$  (Theorem 4.2).
- The fact that the feasible set of  $(\text{D-SDP}_{\mathbf{n}}^{m, \leq})$  is compact with coefficients bounded independently of  $m$  (Eq. (4.128)) which exploits an analytical feasible solution of the primal program.

In what follows, we generalise these two results to the multimode setting by obtaining multimode analytical feasible solutions from products of single-mode ones.

**Lemma 4.15.** For all  $m, n \in \mathbb{N}$  with  $m \geq n$ , suppose that  $Q(m, n) \in \text{Sym}_{m+1}$  and  $\mathbf{F}(m, n) = (F_i(m, n))_{0 \leq i \leq m} \in \mathbb{R}^{m+1}$  are feasible solutions of  $(\text{SDP}_n^{m, \leq})$ . Let  $\mathbf{m} = (m_1, \dots, m_M) \in \mathbb{N}^M$  and  $\mathbf{n} = (n_1, \dots, n_M) \in \mathbb{N}^M$  with  $\mathbf{m} \geq \mathbf{n}$ . Define  $Q := Q(m_1, n_1) \otimes \dots \otimes Q(m_M, n_M) \in \text{Sym}_{\pi_{\mathbf{m}}}$  and  $\mathbf{F} = (F_{\mathbf{k}})_{\mathbf{k} \leq \mathbf{m}} \in \mathbb{R}^{\pi_{\mathbf{m}}}$ , where for all  $\mathbf{k} = (k_1, \dots, k_M) \leq \mathbf{m}$ ,  $F_{\mathbf{k}} := \prod_{i=1}^M F_{k_i}(m_i, n_i)$ .

Then,  $(Q, \mathbf{F})$  is a feasible solution of  $(\text{SDP}_n^{m, \leq})$ . Moreover, if  $(Q(m_i, n_i), \mathbf{F}(m_i, n_i))$  is strictly feasible for all  $i = 1, \dots, M$  then  $(Q, \mathbf{F})$  is a strictly feasible solution of  $(\text{SDP}_n^{m, \leq})$ .

*Proof.* With the notations of the Lemma, we show the feasibility of  $(Q, \mathbf{F})$  (resp. strict feasibility). We immediately have  $Q \succeq 0$ ,  $F_{\mathbf{k}} \geq 0$  (resp.  $Q \succ 0$ ,  $F_{\mathbf{k}} > 0$ ) for all  $\mathbf{k} \leq \mathbf{m}$ , and  $Q_{ij} = \prod_{p=1}^M Q_{i_p j_p}(m_p, n_p)$  for all  $\mathbf{i} = (i_1, \dots, i_M) \leq \mathbf{m}$  and  $\mathbf{j} = (j_1, \dots, j_M) \leq \mathbf{m}$ . Hence, for all  $\mathbf{r} = (r_1, \dots, r_M) \leq 2\mathbf{m}$ ,

$$\begin{aligned} \sum_{\mathbf{i}+\mathbf{j}=\mathbf{r}} Q_{ij} &= \sum_{i_1+j_1=r_1, \dots, i_M+j_M=r_M} \prod_{p=1}^M Q_{i_p j_p}(m_p, n_p) \\ &= \prod_{p=1}^M \sum_{i_p+j_p=r_p} Q_{i_p j_p}(m_p, n_p). \end{aligned} \quad (4.153)$$

In particular, if  $r \neq 2l$  for all  $l \leq \mathbf{m}$ , then at least one coefficient  $r_p$  is odd, and the corresponding sum gives 0 since  $(Q(m_p, n_p), \mathbf{F}(m_p, n_p))$  is feasible for  $(\text{SDP}_{n_p}^{m_p, \leq})$ . In that case,  $\sum_{\mathbf{i}+\mathbf{j}=\mathbf{r}} Q_{ij} = 0$ . Otherwise, for all  $\mathbf{l} = (l_1, \dots, l_M) \leq \mathbf{m}$ ,

$$\begin{aligned} \sum_{\mathbf{i}+\mathbf{j}=2\mathbf{l}} Q_{ij} &= \prod_{p=1}^M \sum_{i_p+j_p=2l_p} Q_{i_p j_p}(m_p, n_p) \\ &= \prod_{p=1}^M \sum_{k_p \geq l_p} \frac{(-1)^{k_p+l_p}}{l_p!} \binom{k_p}{l_p} F_{k_p}(m_p, n_p) \\ &= \sum_{l_1 \leq k_1 \leq m_1, \dots, l_M \leq k_M \leq m_M} \prod_{p=1}^M \frac{(-1)^{k_p+l_p}}{l_p!} \binom{k_p}{l_p} F_{k_p}(m_p, n_p) \\ &= \sum_{\mathbf{k} \geq \mathbf{l}} \frac{(-1)^{|\mathbf{k}|+|\mathbf{l}|}}{l!} \binom{\mathbf{k}}{\mathbf{l}} F_{\mathbf{k}}, \end{aligned} \quad (4.154)$$

where we used the feasibility of  $(Q(m_p, n_p), \mathbf{F}(m_p, n_p))$  in the second line. Finally,

$$\begin{aligned} \sum_{\mathbf{k} \leq \mathbf{m}} F_{\mathbf{k}} &= \sum_{k_1 \leq m_1, \dots, k_M \leq m_M} \prod_{i=1}^M F_{k_i}(m_i, n_i) \\ &= \prod_{i=1}^M \sum_{k_i=0}^{m_i} F_{k_i}(m_i, n_i) \\ &= 1, \end{aligned} \quad (4.155)$$

since  $\sum_{k=0}^m F_k(m, n) = 1$  for all  $m, n \in \mathbb{N}$  with  $m \geq n$ . This shows that  $(Q, \mathbf{F})$  is a feasible solution of  $(\text{SDP}_n^{m, \leq})$  (resp. strictly feasible).  $\square$

A direct consequence of this construction is the following result:

**Theorem 4.7** (Generalisation of Theorem 4.2). *Strong duality holds between the programs  $(\text{SDP}_n^{m,\leq})$  and  $(\text{D-SDP}_n^{m,\leq})$ .*

*Proof.* The proof of Theorem 4.2 gives a strictly feasible solution  $(Q(m, n), \mathbf{F}(m, n))$  of  $(\text{SDP}_n^{m,\leq})$  for all  $m \geq n$ . By Lemma 4.15, the program  $(\text{SDP}_n^{m,\leq})$  thus has a strictly feasible solution, for all  $m \geq n$ . By Slater's condition, this implies that strong duality holds between the programs  $(\text{SDP}_n^{m,\leq})$  and  $(\text{D-SDP}_n^{m,\leq})$ .  $\square$

In particular, strong duality holds between the programs  $(\text{SDP}_n^{m1,\leq})$  and  $(\text{D-SDP}_n^{m1,\leq})$ . Note that the multimode generalisation of Corollary 4.1 is a direct consequence of Theorem 4.7:

**Corollary 4.3** (Generalisation of Corollary 4.1). *Strong duality holds between the programs  $(\text{SDP}_n^{m,\geq})$  and  $(\text{D-SDP}_n^{m,\geq})$ .*

*Proof.* the strictly feasible solution of  $(\text{SDP}_n^{m,\leq})$  derived in the proof of Theorem 4.7 yields a strictly feasible solution for  $(\text{SDP}_n^{m,\geq})$ . With Slater's condition, this shows again that strong duality holds between the programs  $(\text{SDP}_n^{m,\geq})$  and  $(\text{D-SDP}_n^{m,\geq})$ .  $\square$

In particular, strong duality holds between the programs  $(\text{SDP}_n^{m1,\geq})$  and  $(\text{D-SDP}_n^{m1,\geq})$ .

We recall the following definition from Subsection 4.3.5: for all  $n \in \mathbb{N}$ ,  $\mathbf{F}^n = (F_k^n)_{k \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  where

- if  $n$  is even:

$$F_k^n := \begin{cases} \frac{1}{2^n} \binom{k}{\frac{k}{2}} \binom{n-k}{\frac{n-k}{2}} & \text{when } k \leq n, k \text{ even,} \\ 0 & \text{otherwise,} \end{cases} \quad (4.156)$$

- if  $n$  is odd:

$$F_k^n := \begin{cases} \frac{1}{2^n} \frac{\binom{n}{\lfloor \frac{n}{2} \rfloor} \binom{\lfloor \frac{n}{2} \rfloor}{\lfloor \frac{k}{2} \rfloor}}{\binom{n}{k}}, & \text{when } k \leq n, \\ 0 & \text{otherwise.} \end{cases} \quad (4.157)$$

Let us define, for all  $\mathbf{n} = (n_1, \dots, n_M) \in \mathbb{N}^M$ ,  $\mathbf{F}^{\mathbf{n}} = (F_{\mathbf{k}}^{\mathbf{n}})_{\mathbf{k} \in \mathbb{N}^M} \in \mathbb{R}^{\mathbb{N}^M}$  where

$$F_{\mathbf{k}}^{\mathbf{n}} := \begin{cases} \prod_{i=1}^M F_{k_i}^{n_i} & \text{when } \mathbf{k} \leq \mathbf{n}, \\ 0 & \text{otherwise.} \end{cases} \quad (4.158)$$

By (4.92), for all  $n \in \mathbb{N}$ ,  $F_n^n \geq \frac{1}{n+1}$ , so for all  $\mathbf{n} = (n_1, \dots, n_M) \in \mathbb{N}^M$ ,

$$F_{\mathbf{n}}^{\mathbf{n}} \geq \frac{1}{\pi_{\mathbf{n}}}. \quad (4.159)$$

Like in the single-mode case, the program  $(\text{SDP}_n^{m,\leq})$  is equivalent to

$$(\text{SDP}_n^{m,\leq}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{F} \in \mathbb{R}^{\pi m} \\ \text{maximising } F_n \\ \text{subject to} \\ \sum_{\mathbf{k} \leq \mathbf{m}} F_{\mathbf{k}} = 1 \\ \forall \mathbf{k} \leq \mathbf{m}, F_{\mathbf{k}} \geq 0 \\ f_{\mathbf{F}} \in \mathcal{R}_{m,+}(\mathbb{R}_+^M), \end{array} \right.$$

with  $f_{\mathbf{F}} = \sum_{\mathbf{k} \leq \mathbf{m}} F_{\mathbf{k}} \mathcal{L}_{\mathbf{k}}$  since  $\mathbf{F} \in \mathbb{R}^{\pi m}$ . Its dual program is given by

$$(\text{D-SDP}_n^{m,\leq}) \quad \left\{ \begin{array}{l} \text{Find } y, \boldsymbol{\mu} \in \mathbb{R} \times \mathbb{R}^{\pi m} \\ \text{minimising } y \\ \text{subject to} \\ y \geq 1 + \mu_n \\ \forall \mathbf{k} \leq \mathbf{m}, \mathbf{k} \neq \mathbf{n}, y \geq \mu_{\mathbf{k}} \\ \forall g \in \mathcal{R}_{m,+}(\mathbb{R}_+^M), \langle f_{\boldsymbol{\mu}}, g \rangle \geq 0, \end{array} \right.$$

for all  $m \geq n$ , where  $f_{\boldsymbol{\mu}} = \sum_{\mathbf{k} \leq \mathbf{m}} \mu_{\mathbf{k}} \mathcal{L}_{\mathbf{k}}$ . Moreover, adding the condition  $\mu_{\mathbf{k}} \leq 1$  for all  $\mathbf{k} \leq \mathbf{m}$  does not change the optimal value of the program. We enforce this condition in what follows. With Lemma 4.10 and Lemma 4.15, we thus obtain the following result:

**Lemma 4.16** (Generalisation of Lemma 4.10). *For all  $m \geq n$ ,  $\mathbf{F}^m$  (defined in Eq. (4.158)) is a feasible solution of  $(\text{SDP}_n^{m,\leq})$ .*

In particular, for all  $m \in \mathbb{N}^M$ ,  $\sum_{\mathbf{k} \leq \mathbf{m}} F_{\mathbf{k}}^m \mathcal{L}_{\mathbf{k}} \in \mathcal{R}_{m,+}(\mathbb{R}_+^M)$ . For  $m, n \in \mathbb{N}^M$  with  $m \geq n$ , let  $\boldsymbol{\mu} \in \mathbb{R}^{\pi m}$  be a feasible solution of  $(\text{D-SDP}_n^{m,\leq})$ . Then, for all  $l \leq m$

$$\left\langle \sum_{\mathbf{k} \leq \mathbf{m}} \mu_{\mathbf{k}} \mathcal{L}_{\mathbf{k}}, \sum_{\mathbf{k} \leq \mathbf{l}} F_{\mathbf{k}}^l \mathcal{L}_{\mathbf{k}} \right\rangle \geq 0, \quad (4.160)$$

so that

$$\sum_{\mathbf{k} \leq \mathbf{l}} \mu_{\mathbf{k}} F_{\mathbf{k}}^l \geq 0. \quad (4.161)$$

Hence, for all  $l \geq m$ ,

$$\begin{aligned} \mu_l &\geq -\frac{1}{F_l^l} \sum_{\substack{\mathbf{k} \leq \mathbf{l} \\ \mathbf{k} \neq \mathbf{l}}} \mu_{\mathbf{k}} F_{\mathbf{k}}^l \\ &\geq -\frac{1}{F_l^l} \sum_{\substack{\mathbf{k} \leq \mathbf{l} \\ \mathbf{k} \neq \mathbf{l}}} F_{\mathbf{k}}^l \\ &= 1 - \frac{1}{F_l^l} \\ &\geq 1 - \pi_l, \end{aligned} \quad (4.162)$$

where we used  $F_l^l > 0$  in the first line,  $\mu_k \leq 1$  and  $F_k^l \geq 0$  in the second line,  $\sum_{k \leq l} F_k^l = 1$  in the third line, and Eq. (4.159) in the last line. Eq. (4.162) is the multimode generalisation of Eq. (4.128). It is essential to prove the convergence of the lower bounding hierarchy as (i) it proves that the feasible set of  $(\text{SDP}_n^{m1, \leq})$  is compact and (ii) it shows that the sequence of optimal solutions obtained by diagonal extraction belongs to  $\mathcal{S}(\mathbb{N}^M)$ .

With these additional results, the proof of convergence of the multimode hierarchy of lower bounds  $(\text{SDP}_n^{m1, \leq})$  is then obtained directly from its single-mode counterpart using multi-index notations:

**Theorem 4.8** (Generalisation of Theorem 4.4). *The increasing sequence of optimal values  $\omega_n^{m1, \leq}$  of  $(\text{SDP}_n^{m1, \leq})$  converges to the optimal value  $\omega_n^S$  of  $(\text{LP}_n^S)$ :*

$$\lim_{m \rightarrow +\infty} \omega_n^{m1, \leq} = \omega_n^S. \quad (4.163)$$

With Lemma 4.14, we also obtain

$$\lim_{m \rightarrow +\infty} \omega_n^{m, \leq} = \omega_n^S. \quad (4.164)$$

Like in the single-mode case, Theorem 4.6 and Theorem 4.8 imply strong duality between the linear programs:

**Corollary 4.4** (Generalisation of Corollary 4.2). *Strong duality holds between the programs  $(\text{LP}_n^{L^2})$  and  $(\text{D-LP}_n^{L^2})$  and between programs  $(\text{LP}_n^S)$  and  $(\text{D-LP}_n^S)$ .*

#### 4.4.5 Multimode example

To illustrate the usefulness of our Wigner negativity witnesses in the multimode setting, we consider a lossy Fock state over two-modes:

$$\begin{aligned} \rho_{1,1,\eta} := & (1 - \eta)^2 |1\rangle\langle 1| \otimes |1\rangle\langle 1| + \eta(1 - \eta) |1\rangle\langle 1| \otimes |0\rangle\langle 0| \\ & + \eta(1 - \eta) |0\rangle\langle 0| \otimes |1\rangle\langle 1| + \eta^2 |0\rangle\langle 0| \otimes |0\rangle\langle 0|, \end{aligned} \quad (4.165)$$

with loss parameter  $0 \leq \eta \leq 1$ . Setting  $\eta = 0$  gives  $\rho_{1,1,\eta} = |1\rangle\langle 1| \otimes |1\rangle\langle 1|$  while setting  $\eta = 1$  gives  $\rho_{1,1,\eta} = |0\rangle\langle 0| \otimes |0\rangle\langle 0|$ . This state has a nonnegative Wigner function for  $\eta \geq \frac{1}{2}$ .

We also consider the multimode Wigner negativity witness  $|1\rangle\langle 1| \otimes |1\rangle\langle 1|$ , which is a projector onto the Fock state  $|1\rangle \otimes |1\rangle$ . Solving numerically the corresponding hierarchy  $(\text{SDP}_n^{m, \leq})$  up to  $m = 3$ , we obtain the lower bound 0.266 and solving the hierarchy  $(\text{SDP}_n^{m, \geq})$  up to  $m = 10$ , we obtain the upper bound 0.320.

A direct consequence of the numerical lower bound is that tensor product states are not the closest among Wigner positive states to tensor product states with a negative Wigner function. Indeed, the maximum achievable fidelity with the state  $|1\rangle \otimes |1\rangle$  using Wigner positive tensor

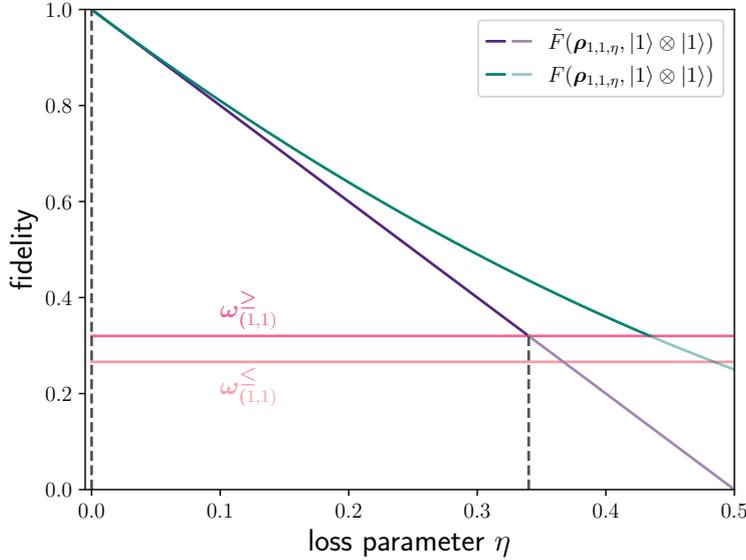


Figure 4.7: Witnessing Wigner negativity of the lossy Fock state  $\rho_{1,1,\eta}$  over two modes using the witness  $|1\rangle\langle 1| \otimes |1\rangle\langle 1|$ . The threshold value for that witness is upper bounded by 0.320 and lower bounded by 0.266. The dashed grey line delimits the interval of loss parameter values where the witness can be used to detect Wigner negativity of  $\rho_{1,1,\eta}$  efficiently, i.e., when the robust bound  $\tilde{F}(\rho_{1,1,\eta}, |1\rangle \otimes |1\rangle)$  (violet curve) on the fidelity from Eq. (4.167) is above the witness upper bound (pink line). When it is below the witness lower bound (light pink line), we are guaranteed that the witness cannot be used to detect Wigner negativity of the state. The fidelity  $F(\rho_{1,1,\eta}, |1\rangle \otimes |1\rangle)$  is also depicted above (green curve). Note that  $\rho_{1,1,\eta}$  has a nonnegative Wigner function for  $\eta \geq 0.5$ .

product states is equal to the square of the maximum achievable fidelity with the state  $|1\rangle$  using single-mode Wigner positive states, that is  $0.5^2 = 0.25 < 0.266$ .

We now use the upper bound to witness the Wigner negativity of the state  $\rho_{1,1,\eta}$  (see Fig. 4.7). The fidelity between  $\rho_{1,1,\eta}$  and  $|1\rangle \otimes |1\rangle$  is given by  $F(\rho_{1,1,\eta}, |1\rangle \otimes |1\rangle) = (1 - \eta)^2$ , for all  $0 \leq \eta \leq 1$ . This fidelity is above the upper bound 0.320 on the threshold value of the witness  $|1\rangle\langle 1| \otimes |1\rangle\langle 1|$  when  $\eta \leq 0.434$ .

However, in practice one would not obtain a precise estimate of the fidelity efficiently, but rather a robust lower bound on the fidelity computed from single-mode fidelities, which satisfies Eq. (4.142). In the worst case, the estimate obtained is closer to  $1 - 2(1 - F(\rho_{1,1,\eta}, |1\rangle \otimes |1\rangle))$  than to  $F(\rho_{1,1,\eta}, |1\rangle \otimes |1\rangle)$ . When the value of this robust lower bound is greater than the threshold value of the witness, this implies that the state has a negative Wigner function.

In the present case, the two single-mode reduced states of  $\rho_{1,1,\eta}$  are the same, given by

$$\text{Tr}_2(\rho_{1,1,\eta}) = (1 - \eta) |1\rangle\langle 1| + \eta |0\rangle\langle 0|, \quad (4.166)$$

so the single-mode fidelities with  $|1\rangle$  are equal for each mode and given by  $1 - \eta$ . Hence, the robust lower bound on  $F(\rho_{1,1,\eta}, |1\rangle \otimes |1\rangle)$  is given by

$$\tilde{F}(\rho_{1,1,\eta}, |1\rangle \otimes |1\rangle) = 1 - 2\eta. \quad (4.167)$$

It is above the upper bound 0.320 on the threshold value of the witness  $|1\rangle\langle 1| \otimes |1\rangle\langle 1|$  when  $\eta \leq 0.340$ .

This example highlights the use of efficient and robust lower bounds on multimode fidelities rather than fidelity estimates [CGKM20], in conjunction with our family of multimode witnesses to detect Wigner negativity of realistic experimental states.

## 4.5 Discussion and open problems

Characterising quantum properties of physical systems is an important step in the development of quantum technologies, and negativity of the Wigner function, a necessary resource for any quantum computational speedup, is no exception. In this chapter, we have derived a complete family of Wigner negativity witnesses which provide an operational quantification of Wigner negativity, both in the single-mode and multimode settings. In the context of quantum optical information processing, the main application of our method is in experimental scenarios, where it leads to robust and efficient certification of negativity of the Wigner function.

What is more, our witnesses also delineate the set of quantum states with positive Wigner function, and it would be interesting to understand whether additional insights on this set can be obtained using these witnesses. This is of particular importance for information theoretic notions like the recently defined quantum Wigner entropy [VHC21] where it is argued that this is the natural measure in order to characterise quantum uncertainty in phase-space even though it is limited to Wigner positive states.

Hierarchies of semidefinite programs (in particular with non-commutative variables [NPA08, KMV20, KMP21]) have found many recent applications in quantum information theory. From an infinite-dimensional linear program, we were able to use numerically both a hierarchy of upper bounds and a hierarchy of lower bounds—thus obtaining a certificate for the optimality of these bounds by looking at their difference—whereas this only works in specific cases for the Lasserre hierarchy of upper bounds [Las11]. Can we find other interesting cases where we can exploit both hierarchies? Moreover, we obtained an analytical sequence of lower bounds for the threshold value of the program ( $\text{LP}_n^{L^2}$ ). Can we also get an analytical sequence of upper bounds? In particular, we anticipate that Fock states  $|n\rangle$  get further away from the set of states having a positive Wigner function as  $n$  increases and that  $\omega_n = \mathcal{O}(\frac{1}{\sqrt{n}})$  as  $n \rightarrow +\infty$ . The question of whether the gap between the optimisation over square integrable functions and Schwartz functions is also left open.

## QUANTUM ADVANTAGE IN INFORMATION RETRIEVAL

**W**HILE the previous chapters focused on continuous-variable systems, we investigate here information retrieval tasks in discrete-variable systems that require Wigner negativity and contextuality to outperform the corresponding classical task. A well-known information retrieval task is random access coding which involves the encoding of a random input string into a shorter message string. The encoding should be such that any element of the input string can be retrieved with high probability from the message string. Such tasks have long been studied as examples in which the communication of quantum information can provide advantage, i.e. enhanced performance, over classical information, e.g. [ANTSV99, PPK<sup>+</sup>09, SBK<sup>+</sup>09, GHH<sup>+</sup>14, THMB15, CKKS16, ABMP18, FK19, Ga01]. However, random access coding concerns only one kind of information retrieval. In this chapter we define a general notion of information retrieval tasks and introduce another example called the Torpedo Game. It is similar to random access coding, but with additional requirements involving the retrieval of relative information about elements of the input string. Taking a geometric perspective it may also be viewed as a pacifist version of the popular strategy game *Battleship* (see Fig. 5.1). Quantum strategies can be implemented in prepare-and-measure scenarios, and outperform classical strategies for the Torpedo Games with bit and trit inputs. In particular, quantum perfect strategies exist in the trit case and provide a greater quantum advantage than for the comparable random access coding task [THMB15].

Critically, optimal quantum strategies emerge from an analysis in terms of the discrete Wigner function and exploits maximum Wigner negativity which has been showed to be necessary for computational speedup [Ga05, CGaG<sup>+</sup>06, ME12]. Yet while negativity is necessary for

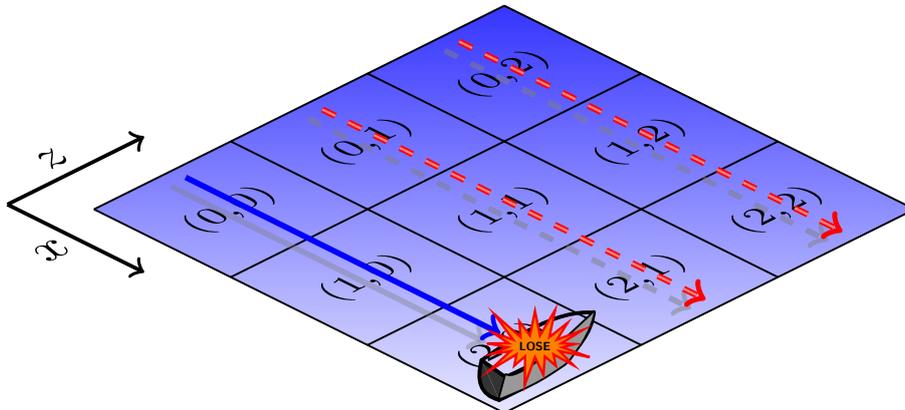


Figure 5.1: The Torpedo Game is a pacifist alternative to *Battleship* where the aim is to avoid sinking Alice’s ship, depicted here in dimension 3.

advantage in the Torpedo game, it does not seem to be sufficient. To more precisely pinpoint the source of quantum advantage we must look further. One natural candidate would be preparation contextuality [Spe05], another signature of non-classicality that has been linked to quantum random access codes in numerous studies [SBK<sup>+</sup>09, CKKS16, ABC<sup>+</sup>19]. It has been shown to be necessary for advantage in a restricted class of random access codes subject to an obliviousness constraint [HTMB17, SHP19]. Here however, we focus on a different characteristic called sequential contextuality [MK18]. It indicates the absence of a hidden-variable model respecting the sequential structure of a given protocol. It has already been used to explain the promotion of linear computation with access to a qubit quantum resource to the complexity class  $P$  in the setting of  $l_2$ -MBQC [MK18, AB09, OGa17]. Subject to an assumption of bounded-memory, we find that this characteristic is necessary and sufficient for quantum advantage, not just in random access coding but in any information retrieval task. Moreover, we show that it quantifies the degree of advantage that can be achieved.

As recalled in Section 1.5, contextuality can exhibit itself at the level of probability distributions (e.g. quantum violations of the CHSH inequality [CHSH69]) but also at the level of the supports of these distributions in some cases. In other words contextuality can be inferred by a series of logical deductions about which events are possible or not, e.g. Hardy’s paradox [Har92, Har93]. This situation has come to be known as logical contextuality [AB11]. In the most extreme cases, known as strong contextuality [AB11], every possible event triggers such a paradox [Man17], e.g. Popescu-Rohrlich box violations of the CHSH inequality [PR94]. For our purposes the quantum perfect strategies we introduce for the Torpedo game display analogous contextuality, and hence paradoxes, of this strongest form in a prepare-and-measure scenario.

Section 5.1 gives an overview of information retrieval tasks including random access coding and the Torpedo Game. Section 5.2 provides background on discrete Wigner functions. Section 5.3 deals with optimal classical and quantum strategies for the Torpedo Game. Finally, Section 5.4 establishes the relationship between sequential contextuality and quantum advantage

in bounded-memory information retrieval tasks. This chapter is based on [EHM20].

## 5.1 Information retrieval tasks

### 5.1.1 Random access codes

For  $n, m \in \mathbb{N}^*$  with  $m < n$ , an  $(n, m)_2$  random access code (RAC)—sometimes denoted  $n \rightarrow m$ —is a communication task in which one aims to encode information about a random  $n$ -bit input string into an  $m$ -bit message, in such a way that any one of the input bits may be retrieved from the message with high probability. A  $(n, m)_2$  quantum random access code (QRAC) instead encodes the input into an  $m$ -qubit (quantum) message state.

Such tasks may be considered as two-party cooperative games in which the first party, Alice, receives a random input string from a referee. Alice encodes information about this in a message that is communicated to the second party, Bob. The referee then asks Bob to retrieve the value of the bit at a randomly chosen position in the input string. We will assume that the referee's choices are made uniformly at random.

For instance, for the  $(2, 1)_2$  RAC game [ANTSV99] an optimal classical strategy is for Alice to directly communicate one of the input bits to Bob. If asked for this bit, Bob can always return the correct answer, otherwise Bob guesses and will provide the correct answer with probability  $\frac{1}{2}$ . Thus the game has a classical value of

$$\theta_{2 \rightarrow 1}^C = \frac{1}{2} \left( 1 + \frac{1}{2} \right) = \frac{3}{4}.$$

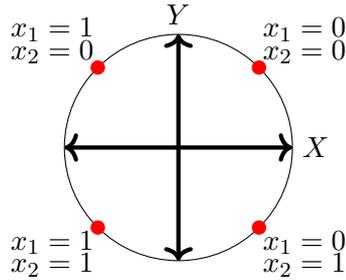


Figure 5.2: The four red dots correspond to the four states  $|\psi_{x_1, x_2}\rangle$  defined in Eq. (5.1) depicted as points on the equator of the Bloch sphere.

Quantum strategies can outperform this classical bound. An optimal quantum strategy is for Alice to communicate the qubit state

$$|\psi_{x_1, x_2}\rangle = \frac{1}{\sqrt{2}} \left( |0\rangle + \frac{1}{\sqrt{2}} ((-1)^{x_1} + (-1)^{x_2} i) |1\rangle \right) \quad (5.1)$$

where  $(x, z)$  is the input bit-string they have received. Bob measures in the  $X$  basis when asked for  $x_1$  and in the  $Y$  basis when asked for  $x_2$  (see Fig. 5.2). If they obtain the  $+1$  eigenvalue they

return the value 1 and if they obtain the  $-1$  eigenvalue they return 0. This yields a quantum value for the game of

$$\theta_{2 \rightarrow 1}^Q = \cos^2\left(\frac{\pi}{8}\right) \approx 0.85.$$

### 5.1.2 General Information Retrieval Tasks

One may also consider more general communication scenarios. In an  $(n, m)_d$  *communication scenario* the input is a random string of  $n$  dits and the message is a string of  $m$  (qu)dits, for  $d \geq 2$ . (Q)RAC tasks have previously been considered in such scenarios, e.g. in [THMB15, CGaS08].

However, we also wish to accommodate for a much wider range of retrieval tasks regarding the input dits. An *information retrieval task* in an  $(n, m)_d$  communication scenario is specified by a tuple  $\langle Q, \{w_q\}_{q \in Q} \rangle$  where

- $Q$  is a finite set of *questions*;
- The  $w_q : \mathbb{Z}_d^n \rightarrow \mathbb{Z}_d$  are *winning relations* which pick out the good answers to question  $q$  given an input string in  $\mathbb{Z}_d^n$ . Note that there may be more than one good answer, or none. It is assumed that inputs and outputs are endowed with the structure of the commutative ring  $\mathbb{Z}_d$ .

Standard  $(n, m)_d$  (Q)RACs are recovered when the questions ask precisely for the respective input dits. In that case the winning relations  $w_i$  reduces simply to projectors onto the respective dits of the input string. However, other interesting tasks arise when the questions also concern relative information about the input string, in the form of parities or linear combinations modulo  $d$  of the input dits. A similar generalisation for  $d = 2$ , using functions rather than relations, has been independently proposed in [DM21]. Below we introduce an example called the Torpedo Game which is distinct from random access coding. We show that Wigner negativity is necessary to outperform the best classical strategy. We later link this advantage to sequential contextuality [MK18].

### 5.1.3 The Torpedo Game

Of particular interest here is an information retrieval task for  $(2, 1)_d$  communication scenarios where we fix the dimension to be  $d = 2$  or  $d = 3$ . We take the game perspective and refer to the task as the dimension  $d$  Torpedo Game (see Fig. 5.1 in dimension 3). Let  $x$  and  $z$  be the two input bits or trits. There are 3 questions in  $Q = \{\infty, 0, 1\}$  dimension 2 and 4 questions  $Q = \{\infty, 0, 1, 2\}$  in dimension 3. The labelling comes from a geometric interpretation to be

elaborated upon shortly. Winning relations for the Torpedo Game are given by

$$\begin{aligned}
 w_\infty(x, z) &= \{a \in \mathbb{Z}_d \mid a \neq x\} \\
 w_0(x, z) &= \{a \in \mathbb{Z}_d \mid a \neq -z\} \\
 w_1(x, z) &= \{a \in \mathbb{Z}_d \mid a \neq x - z\} \\
 w_2(x, z) &= \{a \in \mathbb{Z}_d \mid a \neq 2x - z\}.
 \end{aligned} \tag{5.2}$$

All arithmetic is modulo  $d = 2$  or  $3$  depending on the considered dimension. The last winning relation only features in the dimension 3 Torpedo game.

For  $d = 2$ , the Torpedo Game is equivalent to a  $(2, 1)_2$  (Q)RAC, but with an additional question. Bob may be asked to retrieve either one of the individual input dits, or to retrieve relative information about them in the form of their parity  $x \oplus z$ . Note that in dimension 2, the winning relations are actually functions (there is only one good answer per question).

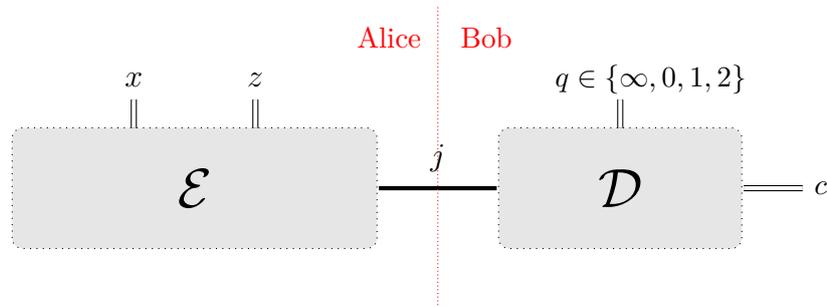


Figure 5.3: Prepare-and-measure protocol for the dimension 3 Torpedo Game: Alice receives trits  $x$  and  $z$  and sends a single message (qu)trit  $j$  via the encoding  $\mathcal{E}$ . Bob is asked a question  $q \in \{\infty, 0, 1, 2\}$ , performs decoding  $\mathcal{D}$ , and outputs  $c$  which should satisfy the winning conditions given by  $w_q(x, z)$  with high probability.

The Torpedo Game may be framed as cooperative, pacifist alternative to the popular game *Battleship*, in which Alice and Bob, finding themselves on opposing sides in a context of naval warfare, wish to subvert the conflict and cooperate to avoid casualties while not directly disobeying orders. The input dits received by Alice designates the coordinates in which Alice is ordered by their commander to position their one-cell ship on the affine plane of order  $d$  for  $d = 2$  or  $d = 3$ . We may think of the affine plane as a toric  $d \times d$  grid, with  $x$  designating the row and  $z$  the column. For instance in Fig. 5.4 we identify the top edge with the bottom edge and the left edge with the right edge. Bob is a naval officer on the opposing side who is ordered by their commander to shoot a torpedo along a line of the grid with slope specified by  $q \in Q$ . The  $\infty$  question requires Bob to shoot along some row, and the 0 question requires Bob to shoot along some column, etc. However, Bob retains the freedom to choose which row, or column, or diagonal of given slope, as the case may be. In other terms, upon receiving  $q$  Bob must shoot along a lines  $qx - z = c$  (if  $q \neq \infty$ ) or  $x = c$  (if  $q = \infty$ ) but is free to choose the constant  $c$ .

Alice and Bob wish to coordinate a strategy for avoiding casualties, while still obeying their explicit orders. Alice may communicate a single (qu)dit to Bob—greater communication may risk revealing their position should it be intercepted—and based on this Bob must choose their  $c$  in such a way that they avoid Alice’s ship.

## 5.2 The discrete Wigner function and information retrieval

Below we work in dimension 3 as the construction we present will only provide a deterministic strategy for the dimension 3 Torpedo game. We will see that there exists a quantum strategy inspired by this construction that outperforms the best classical strategy for the dimension 2 Torpedo game though it is not deterministic. Wootters’s geometric construction [GHW04, Woo87] of discrete Wigner functions (DWF) based on finite fields is useful for visualising our Torpedo Game as exemplified in Figure 5.4, where each distinct orthonormal basis corresponds to a set of 3 parallel (non-intersecting) lines.

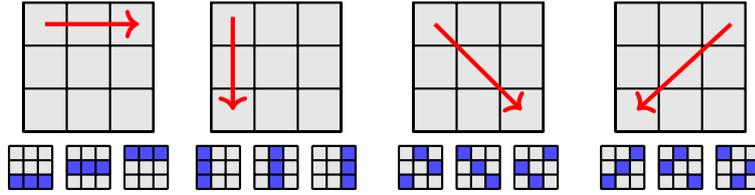


Figure 5.4: The red arrows depict the directions or slopes  $(\infty, 0, 1, 2, \text{ respectively})$  along which Bob may be asked to shoot in the  $d = 3$  Torpedo Game. For each direction, Bob has three possibilities, depicted by the blue lines. In the affine plane of order 3, each of these groups of three blue cells forms a line.

We recall from Subsection 1.2.1 (see Eq. (1.13)) the expression of the phase point operator at the origin of phase space  $\hat{\Pi} = \hat{A}_{0,0}$  (also called the parity operator):

$$\hat{A}_{0,0} := \sum_{k \in \mathbb{Z}_3} |-k\rangle\langle k|,$$

The other phase-point operators are found by conjugation with displacement operators (see Eq. (1.12)). The phase-point operator at point  $(x, z) \in \mathbb{Z}_3^2$  reads:

$$\hat{A}_{x,z} := \hat{D}_{x,z} \hat{A}_{0,0} \hat{D}_{x,z}^\dagger. \quad (5.3)$$

We recall the definition of the Wigner function (Eq. (1.26)). An example of representation of the DWF in dimension 3 is given in Figure 5.5. For a density matrix  $\rho$  and for a phase-space point  $(x, z) \in \mathbb{Z}_3^2$ , it is given by:

$$W_\rho(x, z) = \frac{1}{d} \text{Tr}(\hat{A}_{x,z} \rho). \quad (5.4)$$

Crucially phase operators may also be constructed from projectors onto mutually unbiased bases (MUBs) [Woo87]. For  $x, z \in \mathbb{Z}_3$ :

$$\hat{A}_{x,z} = \Pi_\infty^x + \Pi_0^{-z} + \Pi_1^{x-z} + \Pi_2^{2x-z} - \mathbb{1} \quad (5.5)$$

where  $\Pi_q^i$  is the projector corresponding to dit value  $i$  in Bob's  $q^{\text{th}}$  measurement setting for  $q \in \{\infty, 0, 1, 2\}$ .

The eigenvectors of phase point operators are the objects of interest. The maximizing eigenvectors of the phase point operators in Eq. (5.3) (and additional ones from different choices of DWF) were used in Casaccino *et al.* [CGaS08] as the encoded messages of a  $(d+1, 1)_d$  QRAC (for any odd power-of-prime dimension  $d$ ). This is natural given the use of MUBs in constructing DWFs, and prominence of MUBs in the QRAC literature [ABMP18]. If Alice receives input  $\mathbf{k} = (k_1, k_2, \dots, k_{d+1}) \in \mathbb{Z}_d^{d+1}$  that they encodes in  $\rho_{\mathbf{k}}$  and transmits to Bob, then the average probability of success for the Casaccino *et al.* QRAC is

$$\frac{1}{(d+1)d^{d+1}} \sum_{\mathbf{k} \in \mathbb{Z}_d^{d+1}} \text{Tr} \left[ \rho_{\mathbf{k}} (\Pi_1^{k_1} + \dots + \Pi_{d+1}^{k_{d+1}}) \right] . \quad (5.6)$$

where  $\Pi_q^i$  is the projector corresponding to dit value  $i$  in Bob's  $q$ -th measurement setting. Since phase point operators can also be constructed using sums of projectors onto MUBs (see Eq. (5.5)), the use of a maximizing eigenvector of a phase point operator for  $\rho_{\mathbf{k}}$  is natural to maximize Eq. (5.6).

Here we instead make use of the *minimizing* eigenvectors of phase point operators. The rationale for this is two-fold (i) these eigenvectors display remarkable geometric properties with respect to the measurements in (their constituent) mutually unbiased bases, and (ii) negativity is the hallmark of non-classicality which has already been identified with contextuality (with the previously mentioned caveat that an additional ‘‘spectator’’ subsystem was required). These will be seen to lead to a perfect quantum strategy for the Torpedo Game, in which the goal is to avoid certain answers.

Any state in the  $-1$  eigenspace of phase point operators defined in Eq. (5.3) has an outcome that is forbidden [vDH11, BBC12] in each of a complete set of MUBs. For example, let

$$|\psi_{0,0}\rangle = \frac{|1\rangle - |2\rangle}{\sqrt{2}} \quad (5.7)$$

satisfying

$$\hat{A}_{0,0} |\psi_{0,0}\rangle = -|\psi_{0,0}\rangle . \quad (5.8)$$

This state, which is indeed an eigenvector of  $A_{0,0}$  with eigenvalue  $-1$ , obeys

$$\text{Tr}(\Pi_q^0 |\psi_{0,0}\rangle \langle \psi_{0,0}|) = 0 \quad (5.9)$$

where  $\Pi_q^0$  is the projector on the  $0^{\text{th}}$  eigenvector in the  $q^{\text{th}}$  basis for  $q \in \{\infty, 0, 1, 2\}$ . More specifically,  $\Pi_q^0$  is the projector corresponding to the  $\omega^0 = +1$  eigenvector of the  $q^{\text{th}}$  displacement

operator from  $\{\hat{D}_{0,1}, \hat{D}_{1,0}, \hat{D}_{1,1}, \hat{D}_{1,2}\}$ . These displacement operators have eigenvectors leading to mutually unbiased measurement bases  $q \in \{\infty, 0, 1, 2\}$  respectively. For  $x, z \in \mathbb{Z}_3$ , the related states  $|\psi_{x,z}\rangle = \hat{D}_{x,z} |\psi_{0,0}\rangle$ , which are eigenstates  $\hat{A}_{x,z} |\psi_{x,z}\rangle = -|\psi_{x,z}\rangle$ , obey

$$\text{Tr} \left[ (\Pi_{\infty}^x + \Pi_0^{-z} + \Pi_1^{x-z} + \Pi_2^{2x-z}) |\psi_{x,z}\rangle \langle \psi_{x,z}| \right] = 0. \quad (5.10)$$

Eq. (5.10) implies that the probability of the relevant outcome (outcome  $x$  in the first basis,  $-z$  in the second basis, etc.) in each of the MUBs is zero (see the winning relations of the Torpedo game in Eq. 5.2). The general expression in Eq. (5.10) for any odd power-of-prime  $d$  is proven in [How15, ABC08].

### 5.3 Optimal strategies for the Torpedo Game

Here we gather the optimal classical, quantum and (in one case) post-quantum strategies for the Torpedo Game. The dimension 3 Torpedo game is notable for the existence of a perfect quantum strategy due to the fact that there exist  $d + 1 = 4$  MUBs for this dimensions (it is true in general for odd power-of-prime dimensions though there also exists a perfect classical strategy for these dimensions). The dimension 2 Torpedo Game also has a quantum advantage though the quantum strategy is not perfect. The classical optimum is also established rigorously for dimension 2 and 3. We obtain a quantum advantage in the two cases.

#### 5.3.1 Optimal quantum and post-quantum strategies

**Quantum perfect strategy for the dimension 3 Torpedo Game.** From Eq. (5.10) it follows that there is a perfect quantum strategy for the dimension 3 Torpedo game:

1. Upon receiving trits  $x$  and  $z$  Alice sends the following state to Bob:

$$|\psi_{x,z}\rangle = \hat{D}_{x,z} |\psi_{0,0}\rangle = \hat{D}_{x,z} \left( \frac{|1\rangle - |2\rangle}{\sqrt{2}} \right). \quad (5.11)$$

2. Bob receives  $|\psi_{x,z}\rangle$  and is asked a question  $q \in \{\infty, 0, 1, 2\}$ . They measure the state in the MUB corresponding to  $q$  and outputs the dit corresponding to the measurement outcome.

This quantum strategy wins the Torpedo Game deterministically, i.e. with probability 1. In Figure 5.5 we also provide geometric intuition for why this strategy is perfect in dimension 3.

**Optimal quantum strategy for the dimension 2 Torpedo Game.** An analogous strategy to the qutrit case can be employed for the qubit Torpedo Game, using message states  $|\psi_{x,z}\rangle = \mathbf{X}^x \mathbf{Z}^z |\psi_{0,0}\rangle$  where  $|\psi_{0,0}\rangle \langle \psi_{0,0}| = \frac{1}{2} (\mathbb{I} - (\mathbf{X} + \mathbf{Y} + \mathbf{Z})/\sqrt{3})$ . For  $d = 2$ , while this does not constitute a perfect strategy it still achieves an advantage over classical strategies. In fact, it turns out to be an optimal strategy: this strategy achieves a winning probability of

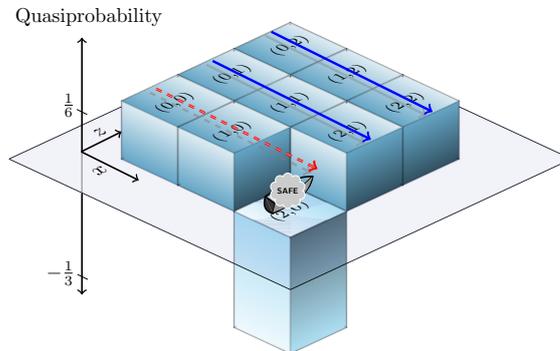


Figure 5.5: The perfect quantum strategy can be understood by plotting the discrete Wigner function (see Eq. (1.26)) of the message state sent by Alice. In the above qutrit case, Alice is in coordinate  $(x, z) = (2, 0)$ , so Alice sends Bob the state  $|\psi_{2,0}\rangle$  whose Wigner function is  $-1/3$  at coordinate  $(2, 0)$  and  $1/6$  otherwise. Bob measures this state along any of the four allowed directions, wherein the probability of each outcome is given by sum of quasiprobabilities along the corresponding line. Hence, the only outcomes with non-zero probability of occurring correspond to lines not passing through  $(2, 0)$ . Whichever outcome Bob sees, they may fire their torpedo along the corresponding line, safe in the knowledge that it will not intersect Alice's ship. In this figure the solid blue lines correspond to the possible outcomes for the  $q = 0$  direction, but the same argument holds for all the other directions.

approximately 0.79 and we show that this is optimal. As explained in Subsection 1.4.4, this problem is very hard in general (NP-hard) since it is a nonconvex bilinear quadratic problem. To show it achieves 0.79, first we can leverage the fact that the  $(3, 1)_2$  (Q)RAC attributed to Isaac Chuang is at least as hard to win as the Torpedo Game. This  $(3, 1)_2$  (Q)RAC has 3 independent inputs while the dimension 2 Torpedo game can be viewed as a modified  $(3, 1)_2$  (Q)RAC where the last input is the parity of the two first input bits. The  $(3, 1)_2$  (Q)RAC is then a harder task and it has been shown to have an optimal quantum value of  $\frac{1}{2} \left(1 + \frac{1}{\sqrt{3}}\right)$ . Thus we get a lower bound of  $\frac{1}{2} \left(1 + \frac{1}{\sqrt{3}}\right) \approx 0.79$  on the optimal quantum value. To obtain a matching upper bound, we implemented numerically the NPA hierarchy [NPA08] which is a hierarchy of semidefinite programs converging from the exterior to the correlations arising from quantum systems. Because the message sent from Alice to Bob is of finite dimension we relied mostly on [NFAV15] which permits a characterisation of correlations arising from finite-dimensional quantum systems. We found a matching upper bound proving that indeed  $\theta_{d=2}^Q \approx 0.79$ .

**Perfect Post-quantum Strategy for for the dimension 2 Torpedo Game.** The average probability of success for the Casaccino et al. QRAC, see Eq. (5.6), can be maximized by using

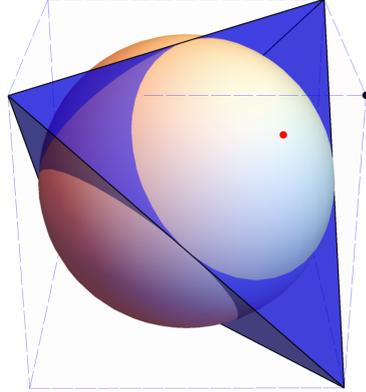


Figure 5.6: The qubit version of the Torpedo Game has a perfect strategy when allowed access to post-quantum “states”. The red point on the surface of the Bloch sphere represents the optimal message state  $|\psi_{0,0}\rangle$ , achieving of a value of 0.79 for the Torpedo game. The black point representing  $\frac{1}{2}(\mathbb{I} - (X + Y + Z))$  is not a valid density matrix, but achieves a value of 1 in the Torpedo Game.

a post-quantum “state” of the form  $\Pi_1^{k_1} + \Pi_2^{k_2} + \dots + \Pi_{d+1}^{k_{d+1}} - \mathbb{I}$  for any odd power-of-prime dimension  $d$ , where scare quotes reflect the fact that, although it is Hermitian and has unit trace, its spectrum is not necessarily nonnegative. In fact the “state” above is a phase point operator,  $\hat{A}_{\mathbf{k}}$ , for one of Wootters’ discrete Wigner functions. Since phase point operators obey  $\text{Tr}(\hat{A}\mathbb{1}) = 1$  and  $\text{Tr}(\hat{A}\hat{A}) = d$  then  $\frac{1}{(d+1)d^{d+1}} \sum_{\mathbf{k} \in \mathbb{Z}_d^{d+1}} \text{Tr}[\hat{A}_{\mathbf{k}}(\hat{A}_{\mathbf{k}} + \mathbb{1})] = 1$ . In other words, there is a perfect strategy by using post-quantum states. Seen in this way, phase point operators in a  $(d+1, 1)_d$  QRAC scenario are similar to Popescu-Rohrlich [PR94] boxes in the CHSH scenario. As such, our Torpedo Game has a perfect strategy within quantum mechanics for all odd power-of-prime dimensions, by construction. In contrast, we saw that the qubit Torpedo Game only has quantum value of roughly 0.79. To reach a perfect strategy, we may once again use a phase point operator as the non-physical “state” that Alice sends to Bob, see Figure 5.6.

### 5.3.2 Optimal classical strategies

In what follows  $d \in \mathbb{N}^*$  is arbitrary and we derive the classical value of a generic information retrieval task in a communication scenario  $(2, 1)_d$ . We describe an encoding map  $\mathcal{E} = \{p_{\mathcal{E}}(\cdot|x, z)\}_{x, z}$  as specifying a probability distribution over messages  $j \in \mathbb{Z}_d$  for each combination of inputs  $x, z \in \mathbb{Z}_d$ . Similarly a decoding map  $\mathcal{D} = \{p_{\mathcal{D}}(\cdot|j, q)\}_{j, q}$  specifies a probability distribution over outputs  $c \in \mathbb{Z}_d$  for each combination of a message and question,  $j \in \mathbb{Z}_d$  and  $q \in Q$  respectively.

Combining an encoding  $\mathcal{E}$  and a decoding  $\mathcal{D}$  results in an empirical behaviour that we can write  $e = \{p_e(\cdot|x, z, q)\}_{x, z, q}$ . This is a set of probability distributions over outputs  $c \in \mathbb{Z}_d$ , one for each combination of the referee variables  $x, z \in \mathbb{Z}_d, q \in Q$  such that

$$p_e(c|x, z, q) = \sum_{j \in \mathbb{Z}_d} p_{\mathcal{D}}(c|j, q) p_{\mathcal{E}}(j|x, z). \quad (5.12)$$

By comparison, quantum mechanical empirical behaviours arise via the Born rule that is  $p_e(c|x, z, q) = \text{Tr}(\rho_{x,z}\Pi_q^c)$ .

Assuming the referee variables to be uniformly distributed, a strategy has a winning probability given in terms of its empirical probabilities as

$$\frac{1}{d^2(d+1)} \sum_{x,z,q} p_e(w_q(x, z) | x, z, q).$$

The classical value of the Torpedo Game can thus be expressed as

$$\theta_d^C = \max_{\mathcal{E}, \mathcal{D}} \left[ \frac{1}{d^2(d+1)} \sum_{x,z,q} p_e(w_q(x, z) | x, z, q) \right]. \quad (5.13)$$

To evaluate this expression note that it suffices to consider deterministic encodings and decodings. In the presence of shared randomness, nondeterministic strategies can always be obtained as convex combinations of deterministic ones and the expression is convex linear [GBHA10]. Furthermore, for each encoding there exists a decoding that is optimal with respect to it. This fact was also observed for one-way communication tasks with messages of bounded dimension in [SHP19]. Thus it is possible to evaluate the classical value by enumerating over deterministic encodings only.

**Proposition 5.1.** *The classical value of an information retrieval task in a  $(2, 1)_d$  communication scenario can be expressed as a maximum over encodings as*

$$\theta^C = \max_{\mathcal{E}} \left[ \frac{1}{d^2(d+1)} \sum_{j,q} \max_c \sum_{\substack{(x,z) \text{ s.t.} \\ c \in w_q(x,z)}} p_{\mathcal{E}}(j|x, z) \right]. \quad (5.14)$$

*Proof.* Starting from Eq. (5.13),

$$\begin{aligned} \theta^C &= \max_{\mathcal{E}, \mathcal{D}} \left[ \frac{1}{d^2(d+1)} \sum_{x,z,q} p_e(w_q(x, z) | x, z, q) \right] \\ &= \max_{\mathcal{E}, \mathcal{D}} \left[ \frac{1}{d^2(d+1)} \sum_{x,z,q} \sum_{c \in w_q(x,z)} p_e(c | x, z, q) \right] \\ &= \max_{\mathcal{E}, \mathcal{D}} \left[ \frac{1}{d^2(d+1)} \sum_{q,c} \sum_{\substack{(x,z) \text{ s.t.} \\ c \in w_q(x,z)}} p_e(c | x, z, q) \right] \\ &= \max_{\mathcal{E}, \mathcal{D}} \left[ \frac{1}{d^2(d+1)} \sum_{j,q,c} \sum_{\substack{(x,z) \text{ s.t.} \\ c \in w_q(x,z)}} p_{\mathcal{D}}(c|j, q) p_{\mathcal{E}}(j | x, z) \right] \\ &= \max_{\mathcal{E}} \left[ \frac{1}{d^2(d+1)} \sum_{j,q} \max_c \sum_{\substack{(x,z) \text{ s.t.} \\ c \in w_q(x,z)}} p_{\mathcal{E}}(j | x, z) \right], \end{aligned}$$

where the last line follows by using a deterministic decoding that is optimal with respect to the encoding.  $\square$

A useful way of representing any deterministic encoding is as a colouring of the  $d \times d$  affine plane using no more than  $d$  colours. Observe that a deterministic encoding can alternatively be expressed as a function  $f_{\mathcal{E}} : \mathbb{Z}_d \times \mathbb{Z}_d \rightarrow \mathbb{Z}_d$ , where  $f_{\mathcal{E}}(x, z)$  is the message dit to be sent (with probability 1) given inputs  $x, z$ . Thinking of the inputs as coordinates in the  $d \times d$  affine plane a deterministic encoding is equivalent to a partition of the plane into no more than  $d$  equivalence classes, or a colouring using no more than  $d$  colours.

**Optimal classical strategies for  $d = 2$  and  $d = 3$  Torpedo Game.** In general there are  $d^{d^2}$  partitions of a  $d \times d$  grid. For low dimensions the expression in Eq. (5.14) can be evaluated by exhaustive search over partitions. For dimension 2 and 3 we find

$$\theta_{d=2}^C = \frac{3}{4} \quad \text{and} \quad \theta_{d=3}^C = \frac{11}{12}. \tag{5.15}$$

Example of strategies that attain these values are depicted below in Figure 5.7 and in Figure 5.8.

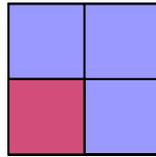


Figure 5.7: An optimal classical strategy for the  $d = 2$  Torpedo Game. Alice uses their bit of communication to indicate in which class of the partition that they find themselves. Classes are represented here by colours.

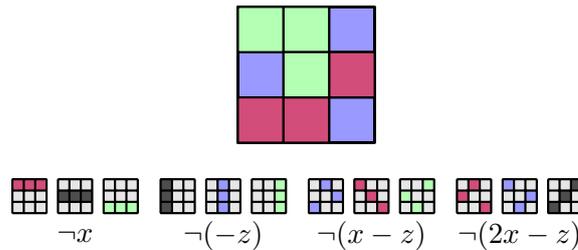


Figure 5.8: An optimal classical strategy for the  $d = 3$  Torpedo Game. Alice uses their dit of communication to indicate in which equivalence class (represented by same coloured cells) of the large grid partition they find themselves. The smaller grids (cf. Fig. 5.4) show where Bob chooses to shoot, given a direction and a colour. For the first direction, when asked to shoot horizontally in the grid, notice that Bob may avoid Alice with certainty if Alice is in either of the red or green partitions. Lines that avoid Alice with certainty are depicted in the corresponding colour, whereas black lines intersect with Alice’s position with probability  $\frac{1}{3}$ . Overall, this strategy wins the Torpedo Game with probability  $\frac{1}{4}(\frac{8}{9} + \frac{8}{9} + 1 + \frac{8}{9}) = \frac{11}{12}$ .

### 5.3.3 Comparison of quantum and classical game values

Recall the optimal quantum values established in Subsection 5.3.1,

$$\theta_{d=2}^Q \simeq 0.79 \quad \text{and} \quad \theta_{d \geq 3}^Q = 1. \quad (5.16)$$

Comparing these with the classical bounds from Subsection 5.3.2 we obtain the ratios

$$\frac{\theta_{d=2}^Q}{\theta_{d=2}^C} \simeq 1.053 \quad \text{and} \quad \frac{\theta_{d=3}^Q}{\theta_{d=3}^C} \simeq 1.091. \quad (5.17)$$

By comparison, it was shown in [THMB15] that the classical and quantum values of the  $(4, 1)_3$  (Q)RAC are  $\frac{16}{27}$  and 0.637, respectively, giving a ratio of  $\theta_{d=3}^Q/\theta_{d=3}^C \simeq 1.075$ . Accordingly, we note that the  $d = 3$  Torpedo Game admits a greater quantum-over-classical advantage than the standard random access coding task whose optimal QRAC also exploits the 4 mutually unbiased bases available in dimension 3.

### 5.3.4 Dimensional witness

The Torpedo Game can be used as a dimensional witness [BPA<sup>+</sup>08] for qubits and qutrits. In the following, we modify slightly the setting of the game to allow the message to be of arbitrary dimension. In particular, we no longer require that the message between Alice and Bob is of the same dimension as the inputs. For instance, we will allow Alice to send a (qu)trit while they receive two input bits. We will thus specify the *dimension of the inputs* as well as the *dimension of the message*. Questions are defined as before and are fixed by the dimension of the inputs:  $d + 1$  questions for inputs of dimension  $d$ .

Input dimension	Message dimension	$\theta^C$	$\theta^Q$
2	2	0.75	0.789
2	3	0.833	0.875
3	2	0.833	> 0.867
3	3	0.917	1

Table 5.1: Classical and quantum optimal values of the Torpedo Game when we allow the message dimension to differ from that of the inputs. Classical values were computed by exhaustive search. Quantum values were obtained by a combination of a seesaw algorithm and implementation of the NV hierarchy [NFAV15] through the interface QDimSum [TRR19]. Note that only the seesaw algorithm succeeded for the trit-input, qubit-message Torpedo Game as the NV hierarchy does not perform well with POVMs.

Following Table 5.1, we can use the Torpedo Game to discriminate between qubits and qutrits. Moreover these witnesses can distinguish between classical and quantum systems of the same dimension.

## 5.4 Sequential contextuality

When quantum advantage is observed in a bounded-memory information retrieval task like a (Q)RAC task or the Torpedo Game, it highlights a difference between the information carrying capacities of qudits compared to dits for a fixed dimension. It can be remarked that such a difference is a consequence of the different geometries of the respective state spaces. In this section, however, we seek a sharper, quantified analysis of the source of the advantage in terms of contextuality.

We will be using the notion of sequential contextuality that was introduced in [MK18] to extend structural treatments of Bell-Kochen-Specker contextuality [AB11] to sequential operational scenarios. As such, it is a behavioural characteristic that can arise in experiments involving sequences of operations. While [MK18] was concerned specifically with sequences of transformations, here we take a broader view that also includes the operations of preparation and measurement. In the special case of prepare-and-measure scenarios, sequential noncontextuality recovers a natural notion of classicality in terms of realisability by hidden-variables. For instance in the bounded-memory regime we are interested in, sequential contextuality also matches the characteristic introduced by Żukowski in [Żuk14].

We note that sequential contextuality is distinct from the notion of contextuality due to Spekkens [Spe05], as discussed in [MK18]. It is also distinct from the analyses of [KZG<sup>+</sup>09, ARBC09, LLS<sup>+</sup>11, GKC<sup>+</sup>10], which sought to close potential loopholes created by sequentiality of measurements in experimental tests of the more traditional Bell-Kochen-Specker form of contextuality. To determine its precise relationship with the latter analyses is however an interesting open question.

The study of contextuality arose in quantum foundations, where a major theme is the attempt to understand empirical behaviours that may appear non-intuitive from a classical perspective, e.g. the EPR paradox [EPR35]. The typical approach is to look for a description of physical systems at a deeper level than the quantum one at which more classically intuitive properties may be restored. Such a description is usually formalised as a hidden-variable model for the behaviour (sometimes also referred to as an ontological model [Spe05]). The great significance of the celebrated no-go theorems of quantum foundations, like Bell's Theorem [Bel64] and the Bell-Kochen-Specker Theorem [Bel66, KS75], was to prove that certain 'non-classical' features of the empirical behaviours of quantum systems are necessarily inherited by any underlying model under minimal assumptions.

Non-classical features of quantum systems like contextuality are also increasingly investigated for their practical utility. For instance, contextuality of the Bell-Kochen-Specker kind was shown to be a prerequisite for quantum speed-up [HWVE14] and to quantify quantum-over-classical advantage in a variety of information processing tasks [ABM17].

Bell-Kochen-Specker contextuality essentially concerns the statistics that arise under varied measurements on a physical system. In contrast, our notion of contextuality concerns

the statistics that arise from sequences of operations – preparations, transformations and measurements – all of which can vary. As a behavioural feature, sequential contextuality signifies the absence of any hidden-variable model that would preserve a compositional description of operations performed in sequence. In other words, sequential *non*contextuality requires a hidden-variable model in which each operation has an independent, modular description (as a transformation on hidden variables) such that to describe a sequence of operations one simply composes their hidden-variable descriptions. We provide a more rigorous mathematical description in the following subsections.

Rather than focusing on characteristics that must be inherited by all hidden-variable models, as is common in foundational works, we also take a practical perspective and shift focus to characteristics that must be inherited by bounded-memory models – a constraint that matches the information retrieval problem at hand. In this respect, the significance of sequential contextuality in what follows can also be viewed through a practical rather than a foundational lens, as a characteristic that quantifies quantum advantage.

#### 5.4.1 Empirical behaviours and hidden-variable Models

Recall from Section 5.3.2 that any strategy for a general information retrieval task gives rise to an empirical behaviour  $e = \{p_e(\cdot|i, q)\}_{i, q}$  (see Eq. (5.12)). In other words, for each combination of input string  $i \in \mathbb{Z}_d^n$  and question  $q \in Q$  there is a resulting probability distribution over outputs. This is true regardless of whether the strategy is classical, quantum, post-quantum, or other (like generalised probabilistic theories [Bar07]). The combination of input string and question fully specifies the precise operations that are performed in the sequence. This is what we refer to as the context, just as a context in a (Bell–Kochen–Specker) measurement scenario specifies a set compatible measurements to be performed jointly. We have also chosen a formal description of empirical behaviour that echoes the formalism of empirical models for measurement scenarios in [AB11] (see Subsection 1.5).

Given an empirical behaviour, one can ask whether it can be simulated by a bounded-memory hidden-variable model. In particular we will be interested in models that respect the sequential structure of the strategy, bearing in mind that the inputs and questions specify a sequence of operations: either preparations, transformations, or measurements. Our main focus is on prepare-and-measure scenarios, in which sequences arise from a combination of a preparation and a measurement. To match the constraints of information retrieval tasks, memory is bounded by the dimension of the message string. This is further motivated by the Holevo bound [Hol73], according to which one can faithfully retrieve no more than  $n$  dits of classical information from  $n$  qudits.

In a bounded-memory model, the hidden variable is restricted to take values in  $\mathbb{Z}_d$  with  $d$  fixed by the communication scenario. A state preparation  $P$  is modelled by a probability distribution  $p_\pi(\cdot | P)$  over the hidden-variable space  $\mathbb{Z}_d$ . Similarly, a measurement  $M$  is modelled

by a family of probability distributions  $\{p_\mu(\cdot \mid \lambda, M)\}_\lambda$  over the outcome space, which, to match the communication scenario, is also  $\mathbb{Z}_d$ . A hidden-variable model—e.g. for a prepare-and-measure sequence of operations  $M_q \circ P_{x,z}$  with inputs  $x, z \in \mathbb{Z}_d$  and a question  $q \in Q$ , as in Fig. 5.3—simulates an empirical behaviour as

$$p_e(\cdot \mid x, z, q) = \sum_\lambda p_\mu(\cdot \mid \lambda, M_q) \cdot p_\pi(\lambda \mid P_{x,z}). \quad (5.18)$$

Each operation is thus modelled in sequence as an operation on the hidden-variable space.

With Eq. (5.18) in mind, the bounded-memory classical strategies of Section 5.3 can also be interpreted as bounded-memory hidden-variable models themselves. To see this, note that in Eq. (5.12) encoding corresponds to hidden-variable preparation and decoding to hidden-variable measurement.

The above description makes contact with the strategies and empirical behaviours of Section 5.3 and it will be convenient for the remainder of this section to use a simplified notation that meets the one from [MK18]. Let us work in the real vector space  $\mathbb{Z}_d$  with basis given by the hidden-variable states. For all preparations  $P_{x,z}$ , the probability distribution  $p_\pi(\cdot \mid P_{x,z})$  will be more concisely denoted as a probability vector  $\boldsymbol{\lambda}_{x,z}$  in  $\mathbb{Z}_d$ . Deterministic encoding will result in vectors  $\boldsymbol{\lambda}_{x,z}$  indicating in which partition  $x, z$  belong to. For measurements, we express  $\{p_\mu(\cdot \mid \lambda, M)\}_\lambda$  more concisely as, for each question, a choice of a  $d \times d$  left stochastic matrix<sup>1</sup>  $T_q$  acting on  $\boldsymbol{\lambda}_{x,z}$ . If we also denote by  $e_{x,z,q} := p_e(\cdot \mid x, z, q)$  the empirical probability vector over outcomes  $\mathbb{Z}_d$ , then Eq. (5.18) can be rewritten for each outcome  $c \in \mathbb{Z}_d$  in simplified notation as the dot product

$$e_{x,z,q}(c) = T_q \boldsymbol{\lambda}_{x,z} \cdot \boldsymbol{\delta}_c, \quad (5.19)$$

where  $\boldsymbol{\delta}_c$  is the  $d$ -dimensional vector filled with zeros except for the  $c^{\text{th}}$  coordinate which equals 1.

As mentioned in Subsection 5.3.2, it is sufficient to consider deterministic encodings and decodings to achieve optimal strategies. As in [SHP19], if the message is of dimension  $d$ , if there are  $|Q|$  possible questions, and if the outcome is a dit, then there are  $d^{|Q|d}$  different deterministic decoding strategies. This number can be calculated by enumerating the possible choices of  $T_q$  for each question. There are  $d^d$  possible binary left stochastic matrices hence  $d^{d|Q|}$  deterministic decoding strategies.

As our focus is on prepare-and-measure scenarios, we have not discussed hidden-variable modelling of transformations. This can be found in [MK18]. For our main example of an information retrieval task we have also focused on the prepare-and-measure version of the Torpedo Game. Note, however, that it can be equivalently expressed in a sequential scenario with fixed preparation and measurement.

<sup>1</sup>A left-stochastic matrix is a square matrix with each column summing to 1. It is furthermore binary if each column consists of a single 1 and 0 elsewhere.

### 5.4.2 Sequential contextuality in information retrieval tasks

Here we introduce a different notion of contextuality than the one presented in Section 1.5. It concerns the sequence of operations that might be applied to a system. An empirical behaviour is sequential noncontextual if it admits a hidden-variable model that: (i) preserves a modular sequential description of operations, and (ii) the hidden-variable representation of operations is context-independent. We refer readers to [MK18] for a detailed presentation of this concept.

These assumptions have been implicitly built into the above definition of hidden-variable models displayed in Eq. (5.19). For (i), note that each operation has an individual description at the hidden-variable level. For example, to obtain predictions for a prepare-and-measure experiment we compose the individual hidden-variable descriptions of the preparation and of the measurement, as in Eq. (5.19). And for (ii), note for example that regardless of which context the preparation  $P_{x,z}$  appears in it should be modelled by the same vector  $\lambda_{x,z}$ . One could relax these assumptions, in which case it would become trivial to find a hidden-variable model for any behaviour, but it would also entail giving up the intuitive sense of what the model means.

If an empirical behaviour does not admit a sequential noncontextual hidden-variable model it is said to be sequential contextual. In this chapter we will only be considering sequential contextuality with respect to bounded-memory models, though the definition may be applied more generally.

A useful intuition for sequential contextuality is that, within the memory constraints, for any faithful model of the behaviour, the whole (the description of the context) is more than the composition of its parts (the descriptions of the individual operations). A contextual model would always involve additional memory and communication to track the context, which would be outside of the constraints of the task – involving, e.g., a contextuality demon analogous to Maxwell’s demon in thermodynamics. Indeed it was shown in [HCB<sup>+</sup>18] that a related characteristic incurs a simulation cost as measured by Landauer erasure.

### 5.4.3 Quantified Contextual Advantage in Information Retrieval Tasks

The following proposition can be understood as a no-go theorem stating that winning the Torpedo Game deterministically for  $d = 2$  and  $d = 3$  is incompatible with the assumptions of sequential noncontextuality and bounded memory. If such a performance is observed then one is forced to abandon at least one of the assumptions, and we note that the Holevo bound gives an argument that perhaps the noncontextuality assumption is the weaker of these.

**Proposition 5.2.** *For  $d = 2$  and  $d = 3$ , strong sequential contextuality with respect to bounded memory is necessary and sufficient to win the Torpedo Game deterministically.*

*Proof.* Suppose a bounded-memory hidden-variable model realises an empirical model that wins the Torpedo Game deterministically. Input-question combinations  $(x, z, q)$  label the contexts.

Recall that the winning relation is  $\omega_q(x, z)$ , and the winning condition for the Torpedo Game is

$$p_e(c \notin \omega_q(x, z) \mid x, z, q) = 0. \quad (5.20)$$

Using notations introduced in Subsection 5.4.1, the hidden-variable model must specify probability vectors  $\{\lambda_{x,z}\}_{x,z \in \mathbb{Z}_d}$  and left stochastic matrices  $\{T_q\}_{q \in Q}$  such that

$$T_q \lambda_{x,z} \cdot \delta_c = 0, \quad (5.21)$$

for all  $x, z \in \mathbb{Z}_d, \forall q \in Q$  and  $c \notin \omega_q(x, z)$ .

As mentioned in Subsection 5.3.2, it suffices to consider deterministic strategies. Eq. (5.21) reduces to a set of binary linear equations (36 equations for  $d = 3$ ) that any sequentially noncontextual realisation must jointly satisfy.

This cannot be possible since it would provide a perfect classical strategy for the  $d = 2$  and  $d = 3$  Torpedo Games, violating the optimal bounds Eq. (5.15) that were obtained by exhaustive search. On the other hand, it is always possible to obtain a contextual realisation, by taking context-wise solutions to Eq. (5.21): e.g. where the choice of  $\lambda_{x,z}$  is not only a function of  $x$  and  $z$ , but also of  $q$ .

It can further be observed that if any fraction of an empirical model  $e$  can be described noncontextually, i.e.  $\text{NCF}(e) = p > 0$ , then with an average probability at least  $p$  the empirical model  $e$  fails in the Torpedo Game. Therefore, to win the Torpedo Game deterministically requires strong contextuality. □

An explicit noncontextual memory-bounded hidden-variable model that fails to fully realise the empirical predictions but that satisfies the maximum of 33 out of 36 constraints—thus reaching the maximum probability  $\frac{11}{12}$  for the Torpedo Game with input trits—from Eq. (5.21) for  $d = 3$  is the following. The state vectors are

$$\lambda_{0,0} = \lambda_{0,1} = \lambda_{1,1} = \delta_0,$$

$$\lambda_{1,0} = \lambda_{0,2} = \lambda_{2,2} = \delta_1,$$

$$\lambda_{2,0} = \lambda_{2,1} = \lambda_{1,2} = \delta_2.$$

The measurement left-stochastic matrices are

$$T_\infty = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, T_0 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, T_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, T_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

This corresponds to the strategy depicted in Fig. 5.8 with the green section corresponding to sending the message  $\delta_0$ , the blue one to  $\delta_1$  and the red one to  $\delta_2$ .

We also obtain the following more general result, of which Proposition 5.2 is a special case. This is an extension of [ABM17, Theorem 3] to the case of information retrieval tasks with sequential contextuality rather than measurement contextuality.

**Theorem 5.1.** *Given any information retrieval task and strategy with empirical behaviour  $e$ ,*

$$\varepsilon \geq \text{NCF}(e) \nu$$

*where  $\varepsilon$  is the probability of failure, averaged over inputs and questions,  $\text{NCF}(e)$  is the bounded-memory noncontextual fraction of  $e$  with memory size  $d$  fixed by the scenario, and  $\nu := 1 - \theta^C$  measures the hardness of the task,  $\theta^C$  being the classical value.*

*Proof.* The empirical behaviour can be decomposed as:

$$e = \text{NCF}(e)e^{\text{NC}} + \text{CF}(e)e'$$

where  $e'$  is necessarily strongly contextual. From this convex decomposition, we obtain that the probability of success using the empirical model  $e$  reads:

$$p_{S,e} = \text{NCF}(e)p_{S,e^{\text{NC}}} + \text{CF}(e)p_{S,e'}$$

where  $p_{S,e^{\text{NC}}}$  and  $p_{S,e'}$  are the average probabilities associated with empirical models  $e^{\text{NC}}$  and  $e'$  respectively. At best,  $e'$  wins with probability 1 and thus:

$$\begin{aligned} p_{S,e} &\leq \text{NCF}(e)p_{S,e^{\text{NC}}} + \text{CF}(e) \\ \varepsilon &\geq \text{NCF}(e)\varepsilon_{e^{\text{NC}}} \end{aligned}$$

and  $\varepsilon_{e^{\text{NC}}} = 1 - p_{S,e^{\text{NC}}}$  is the average probability of failure associated to  $e^{\text{NC}}$ . Since  $e^{\text{NC}}$  is noncontextual, we know that the minimum probability of failure is  $\nu = 1 - \theta^C$ , where  $\theta^C$  is the classical value of the game. Then  $\varepsilon_{e^{\text{NC}}} \geq \nu$ , from which we obtain the desired inequality:

$$\varepsilon \geq \text{NCF}(e)\nu$$

□

This provides a quantifiable relationship between quantum advantage and sequential contextuality. Inequalities of this form are also known to arise for a variety of other informational tasks that admit quantum advantage, with hardness measures and notions of non-classicality adapted to the particular task [ABM17, MK18, Wes18].

## 5.5 Discussion and open problems

We have formalised a class of information retrieval tasks in communication scenarios, of which the much-studied problem of (quantum) random access coding is a special case. We showed that quantum-over-classical advantage is explained by quantum (sequential) contextuality. We have identified a distinct information retrieval task that we have presented as the Torpedo Game, which admits a greater quantum-over-classical advantage than the comparable QRAC for

qutrits by exploiting Wigner negativity. Remarkably, the qutrit torpedo strategy is maximally contextual, meaning that no fraction of it can be explained by an underlying noncontextual model. By choosing measurements that are associated with a particular discrete Wigner function, Wigner negativity seems a necessary ingredient for a quantum strategy to perform better than a classical one as the optimal quantum strategy relies on it. However post-quantum strategies (e.g. using phase point operators) might be optimal while remaining Wigner positive. A more thorough investigation of the precise relationship between negativity and advantage is an interesting direction to pursue.

To obtain perfect quantum strategies for the Torpedo Game we have derived a prepare-and-measure scenario for which quantum mechanics exhibits logically paradoxical behaviour (with respect to noncontextual hidden variable assumptions). More generally, we have identified this as a characteristic that quantifies quantum advantage for any bounded-memory information retrieval task.

In the specific case of random access coding, some works have imposed obliviousness constraints as part of the task as opposed to bounded-memory. These restrict what information the receiver can be allowed to infer about the input string. Whereas preparation contextuality is known to be necessary and sufficient for quantum advantage in oblivious tasks [HTMB17, SHP19], we have shown sequential contextuality to be necessary and sufficient characteristic for bounded-memory tasks.

Importantly, a nonlocal version of the Torpedo Game was derived after our corresponding paper in [HXG<sup>+</sup>21], where it was also implemented experimentally. This example was employed to certify quantum correlations allowing to test for nonlocality, steering and quantum state tomography in a single experiment. Based on this nonlocality experiment, entanglement in prepare-and-measure scenarios was further studied in [PTWP21].

We only restricted the analysis to dimensions 2 and 3 because as the dimension increases, the Torpedo Game gets easier as the number of right answers to each questions grows linearly with the dimension (there are  $d - 1$  correct answers in dimension  $d$ ). The classical optimum can only be established rigorously for small dimensions, owing to the proliferation of possible hidden variable assignments as the dimension increases. We have, however, found perfect classical strategies, i.e. strategies that win with probability 1, for  $d = 4$  up to  $d = 23$  by randomly generating grids (see Figure 5.9 for an example in dimension 5). We thus conjecture that the Torpedo Game has a perfect classical strategy for dimension 4 and above.

However, we briefly comment on a possible generalisation to higher dimensions. In order to re-instate a quantum-over-classical advantage, as we had in dimensions two and three, we may modify the Torpedo game to make it harder to win classically. Note that the following modifications have no effect on the quantum values, which remain  $\theta_{d \geq 3}^Q = 1$ . As previously noted in [Gro06, vDH11], the eigenvectors of phase point operators Eq. (5.3) are degenerate: a  $+1$  eigenspace of dimension  $\frac{d+1}{2}$  and a  $-1$  eigenspace of dimension  $\frac{d-1}{2}$ . Thus it is possible

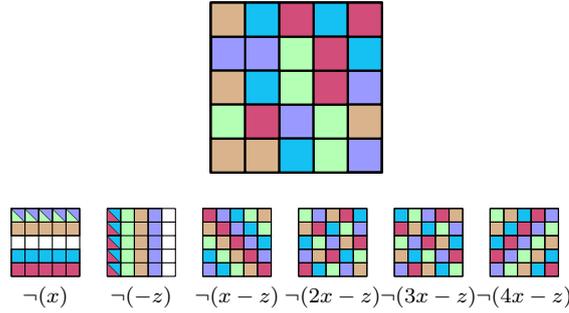


Figure 5.9: A perfect classical strategy for the  $d = 5$  Torpedo Game. As in Figure 5.8, cells of the same color belong to the same partition. The lines that avoid Alice are depicted below for every question Bob can be asked.

to enlarge Alice's input from  $d^2$  to  $\frac{d^2(d-1)}{2}$ . Formally, let  $0 \leq \ell < \frac{d-1}{2}$ , so Alice sends Bob  $|\psi_{x,z,\ell}\rangle = X^x Z^z (|\ell+1\rangle + |-(\ell+1)\rangle) / \sqrt{2}$ , instead of just  $|\psi_{x,z,\ell=0}\rangle := |\psi_{x,z}\rangle$  as before. The modification changes a single relation from  $w_\infty(x, z) = \{a \in \mathbb{Z}_d \mid a \neq x\}$  to

$$w_\infty(x, z, \ell) = \{a \in \mathbb{Z}_d \mid a \in \{x + \ell + 1, x - \ell - 1\}\},$$

whereas the remaining conditions persist i.e.,  $w_q(x, z, \ell) = w_q(x, z)$  in Eq. (5.2) for  $q \in \{0, 1, \dots, d-1\}$ . It seems reasonable that such a game, with more restrictive winning conditions, should be harder to win classically. Indeed, we were unable to find any perfect classical strategy by sampling, although we cannot rule out its existence since we were unable to exhaustively check all classical strategies. More generally, we have motivated how our perfect quantum strategies for this information retrieval task arise from a remarkable geometric feature of maximally negative states (c.f. Eq. (5.10)), and we expect that this insight can be further mined for obtaining other examples of quantum advantage.

## CONCLUSION AND OUTLOOK

Since the seminal papers of Feynman [Fey82] and Deutsch [Deu85], quantum information scientists have asked some fundamental questions: are there computational problems that are provably hard for classical computers but which are efficiently solvable on a quantum computer? what really makes a quantum computer non-classical? what is the best route towards scalable fault-tolerant quantum computers? Unable to answer any of these, in this thesis we have considered how contextuality and Wigner negativity play a fundamental role as non-classical features.<sup>2</sup>

In this dissertation, we have introduced a robust framework for contextuality in continuous variables along the lines of the discrete-variable one introduced in [AB11]. The Fine–Abramsky–Brandenburger (FAB) theorem extends to continuous-variables, even in the case where there are an uncountably infinite set of measurement labels. The contextual fraction [ABM17] can be defined with infinite-dimensional linear programming and it can be approximated through Lasserre-type of semidefinite relaxations [Las09].

Thanks to the FAB theorem, we have generalised the seminal result from [HWVE14]—namely that Wigner negativity corresponds to measurement contextuality with respect to Pauli measurements—to the continuous-variable realm. These first two contributions have closed a portion of the gap between discrete variables and continuous variables by showing how some important concepts and results extend between them.

We have introduced a reliable method for witnessing Wigner negativity that can be efficiently implemented experimentally. Each witness is associated with a threshold value that can be computed using infinite-dimensional linear programming. Relaxing and restricting these programs yield two converging hierarchies of semidefinite programs that provide tighter and tighter upper and lower bounds on the threshold values for increasing ranks in the hierarchies. Importantly, we have shown how well the hierarchies perform numerically on realistic examples.

We have then sought to extend the range of tasks for which non-classicality is known to be a useful resource, by tying it to and using it to design information retrieval tasks with a quantum-over-classical advantage We have introduced a game—the Torpedo Game—where a quantum advantage can be derived for finite-dimensional systems of dimensions 2 and 3. Crucially, the optimal quantum strategies make use of quantum states with maximally negative

---

<sup>2</sup>These ironic few last sentences are a pastiche of the introduction in [RT21].

discrete Wigner functions. To pinpoint the source of quantum advantage, we have shown that, subject to an assumption of bounded memory, sequential contextuality is necessary and sufficient for quantum advantage in any information retrieval task and in particular the Torpedo Game. A nonlocal version of the Torpedo Game was later derived in [HXG<sup>+</sup>21], where it was experimentally tested to certify quantum correlations, allowing to test for nonlocality, steering and quantum state tomography in a single experiment.

As shown in this dissertation, optimisation theory can lead to interesting results that help answering questions arising in quantum information. It has also been used to obtain information about the ground-state energy of a Hamiltonian [BH13] or to obtain bounds on the maximum quantum violations of Bell inequalities [NFAV15]. It is natural to wonder what other problems from quantum information might benefit from optimisation theory and in particular from nonlinear polynomial optimisation with noncommutative variables [NPA07, KMV20].

Perhaps a less obvious outlook is the converse question, that is to wonder whether optimisation theory may benefit from quantum information from a theoretical point of view<sup>3</sup>. In order to solve a Global Moment Problem (see (GMP)) for *sparse instances* with the Lasserre hierarchy, the original problem is split into several sub-problems [WLT18, KMP21]. Solving a sparse GMP thus amounts to solving smaller instances of the problem i.e. finding optimal measures on reduced space and then combining the solutions into a global solution. Of course this can only be implemented if we are certified that the solutions of the smaller problems can be combined—or glued—consistently. In words used throughout this dissertation, this requires that the optimal measures on sub-problems can be extended consistently to a global probability measure. Because the smaller optimisation problems can pick out any solution of the corresponding feasible space, being able to glue them consistently requires that the splitting into sub-problems corresponds to a noncontextual scenario for otherwise a consistent global measure may not exist. Interestingly, the problem of gluing measures consistently was first introduced by Vorob’ev [Vor62] and of course it bears a close relationship to noncontextuality [Bar14]. It is also extensively studied for optimal transport [Vil09].

It is expected that results linking contextuality and/or Wigner negativity to advantage will continue to emerge. We have provided some answers in this dissertation, but it will be interesting to see what other tools, results, and insights can be transferred from discrete variables to continuous variables and vice versa.

The equivalence between contextuality and Wigner negativity in continuous variables for at least two systems was shown using all possible Weyl measurements. Of course the set of all possible Weyl measurements is a continuum and one has to wonder if the equivalence still holds in a realistic experimental setting where only a finite number of measurements can be implemented. One problem that may arise is that the fact that the Wigner function is the

<sup>3</sup>Of course having a quantum computer might help greatly for solving some optimisation problems but this is not what we are wondering here.

unique quasiprobability distribution yielding the correct marginal distributions can only be established using all possible Weyl measurements [BB87, BG20].

From this equivalence, it is also very tempting to understand the precise relationship between measures of quantumness. In particular, is there a direct link between the negativity volume and the contextual fraction? This dissertation may provide a jumping off point for understanding how to derive a general resource theory of non-classicality along the lines of the resource theory of contextuality presented in [ABKM19], for which all of these measures would be non-classicality monotones.

Also we know that the violation of the threshold value associated to a Wigner negativity witness introduced in Chapter 4 provide a lower bound on the distance to the set of Wigner positive states. One may wonder if there is a link between this distance and the contextual fraction so that the aforementioned violation may provide a lower bound on the contextual fraction. This dissertation may also be useful for understanding how to derive a general resource theory of non-classicality.

Different and more fine-grained notions of contextuality or non-classicality can be useful for example in obtaining an ever finer understanding of quantum advantages. Sequential contextuality is one example that we have shown to provide a good characterisation of advantage in information retrieval. We expect that this this notion and other finer or more generalised notions will will continue to give rise to communication or computation tasks with a quantum-over-classical advantage.

Finally, it is still an open question to understand precisely and quantifiably what is the non-classical phenomenon at the root of recent, and much mediatised, demonstrations of quantum computational advantage [AAB<sup>+</sup>19, ZWD<sup>+</sup>20]. This dissertation might provide partial answers to start answering this very challenging question.

# BIBLIOGRAPHY

- [AAB<sup>+</sup>19] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, et al.  
Quantum supremacy using a programmable superconducting processor.  
*Nature*, 574(7779):505–510, 2019.  
[doi:10.1117/12.2603523](https://doi.org/10.1117/12.2603523).  
(page 182)
- [AB09] J. Anders and D. E. Browne.  
Computational power of correlations.  
*Physical Review Letters*, 102:050502, Feb 2009.  
[doi:10.1103/PhysRevLett.102.050502](https://doi.org/10.1103/PhysRevLett.102.050502).  
(page 160)
- [AB11] S. Abramsky and A. Brandenburger.  
The sheaf-theoretic structure of non-locality and contextuality.  
*New Journal of Physics*, 13(11):113036, 2011.  
[doi:10.1088/1367-2630/13/11/113036](https://doi.org/10.1088/1367-2630/13/11/113036).  
(pages 2, 3, 4, 5, 35, 39, 40, 42, 43, 46, 48, 51, 54, 72, 73, 75, 76, 160, 172, 173, and 180)
- [ABC08] D. Appleby, I. Bengtsson, and S. Chaturvedi.  
Spectra of phase point operators in odd prime dimensions and the extended Clifford group.  
*Journal of mathematical physics*, 49(1):012102, 2008.  
[doi:10.1063/1.2824479](https://doi.org/10.1063/1.2824479).  
(page 166)
- [ABC<sup>+</sup>19] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai.  
Parity oblivious d-level random access codes and class of noncontextuality inequalities.  
*Quantum Information Processing*, 18(4):111, 2019.  
[doi:10.1007/s11128-019-2228-3](https://doi.org/10.1007/s11128-019-2228-3).  
(page 160)
- [ABdSZ17] S. Abramsky, R. S. Barbosa, N. de Silva, and O. Zapata.  
The quantum monad on relational structures.  
In K. G. Larsen, H. L. Bodlaender, and J.-F. Raskin, editors, *42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)*, volume 83 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 35:1–35:19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.  
[doi:10.4230/LIPIcs.MFCS.2017.35](https://doi.org/10.4230/LIPIcs.MFCS.2017.35).  
(pages 3 and 74)
- [ABK<sup>+</sup>15] S. Abramsky, R. S. Barbosa, K. Kishida, R. Lal, and S. Mansfield.  
Contextuality, cohomology and paradox.  
In S. Kreutzer, editor, *24th EACSL Annual Conference on Computer Science Logic (CSL 2015)*, volume 41 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 211–228. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2015.  
[doi:10.4230/LIPIcs.CSL.2015.211](https://doi.org/10.4230/LIPIcs.CSL.2015.211).  
(pages 35, 73, and 74)
- [ABKM19] S. Abramsky, R. S. Barbosa, M. Karvonen, and S. Mansfield.

- A comonadic view of simulation and quantum resources, 2019.  
[doi:10.1109/lics.2019.8785677](https://doi.org/10.1109/lics.2019.8785677).  
To appear in Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2019).  
(pages 6, 54, 55, and 182)
- [ABM17] S. Abramsky, R. S. Barbosa, and S. Mansfield.  
Contextual fraction as a measure of contextuality.  
*Physical Review Letters*, 119(5):050504, 2017.  
[doi:10.1103/PhysRevLett.119.050504](https://doi.org/10.1103/PhysRevLett.119.050504).  
(pages 3, 6, 26, 35, 43, 45, 54, 55, 57, 64, 65, 85, 172, 176, 177, and 180)
- [ABMP18] E. A. Aguilar, J. J. Borkala, P. Mironowicz, and M. Pawłowski.  
Connections between mutually unbiased bases and quantum random access codes.  
*Physical Review Letters*, 121(5):050501, 2018.  
[doi:10.1103/PhysRevLett.121.050501](https://doi.org/10.1103/PhysRevLett.121.050501).  
(pages 159 and 165)
- [Abr13a] S. Abramsky.  
Relational databases and Bell’s theorem.  
In V. Tannen, L. Wong, L. Libkin, W. Fan, W.-C. Tan, and M. Fourman, editors, *In search of elegance in the theory and practice of computation: Essays dedicated to Peter Buneman*, volume 8000 of *Lecture Notes in Computer Science*, pages 13–35. Springer, 2013.  
[doi:10.1007/978-3-642-41660-6\\_2](https://doi.org/10.1007/978-3-642-41660-6_2).  
(page 74)
- [Abr13b] S. Abramsky.  
Relational hidden variables and non-locality.  
*Studia Logica*, 101(2):411–452, 2013.  
[doi:10.1007/s11225-013-9477-4](https://doi.org/10.1007/s11225-013-9477-4).  
(page 72)
- [Abr15] S. Abramsky.  
Contextual semantics: From quantum mechanics to logic, databases, constraints, and complexity.  
In E. Dzhamfarov, S. Jordan, R. Zhang, and V. Cervantes, editors, *Contextuality from Quantum Physics to Psychology*, volume 6 of *Advanced Series on Mathematical Psychology*, pages 23–50. World Scientific, 2015.  
[doi:10.1142/9789814730617\\_0002](https://doi.org/10.1142/9789814730617_0002).  
(page 73)
- [ABS<sup>+</sup>15] A. Asadian, C. Budroni, F. E. Steinhoff, P. Rabl, and O. Gühne.  
Contextuality in phase space.  
*Physical Review Letters*, 114(25):250403, 2015.  
[doi:10.1103/PhysRevLett.114.250403](https://doi.org/10.1103/PhysRevLett.114.250403).  
(page 47)
- [ADR82] A. Aspect, J. Dalibard, and G. Roger.  
Experimental test of bell’s inequalities using time-varying analyzers.  
*Physical Review Letters*, 49:1804–1807, Dec 1982.  
[doi:10.1103/PhysRevLett.49.1804](https://doi.org/10.1103/PhysRevLett.49.1804).  
(page 2)
- [ADVT09] P. Albini, E. De Vito, and A. Toigo.  
Quantum homodyne tomography as an informationally complete positive-operator-valued measure.  
*Journal of Physics A: Mathematical and Theoretical*, 42(29):295302, 2009.  
[doi:10.1088/1751-8113/42/29/295302](https://doi.org/10.1088/1751-8113/42/29/295302).  
(page 76)
- [AFLS15] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz.  
A combinatorial approach to nonlocality and contextuality.  
*Communications in Mathematical Physics*, 334(2):533–628, 2015.

- [doi:10.1007/s00220-014-2260-1](https://doi.org/10.1007/s00220-014-2260-1).  
(page 46)
- [AGA<sup>+</sup>12] L. Aolita, R. Gallego, A. Acín, A. Chiuri, G. Vallone, P. Mataloni, and A. Cabello.  
Fully nonlocal quantum correlations.  
*Physical Review A*, 85:032107, Mar 2012.  
[doi:10.1103/PhysRevA.85.032107](https://doi.org/10.1103/PhysRevA.85.032107).  
(page 43)
- [AGK13] S. Abramsky, G. Gottlob, and P. G. Kolaitis.  
Robust constraint satisfaction and local hidden variables in quantum mechanics.  
In F. Rossi, editor, *23rd International Joint Conference on Artificial Intelligence (IJCAI 2013)*, pages 440–446. AAAI Press, 2013.  
(page 74)
- [AGPF18] F. Albarelli, M. G. Genoni, M. G. Paris, and A. Ferraro.  
Resource theory of quantum non-Gaussianity and Wigner negativity.  
*Physical Review A*, 98(5):052350, 2018.  
[doi:10.1103/PhysRevA.98.052350](https://doi.org/10.1103/PhysRevA.98.052350).  
(pages 3 and 20)
- [AH12] S. Abramsky and L. Hardy.  
Logical Bell inequalities.  
*Physical Review A*, 85(6):062114, 2012.  
[doi:10.1103/PhysRevA.85.062114](https://doi.org/10.1103/PhysRevA.85.062114).  
(page 74)
- [AMB12] S. Abramsky, S. Mansfield, and R. S. Barbosa.  
The cohomology of non-locality and contextuality.  
In B. Jacobs, P. Selinger, and B. Spitters, editors, *Proceedings of 8th International Workshop on Quantum Physics and Logic (QPL 2011)*, volume 95 of *Electronic Proceedings in Theoretical Computer Science*, pages 1–14. Open Publishing Association, 2012.  
[doi:10.4204/EPTCS.95.1](https://doi.org/10.4204/EPTCS.95.1).  
(page 73)
- [ANTSV99] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani.  
Dense quantum coding and a lower bound for 1-way quantum automata.  
In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 376–383, 1999.  
[doi:10.1145/301250.301347](https://doi.org/10.1145/301250.301347).  
(pages 159 and 161)
- [ApS19] M. ApS.  
*MOSEK Optimizer API for Python 9.2.36*, 2019.  
URL <https://docs.mosek.com/9.2/pythonapi/index.html>.  
(pages 28 and 99)
- [ARBC09] E. Amsalem, M. Rådmark, M. Bourennane, and A. Cabello.  
State-independent quantum contextuality with single photons.  
*Physical Review Letters*, 103(16):160405, 2009.  
(page 172)
- [ARL14] G. Adesso, S. Ragy, and A. R. Lee.  
Continuous variable quantum information: Gaussian states and beyond.  
*Open Systems & Information Dynamics*, 21(01n02):1440001, 2014.  
[doi:10.1142/s1230161214400010](https://doi.org/10.1142/s1230161214400010).  
(pages 15, 16, and 76)
- [AS14a] S. Abramsky and M. Sadrzadeh.  
Semantic unification.  
In *Categories and Types in Logic, Language, and Physics*, pages 1–13. Springer, 2014.  
[doi:10.1007/978-3-642-54789-8\\_1](https://doi.org/10.1007/978-3-642-54789-8_1).  
(page 2)

- [AS14b] S. Abramsky and M. Sadrzadeh.  
Semantic unification: A sheaf-theoretic approach to natural language.  
In C. Casadio, B. Coecke, M. Moortgat, and P. Scott, editors, *Categories and Types in Logic, Language, and Physics: Essays dedicated to Jim Lambek on the occasion of his 90th birthday*, pages 1–13. Springer, 2014.  
[doi:10.1007/978-3-642-54789-8\\_1](https://doi.org/10.1007/978-3-642-54789-8_1).  
(pages 2 and 73)
- [Bar02] A. Barvinok.  
*A course in convexity*, volume 54 of *Graduate Studies in Mathematics*.  
American Mathematical Society, 2002.  
[doi:10.1090/gsm/054](https://doi.org/10.1090/gsm/054).  
(pages 25, 27, 57, 59, 110, and 111)
- [Bar07] J. Barrett.  
Information processing in generalized probabilistic theories.  
*Physical Review A*, 75:032304, Mar 2007.  
[doi:10.1103/PhysRevA.75.032304](https://doi.org/10.1103/PhysRevA.75.032304).  
(page 173)
- [Bar14] R. S. Barbosa.  
On monogamy of non-locality and macroscopic averages: examples and preliminary results.  
In B. Coecke, I. Hasuo, and P. Panangaden, editors, *Proceedings of the 11th workshop on Quantum Physics and Logic, QPL 2014, Kyoto, Japan, 4-6th June 2014*, volume 172 of *EPTCS*, pages 36–55, 2014.  
[doi:10.4204/EPTCS.172.4](https://doi.org/10.4204/EPTCS.172.4).  
(page 181)
- [Bar15] R. S. Barbosa.  
*Contextuality in quantum mechanics and beyond*.  
PhD thesis, University of Oxford, 2015.  
(page 74)
- [BAV<sup>+</sup>21] J. E. Bourassa, R. N. Alexander, M. Vasmer, A. Patil, I. Tzitrin, T. Matsuura, D. Su, B. Q. Baragiola, S. Guha, G. Dauphinais, K. K. Sabapathy, N. C. Menicucci, and I. Dhand.  
Blueprint for a Scalable Photonic Fault-Tolerant Quantum Computer.  
*Quantum*, 5:392, Feb. 2021.  
[doi:10.22331/q-2021-02-04-392](https://doi.org/10.22331/q-2021-02-04-392).  
(page 3)
- [BB87] J. Bertrand and P. Bertrand.  
A tomographic approach to wigner’s function.  
*Foundations of Physics*, 17(4):397–405, 1987.  
[doi:10.1007/BF00733376](https://doi.org/10.1007/BF00733376).  
(pages 75 and 182)
- [BBC12] I. Bengtsson, K. Blanchfield, and A. Cabello.  
A Kochen–Specker inequality from a sic.  
*Physics Letters A*, 376(4):374–376, 2012.  
[doi:10.1016/j.physleta.2011.12.011](https://doi.org/10.1016/j.physleta.2011.12.011).  
(page 165)
- [BBV04] S. Boyd, S. P. Boyd, and L. Vandenberghe.  
*Convex optimization*.  
Cambridge university press, 2004.  
[doi:10.1017/cbo9780511804441](https://doi.org/10.1017/cbo9780511804441).  
(pages 23, 24, and 25)
- [BCE21] R. I. Booth, U. Chabaud, and P.-E. Emeriau.  
Equivalence between contextuality and Wigner negativity for generalised continuous-variable quadrature measurements.  
in preparation, 2021.

- (pages 7 and 76)
- [BDE<sup>+</sup>19] R. S. Barbosa, T. Douce, P.-E. Emeriau, E. Kashefi, and S. Mansfield.  
Continuous-variable nonlocality and contextuality.  
Preprint [arXiv:1905.082671 \[quant-ph\]](https://arxiv.org/abs/1905.082671), 2019.  
(pages 7 and 47)
- [Bel64] J. S. Bell.  
On the Einstein Podolsky Rosen paradox.  
*Physics Physique Fizika*, 1(3):195–200, Nov. 1964.  
[doi:10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195).  
Publisher: American Physical Society.  
(pages 2, 3, 5, 35, and 172)
- [Bel66] J. S. Bell.  
On the problem of hidden variables in quantum mechanics.  
*Reviews of Modern Physics*, 38(3):447–452, 1966.  
[doi:10.1103/RevModPhys.38.447](https://doi.org/10.1103/RevModPhys.38.447).  
(pages 3 and 172)
- [BG20] A. Blass and Y. Gurevich.  
Negative probabilities: What they are and what they are for.  
Preprint [arXiv:2009.10552 \[quant-ph\]](https://arxiv.org/abs/2009.10552), 2020.  
(pages 75, 76, and 182)
- [BH13] F. G. Brandao and A. W. Harrow.  
Product-state approximations to quantum ground states.  
In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pages 871–880, 2013.  
[doi:10.1145/2488608.2488719](https://doi.org/10.1145/2488608.2488719).  
(page 181)
- [Bil08] P. Billingsley.  
*Probability and measure*.  
John Wiley & Sons, 2008.  
[doi:10.2307/2987970](https://doi.org/10.2307/2987970).  
(pages 20 and 84)
- [BK13] A. Brandenburger and H. J. Keisler.  
Use of a canonical hidden-variable space in quantum mechanics.  
In B. Coecke, L. Ong, and P. Panangaden, editors, *Computation, Logic, Games, and Quantum Foundations. The Many Facets of Samson Abramsky: Essays dedicated to Samson Abramsky on the occasion of his 60th birthday*, pages 1–6. Springer, 2013.  
[doi:10.1007/978-3-642-38164-5\\_1](https://doi.org/10.1007/978-3-642-38164-5_1).  
(pages 41 and 51)
- [BKM21] R. S. Barbosa, M. Karvonen, and S. Mansfield.  
Closing Bell: Boxing black box simulations in the resource theory of contextuality.  
Preprint [arXiv:2104.11241 \[quant-ph\]](https://arxiv.org/abs/2104.11241), 2021.  
(pages 36 and 39)
- [BKP06] J. Barrett, A. Kent, and S. Pironio.  
Maximally nonlocal and monogamous quantum correlations.  
*Physical Review Letters*, 97:170409, Oct 2006.  
[doi:10.1103/PhysRevLett.97.170409](https://doi.org/10.1103/PhysRevLett.97.170409).  
(page 43)
- [BKS<sup>+</sup>09] H. Bartosik, J. Klepp, C. Schmitzer, S. Sponar, A. Cabello, H. Rauch, and Y. Hasegawa.  
Experimental test of quantum contextuality in neutron interferometry.  
*Physical Review Letters*, 103:040403, Jul 2009.  
[doi:10.1103/PhysRevLett.103.040403](https://doi.org/10.1103/PhysRevLett.103.040403).  
(page 3)

- [Bog07] V. I. Bogachev.  
*Measure theory*.  
Springer, 2007.  
[doi:10.1007/978-3-540-34514-5](https://doi.org/10.1007/978-3-540-34514-5).  
(page 61)
- [BPA<sup>+</sup>08] N. Brunner, S. Pironio, A. Acín, N. Gisin, A. A. Méthot, and V. Scarani.  
Testing the dimension of Hilbert spaces.  
*Physical Review Letters*, 100:210503, May 2008.  
[doi:10.1103/PhysRevLett.100.210503](https://doi.org/10.1103/PhysRevLett.100.210503).  
(page 171)
- [BPA<sup>+</sup>19] B. Q. Baragiola, G. Pantaleoni, R. N. Alexander, A. Karanjai, and N. C. Menicucci.  
All-gaussian universality and fault tolerance with the Gottesman-Kitaev-Preskill code.  
*Physical Review Letters*, 123:200502, Nov 2019.  
[doi:10.1103/PhysRevLett.123.200502](https://doi.org/10.1103/PhysRevLett.123.200502).  
(page 86)
- [BSBN02] S. D. Bartlett, B. C. Sanders, S. L. Braunstein, and K. Nemoto.  
Efficient classical simulation of continuous variable quantum information processes.  
*Physical Review Letters*, 88:097904, 2002.  
[doi:10.1103/PhysRevLett.88.097904](https://doi.org/10.1103/PhysRevLett.88.097904).  
(page 4)
- [BTN01] A. Ben-Tal and A. Nemirovski.  
*Lectures on modern convex optimization: analysis, algorithms, and engineering applications*.  
SIAM, 2001.  
[doi:10.1137/1.9780898718829](https://doi.org/10.1137/1.9780898718829).  
(pages 24 and 25)
- [BVDB<sup>+</sup>17] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf.  
Contextuality as a resource for models of quantum computation with qubits.  
*Physical Review Letters*, 119(12):120505, 2017.  
[doi:10.1103/PhysRevLett.119.120505](https://doi.org/10.1103/PhysRevLett.119.120505).  
(pages 3 and 6)
- [BVL05] S. L. Braunstein and P. Van Loock.  
Quantum information with continuous variables.  
*Reviews of modern physics*, 77(2):513, 2005.  
[doi:10.1103/RevModPhys.77.513](https://doi.org/10.1103/RevModPhys.77.513).  
(pages 3 and 16)
- [BW98] K. Banaszek and K. Wódkiewicz.  
Nonlocality of the Einstein-Podolsky-Rosen state in the Wigner representation.  
*Physical Review A*, 58:4345–4347, Dec 1998.  
[doi:10.1103/PhysRevA.58.4345](https://doi.org/10.1103/PhysRevA.58.4345).  
(page 76)
- [BY08] A. Brandenburger and N. Yanofsky.  
A classification of hidden-variable properties.  
*Journal of Physics A: Mathematical and Theoretical*, 41(42):425302, 2008.  
[doi:10.1088/1751-8113/41/42/425302](https://doi.org/10.1088/1751-8113/41/42/425302).  
(page 50)
- [Car15] G. Carù.  
Detecting contextuality: Sheaf cohomology and all vs nothing arguments.  
Master’s thesis, University of Oxford, 2015.  
Available at <http://www.cs.ox.ac.uk/files/7608/Dissertation.pdf>.  
(page 73)
- [Car17] G. Carù.

- On the cohomology of contextuality.  
In R. Duncan and C. Heunen, editors, *13th International Conference on Quantum Physics and Logic (QPL 2016)*, volume 236 of *Electronic Proceedings in Theoretical Computer Science*, pages 21–39. Open Publishing Association, 2017.  
[doi:10.4204/EPTCS.236.2](https://doi.org/10.4204/EPTCS.236.2).  
(page 73)
- [Car18] G. Carù.  
Towards a complete cohomology invariant for non-locality and contextuality.  
Preprint [arXiv:1807.04203 \[quant-ph\]](https://arxiv.org/abs/1807.04203), 2018.  
(page 73)
- [CB18] L. Catani and D. E. Browne.  
State-injection schemes of quantum computation in Spekkens’ toy theory.  
*Physical Review A*, 98(5):052108, 2018.  
[doi:10.1103/PhysRevA.98.052108](https://doi.org/10.1103/PhysRevA.98.052108).  
(page 6)
- [CC17] L. Chakhmakhchyan and N. J. Cerf.  
Boson sampling with gaussian measurements.  
*Physical Review A*, 96:032326, Sep 2017.  
[doi:10.1103/PhysRevA.96.032326](https://doi.org/10.1103/PhysRevA.96.032326).  
(page 86)
- [CDG<sup>+</sup>20] U. Chabaud, T. Douce, F. Grosshans, E. Kashefi, and D. Markham.  
Building trust for continuous variable quantum states.  
In *15th Conference on the Theory of Quantum Computation, Communication and Cryptography*, 2020.  
[doi:10.4230/LIPIcs.TQC.2020.3](https://doi.org/10.4230/LIPIcs.TQC.2020.3).  
(pages 87, 88, 89, and 93)
- [CDM<sup>+</sup>17] U. Chabaud, T. Douce, D. Markham, P. van Loock, E. Kashefi, and G. Ferrini.  
Continuous-variable sampling from photon-added or photon-subtracted squeezed states.  
*Physical Review A*, 96:062307, Dec 2017.  
[doi:10.1103/PhysRevA.96.062307](https://doi.org/10.1103/PhysRevA.96.062307).  
(page 86)
- [CE] U. Chabaud and P.-E. Emeriau.  
Zeilberger’s algorithm and hierarchy of semidefinite programs.  
*Software Heritage* repository [swh:1:dir:d98f70e386783ef69bf8c2ecafdb7b328b19b7ec](https://sw.hub.rs/1:dir:d98f70e386783ef69bf8c2ecafdb7b328b19b7ec) containing the numerical tools developed for this 3rd chapter.  
(pages 94, 95, 98, 99, 100, and 128)
- [CEG21] U. Chabaud, P.-E. Emeriau, and F. Grosshans.  
Witnessing Wigner negativity.  
*Quantum*, 5:471, June 2021.  
[doi:10.22331/q-2021-06-08-471](https://doi.org/10.22331/q-2021-06-08-471).  
(pages 7, 89, and 100)
- [CFGM21] U. Chabaud, G. Ferrini, F. Grosshans, and D. Markham.  
Classical simulation of Gaussian quantum circuits with non-Gaussian input states.  
*Physical Review Research*, 3:033018, Jul 2021.  
[doi:10.1103/PhysRevResearch.3.033018](https://doi.org/10.1103/PhysRevResearch.3.033018).  
(page 4)
- [CFRD07] E. G. Cavalcanti, C. J. Foster, M. D. Reid, and P. D. Drummond.  
Bell inequalities for continuous-variable correlations.  
*Physical Review Letters*, 99:210405, Nov 2007.  
[doi:10.1103/PhysRevLett.99.210405](https://doi.org/10.1103/PhysRevLett.99.210405).  
(page 47)
- [CG69] K. E. Cahill and R. J. Glauber.

- Density operators and quasiprobability distributions.  
*Physical Review*, 177(5):1882, 1969.  
[doi:10.1103/PhysRev.177.1882](https://doi.org/10.1103/PhysRev.177.1882).  
(pages 18, 85, and 87)
- [CGaG<sup>+</sup>06] C. Cormick, E. F. Galvão, D. Gottesman, J. P. Paz, and A. O. Pittenger.  
Classicality in discrete Wigner functions.  
*Physical Review A*, 73:012301, Jan 2006.  
[doi:10.1103/PhysRevA.73.012301](https://doi.org/10.1103/PhysRevA.73.012301).  
(pages 19 and 159)
- [CGaS08] A. Casaccino, E. F. Galvão, and S. Severini.  
Extrema of discrete Wigner functions and applications.  
*Physical Review A*, 78:022310, Aug 2008.  
[doi:10.1103/PhysRevA.78.022310](https://doi.org/10.1103/PhysRevA.78.022310).  
(pages 162 and 165)
- [CGKM20] U. Chabaud, F. Grosshans, E. Kashefi, and D. Markham.  
Efficient verification of Boson Sampling.  
Preprint [arXiv:2006.03520 \[quant-ph\]](https://arxiv.org/abs/2006.03520), 2020.  
(pages 87, 90, 98, 142, and 158)
- [Cha21] U. Chabaud.  
Continuous variable quantum advantages and applications in quantum optics.  
PhD thesis [arXiv:2102.05227 \[quant-ph\]](https://arxiv.org/abs/2102.05227), Sorbonne University, 2021.  
(page 3)
- [CHSH69] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt.  
Proposed experiment to test local hidden-variable theories.  
*Physical Review Letters*, 23:880–884, Oct 1969.  
[doi:10.1103/PhysRevLett.23.880](https://doi.org/10.1103/PhysRevLett.23.880).  
(pages 37 and 160)
- [CKKS16] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora.  
Optimal bounds for parity-oblivious random access codes.  
*New Journal of Physics*, 18(4):045003, 2016.  
[doi:10.1088/1367-2630/18/4/045003](https://doi.org/10.1088/1367-2630/18/4/045003).  
(pages 159 and 160)
- [CMG20] U. Chabaud, D. Markham, and F. Grosshans.  
Stellar representation of non-Gaussian quantum states.  
*Physical Review Letters*, 124(6):063605, 2020.  
[doi:10.1103/PhysRevLett.124.063605](https://doi.org/10.1103/PhysRevLett.124.063605).  
(page 3)
- [COR<sup>+</sup>13] A. Crespi, R. Osellame, R. Ramponi, D. J. Brod, E. F. Galvao, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino.  
Integrated multimode interferometers with arbitrary designs for photonic boson sampling.  
*Nature photonics*, 7(7):545–549, 2013.  
[doi:10.1038/nphoton.2013.112](https://doi.org/10.1038/nphoton.2013.112).  
(page 3)
- [CRW<sup>+</sup>21] U. Chabaud, G. Roeland, M. Walschaers, F. Grosshans, V. Parigi, D. Markham, and N. Treps.  
Certification of non-Gaussian states with operational measurements.  
*PRX Quantum*, 2:020333, Jun 2021.  
[doi:10.1103/PRXQuantum.2.020333](https://doi.org/10.1103/PRXQuantum.2.020333).  
(pages 87, 90, 91, 93, 97, 98, and 99)
- [CSW14] A. Cabello, S. Severini, and A. Winter.  
Graph-theoretic approach to quantum correlations.  
*Physical Review Letters*, 112(4):040401, 2014.

- [doi:10.1103/PhysRevLett.112.040401](https://doi.org/10.1103/PhysRevLett.112.040401).  
(pages 3, 46, and 75)
- [DA18] C. Duarte and B. Amaral.  
Resource theory of contextuality for arbitrary prepare-and-measure experiments.  
*Journal of Mathematical Physics*, 59(6):062202, 2018.  
[doi:10.1063/1.5018582](https://doi.org/10.1063/1.5018582).  
(page 54)
- [DAGBR15] N. Delfosse, P. Allard Guerin, J. Bian, and R. Raussendorf.  
Wigner function negativity and contextuality in quantum computation on rebits.  
*Physical Review X*, 5(2):021003, 2015.  
[doi:10.1103/PhysRevX.5.021003](https://doi.org/10.1103/PhysRevX.5.021003).  
(pages 4, 6, and 75)
- [Deu85] D. Deutsch.  
Quantum theory, the church–turing principle and the universal quantum computer.  
*Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 400(1818):97–117, 1985.  
[doi:10.1098/rspa.1985.0070](https://doi.org/10.1098/rspa.1985.0070).  
(page 180)
- [DG06] M. A. De Gosson.  
*Symplectic geometry and quantum mechanics*, volume 166.  
Springer Science & Business Media, 2006.  
[doi:10.1007/3-7643-7575-2](https://doi.org/10.1007/3-7643-7575-2).  
(page 76)
- [Dic98] W. M. Dickson.  
*Quantum chance and non-locality: Probability and non-locality in the interpretations of quantum mechanics*.  
Cambridge University Press, 1998.  
[doi:10.1017/CBO9780511524738](https://doi.org/10.1017/CBO9780511524738).  
(page 41)
- [DJ92] D. Deutsch and R. Jozsa.  
Rapid solution of problems by quantum computation.  
*Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992.  
[doi:10.1098/rspa.1992.0167](https://doi.org/10.1098/rspa.1992.0167).  
(page 1)
- [DKC15] E. N. Dzhafarov, J. V. Kujala, and V. H. Cervantes.  
Contextuality-by-default: A brief overview of ideas, concepts, and terminology.  
In H. Atmanspacher, T. Filk, and E. Pothos, editors, *9th International Conference on Quantum Interaction (QI 2015)*, volume 9535 of *Lecture Notes in Computer Science*, pages 12–23. Springer, 2015.  
[doi:10.1007/978-3-319-28675-4\\_2](https://doi.org/10.1007/978-3-319-28675-4_2).  
(page 46)
- [DM21] J. F. Doriguello and A. Montanaro.  
Quantum random access codes for boolean functions.  
*Quantum*, 5:402, Mar. 2021.  
[doi:10.22331/q-2021-03-07-402](https://doi.org/10.22331/q-2021-03-07-402).  
(page 162)
- [DMK<sup>+</sup>17] T. Douce, D. Markham, E. Kashefi, E. Diamanti, T. Coudreau, P. Milman, P. van Loock, and G. Ferrini.  
Continuous-variable instantaneous quantum computing is hard to sample.  
*Physical Review Letters*, 118:070503, Feb 2017.  
[doi:10.1103/PhysRevLett.118.070503](https://doi.org/10.1103/PhysRevLett.118.070503).  
(page 86)
- [DOBV<sup>+</sup>17] N. Delfosse, C. Okay, J. Bermejo-Vega, D. E. Browne, and R. Raussendorf.

- Equivalence between contextuality and negativity of the Wigner function for qudits.  
*New Journal of Physics*, 19(12):123024, 2017.  
[doi:10.1088/1367-2630/aa8fe3](https://doi.org/10.1088/1367-2630/aa8fe3).  
(pages 4, 6, 75, 77, 80, and 82)
- [DPS03] G. M. D’Ariano, M. G. Paris, and M. F. Sacchi.  
Quantum tomography.  
*Advances in Imaging and Electron Physics*, 128:206–309, 2003, [quant-ph/0302028](https://arxiv.org/abs/quant-ph/0302028).  
(pages 87, 88, 89, and 93)
- [dS17] N. de Silva.  
Graph-theoretic strengths of contextuality.  
*Physical Review A*, 95:032108, Mar 2017.  
[doi:10.1103/PhysRevA.95.032108](https://doi.org/10.1103/PhysRevA.95.032108).  
(page 75)
- [DT06] G. B. Dantzig and M. N. Thapa.  
*Linear programming 2: theory and extensions*.  
Springer Science & Business Media, 2006.  
[doi:10.7249/r366](https://doi.org/10.7249/r366).  
(page 25)
- [DZK16] E. N. Dzhafarov, R. Zhang, and J. Kujala.  
Is there contextuality in behavioural and social systems?  
Theme issue on Quantum probability and the mathematical modelling of decision making, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2058):20150099, 2016.  
[doi:10.1098/rsta.2015.0099](https://doi.org/10.1098/rsta.2015.0099).  
(page 74)
- [EHM20] P.-E. Emeriau, M. Howard, and S. Mansfield.  
Quantum advantage in information retrieval.  
Preprint [arXiv:2007.15643 \[quant-ph\]](https://arxiv.org/abs/2007.15643), 2020.  
(pages 7 and 161)
- [EPR35] A. Einstein, B. Podolsky, and N. Rosen.  
Can quantum-mechanical description of physical reality be considered complete?  
*Physical Review*, 47(10):777–780, May 1935.  
[doi:10.1103/PhysRev.47.777](https://doi.org/10.1103/PhysRev.47.777).  
(pages 1, 2, and 172)
- [EPR92] A. C. Elitzur, S. Popescu, and D. Rohrlich.  
Quantum nonlocality for each pair in an ensemble.  
*Physics Letters A*, 162(1):25–28, 1992.  
[doi:10.1016/0375-9601\(92\)90952-I](https://doi.org/10.1016/0375-9601(92)90952-I).  
(page 43)
- [ESP02] J. Eisert, S. Scheel, and M. B. Plenio.  
Distilling Gaussian states with Gaussian operations is impossible.  
*Physical Review Letters*, 89(13):137903, 2002.  
[doi:10.1103/PhysRevLett.89.137903](https://doi.org/10.1103/PhysRevLett.89.137903).  
(page 4)
- [Fey82] R. P. Feynman.  
Simulating physics with computers.  
*International Journal of Theoretical Physics*, 21(6-7):467–488, jun 1982.  
[doi:10.1007/bf02650179](https://doi.org/10.1007/bf02650179).  
(pages 1 and 180)
- [Fin82] A. Fine.  
Hidden variables, joint probability, and the bell inequalities.

- Physical Review Letters*, 48:291–295, Feb 1982.  
[doi:10.1103/PhysRevLett.48.291](https://doi.org/10.1103/PhysRevLett.48.291).  
(pages 2, 5, 40, and 51)
- [Fiu02] J. Fiurášek.  
Gaussian transformations and distillation of entangled Gaussian states.  
*Physical Review Letters*, 89(13):137904, 2002.  
[doi:10.1103/PhysRevLett.89.137904](https://doi.org/10.1103/PhysRevLett.89.137904).  
(page 4)
- [FK19] M. Farkas and J. Kaniewski.  
Self-testing mutually unbiased bases in the prepare-and-measure scenario.  
*Physical Review A*, 99(3):032316, 2019.  
[doi:10.1103/PhysRevA.99.032316](https://doi.org/10.1103/PhysRevA.99.032316).  
(page 159)
- [FKNY02] K. Fujisawa, M. Kojima, K. Nakata, and M. Yamashita.  
*SDPA (SemiDefinite Programming Algorithm) User’s Manual—Version 6.2. 0*, 2002.  
(pages 28 and 99)
- [FMJ11] R. Filip and L. Mišta Jr.  
Detecting quantum states with a positive Wigner function beyond mixtures of Gaussian states.  
*Physical Review Letters*, 106(20):200401, 2011.  
[doi:10.1103/PhysRevLett.106.200401](https://doi.org/10.1103/PhysRevLett.106.200401).  
(page 4)
- [Fol84] G. B. Folland.  
*Real Analysis: Modern techniques and their applications*.  
Pure and Applied Mathematics. John Wiley & Sons, 1984.  
2nd Edition (1999).  
(page 61)
- [Fol92] G. B. Folland.  
*Fourier analysis and its applications*, volume 4.  
American Mathematical Soc., 1992.  
(page 85)
- [FOP05] A. Ferraro, S. Olivares, and M. G. Paris.  
Gaussian states in continuous variable quantum information.  
Preprint [arXiv:0503237 \[quant-ph\]](https://arxiv.org/abs/0503237), 2005.  
(pages 3 and 79)
- [Fri09] T. Fritz.  
Possibilistic physics.  
FQXI *Essay* Contest, 2009.  
(page 72)
- [Ga01] E. F. Galvão.  
Feasible quantum communication complexity protocol.  
*Physical Review A*, 65:012318, Dec 2001.  
[doi:10.1103/PhysRevA.65.012318](https://doi.org/10.1103/PhysRevA.65.012318).  
(page 159)
- [Ga05] E. F. Galvão.  
Discrete Wigner functions and quantum computational speedup.  
*Physical Review A*, 71:042302, Apr 2005.  
[doi:10.1103/PhysRevA.71.042302](https://doi.org/10.1103/PhysRevA.71.042302).  
(pages 6, 19, and 159)
- [GACFF20] L. García-Álvarez, C. Calcluth, A. Ferraro, and G. Ferrini.  
Efficient simulatability of continuous-variable circuits with large Wigner negativity.  
*Physical Review Research*, 2:043322, Dec 2020.

- [doi:10.1103/PhysRevResearch.2.043322](https://doi.org/10.1103/PhysRevResearch.2.043322).  
(page 4)
- [GBHA10] R. Gallego, N. Brunner, C. Hadley, and A. Acín.  
Device-independent tests of classical and quantum dimensions.  
*Physical Review Letters*, 105:230501, Nov 2010.  
[doi:10.1103/PhysRevLett.105.230501](https://doi.org/10.1103/PhysRevLett.105.230501).  
(page 169)
- [GC02] G. Giedke and J. I. Cirac.  
Characterization of Gaussian operations and distillation of Gaussian states.  
*Physical Review A*, 66(3):032316, 2002.  
[doi:10.1103/PhysRevA.66.032316](https://doi.org/10.1103/PhysRevA.66.032316).  
(page 4)
- [GEH<sup>+</sup>21] T. Giordani, C. Esposito, F. Hoch, G. Carvacho, D. J. Brod, E. F. Galvão, N. Spagnolo, and F. Sciarrino.  
Witnesses of coherence and dimension from multiphoton indistinguishability tests.  
*Physical Review Research*, 3:023031, Apr 2021.  
[doi:10.1103/PhysRevResearch.3.023031](https://doi.org/10.1103/PhysRevResearch.3.023031).  
(page 87)
- [GHH<sup>+</sup>14] A. Grudka, K. Horodecki, M. Horodecki, W. Kłobus, and M. Pawłowski.  
When are Popescu-Rohrlich boxes and random access codes equivalent?  
*Physical Review Letters*, 113(10):100401, 2014.  
[doi:10.1103/PhysRevLett.113.100401](https://doi.org/10.1103/PhysRevLett.113.100401).  
(page 159)
- [GHW04] K. S. Gibbons, M. J. Hoffman, and W. K. Wootters.  
Discrete phase space based on finite fields.  
*Physical Review A*, 70(6):062101, 2004.  
[doi:10.1103/PhysRevA.70.062101](https://doi.org/10.1103/PhysRevA.70.062101).  
(pages 15 and 164)
- [GKC<sup>+</sup>10] O. Gühne, M. Kleinmann, A. Cabello, J.-Å. Larsson, G. Kirchmair, F. Zähringer, R. Gerritsma, and C. F. Roos.  
Compatibility and noncontextuality for sequential measurements.  
*Physical Review A*, 81(2):022121, 2010.  
(page 172)
- [GKP01] D. Gottesman, A. Kitaev, and J. Preskill.  
Encoding a qubit in an oscillator.  
*Physical Review A*, 64:012310, Jun 2001.  
[doi:10.1103/PhysRevA.64.012310](https://doi.org/10.1103/PhysRevA.64.012310).  
(page 86)
- [Gou72] H. W. Gould.  
*Combinatorial Identities: A standardized set of tables listing 500 binomial coefficient summations*.  
Morgantown, W Va, 1972.  
(pages 121 and 133)
- [Gro06] D. Gross.  
Hudson’s theorem for finite-dimensional quantum systems.  
*Journal of mathematical physics*, 47(12):122107, 2006.  
[doi:10.1063/1.2393152](https://doi.org/10.1063/1.2393152).  
(pages 3, 15, 16, 19, and 178)
- [GS07] S. Ghose and B. C. Sanders.  
Non-Gaussian ancilla states for continuous variable quantum computation via Gaussian maps.  
*Journal of Modern Optics*, 54(6):855–869, 2007.  
[doi:10.1080/09500340601101575](https://doi.org/10.1080/09500340601101575).  
(page 4)

- [GT71] M. Guilletot-Teissier.  
Développements des distributions en séries de fonctions orthogonales. séries de Legendre et de Laguerre.  
*Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 25(3):519–573, 1971.  
(pages 103, 138, and 141)
- [GVW<sup>+</sup>15] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger.  
Significant-loophole-free test of bell’s theorem with entangled photons.  
*Physical Review Letters*, 115:250401, Dec 2015.  
[doi:10.1103/PhysRevLett.115.250401](https://doi.org/10.1103/PhysRevLett.115.250401).  
(page 2)
- [Hal13] B. C. Hall.  
*Quantum theory for mathematicians*, volume 267.  
Springer, 2013.  
[doi:10.1007/978-1-4614-7116-5](https://doi.org/10.1007/978-1-4614-7116-5).  
(pages 77, 78, 80, and 85)
- [Har92] L. Hardy.  
Quantum mechanics, local realistic theories, and lorentz-invariant realistic theories.  
*Physical Review Letters*, 68:2981–2984, May 1992.  
[doi:10.1103/PhysRevLett.68.2981](https://doi.org/10.1103/PhysRevLett.68.2981).  
(page 160)
- [Har93] L. Hardy.  
Nonlocality for two particles without inequalities for almost all entangled states.  
*Physical Review Letters*, 71:1665–1668, Sep 1993.  
[doi:10.1103/PhysRevLett.71.1665](https://doi.org/10.1103/PhysRevLett.71.1665).  
(pages 40 and 160)
- [Hav36] E. Haviland.  
On the momentum problem for distribution functions in more than one dimension. ii.  
*American Journal of Mathematics*, 58(1):164–168, 1936.  
[doi:10.2307/2371063](https://doi.org/10.2307/2371063).  
(pages 31, 106, and 152)
- [HBD<sup>+</sup>15] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán, et al.  
Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres.  
*Nature*, 526(7575):682–686, 2015.  
[doi:10.1038/nature15759](https://doi.org/10.1038/nature15759).  
(page 2)
- [HCB<sup>+</sup>18] L. Henaut, L. Catani, D. E. Browne, S. Mansfield, and A. Pappa.  
Tsirelson’s bound and Landauer’s principle in a single-system game.  
*Physical Review A*, 98(6):060302, 2018.  
[doi:10.1103/PhysRevA.98.060302](https://doi.org/10.1103/PhysRevA.98.060302).  
(pages 73 and 175)
- [HCRD10] Q.-Y. He, E. G. Cavalcanti, M. D. Reid, and P. D. Drummond.  
Bell inequalities for continuous-variable measurements.  
*Physical Review A*, 81(6):062106, 2010.  
[doi:10.1103/PhysRevA.81.062106](https://doi.org/10.1103/PhysRevA.81.062106).  
(page 47)
- [Hei85] W. Heisenberg.  
Über den anschaulichen inhalt der quantentheoretischen kinematik und mechanik.  
In *Original Scientific Papers Wissenschaftliche Originalarbeiten*, pages 478–504. Springer, 1985.  
[doi:10.1007/BF01397280](https://doi.org/10.1007/BF01397280).  
(page 17)

- [Hil88] D. Hilbert.  
Über die darstellung definiter formen als summe von formenquadraten.  
*Mathematische Annalen*, 32(3):342–350, 1888.  
[doi:10.1007/BF01443605](https://doi.org/10.1007/BF01443605).  
(pages 29, 114, and 146)
- [HLB<sup>+</sup>06] Y. Hasegawa, R. Loidl, G. Badurek, M. Baron, and H. Rauch.  
Quantum contextuality in a single-neutron optical experiment.  
*Physical Review Letters*, 97:230401, Dec 2006.  
[doi:10.1103/PhysRevLett.97.230401](https://doi.org/10.1103/PhysRevLett.97.230401).  
(page 3)
- [Hol73] A. S. Holevo.  
Bounds for the quantity of information transmitted by a quantum communication channel.  
*Problemy Peredachi Informatsii*, 9(3):3–11, 1973.  
(page 173)
- [How15] M. Howard.  
Classical codes in quantum state space.  
*Journal of Physics A: Mathematical and Theoretical*, 48(49):495303, 2015.  
[doi:10.1088/1751-8113/48/49/495303](https://doi.org/10.1088/1751-8113/48/49/495303).  
(page 166)
- [HR21] F. Hernandez and C. J. Riedel.  
Rapidly decaying functions are Schwartz functions.  
Preprint [arXiv:2103.14183 \[math-ph,quant-ph\]](https://arxiv.org/abs/2103.14183), Mar. 2021.  
(page 113)
- [HSS08] T. Hofmann, B. Schölkopf, and A. J. Smola.  
Kernel methods in machine learning.  
*The Annals of Statistics*, 36(3):1171–1220, 2008.  
[doi:10.1214/009053607000000677](https://doi.org/10.1214/009053607000000677).  
(page 74)
- [HTMB17] A. Hameedi, A. Tavakoli, B. Marques, and M. Bourennane.  
Communication games reveal preparation contextuality.  
*Physical Review Letters*, 119(22):220402, 2017.  
[doi:10.1103/PhysRevLett.119.220402](https://doi.org/10.1103/PhysRevLett.119.220402).  
(pages 160 and 178)
- [Hud74] R. L. Hudson.  
When is the Wigner quasi-probability density non-negative?  
*Reports on Mathematical Physics*, 6(2):249–252, 1974.  
[doi:10.1016/0034-4877\(74\)90007-X](https://doi.org/10.1016/0034-4877(74)90007-X).  
(pages 4, 15, and 19)
- [HUL13] J.-B. Hiriart-Urruty and C. Lemaréchal.  
*Convex analysis and minimization algorithms I: Fundamentals*, volume 305.  
Springer science & business media, 2013.  
[doi:10.1007/978-3-662-02796-7](https://doi.org/10.1007/978-3-662-02796-7).  
(page 24)
- [Hus40] K. Husimi.  
Some formal properties of the density matrix.  
*Proceedings of the Physico-Mathematical Society of Japan. 3rd Series*, 22(4):264–314, 1940.  
[doi:10.11429/ppmsj1919.22.4\\_264](https://doi.org/10.11429/ppmsj1919.22.4_264).  
(page 87)
- [HWVE14] M. Howard, J. Wallman, V. Veitch, and J. Emerson.  
Contextuality supplies the ‘magic’ for quantum computation.  
*Nature*, 510(7505):351–355, 2014.

- doi:10.1038/nature13460.  
(pages 3, 4, 6, 75, 77, 86, 172, and 180)
- [HXG+21] C.-J. Huang, G.-Y. Xiang, Y. Guo, K.-D. Wu, B.-H. Liu, C.-F. Li, G.-C. Guo, and A. Tavakoli.  
Nonlocality, steering, and quantum state tomography in a single experiment.  
*Physical Review Letters*, 127:020401, Jul 2021.  
doi:10.1103/PhysRevLett.127.020401.  
(pages 7, 178, and 181)
- [Jar84] J. P. Jarrett.  
On the physical significance of the locality conditions in the Bell arguments.  
Special issue on the Foundations of Quantum Mechanics, *Noûs*, 18(4):569–589, 1984.  
doi:10.2307/2214878.  
(page 41)
- [Kak41] S. Kakutani.  
Concrete representation of abstract ( $M$ )-spaces (a characterization of the space of continuous functions).  
*Annals of Mathematics*, 42(4):994–1024, 1941.  
doi:10.2307/1968778.  
(pages 26, 56, and 61)
- [Kis16] K. Kishida.  
Logic of local inference for contextuality in quantum physics and beyond.  
In I. Chatzigiannakis, M. Mitzenmacher, Y. Rabani, and D. Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 113:1–113:14. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016.  
doi:10.4230/LIPIcs.ICALP.2016.113.  
(page 74)
- [KLFA18] A. Ketterer, A. Laversanne-Finot, and L. Aolita.  
Continuous-variable supraquantum nonlocality.  
*Physical Review A*, 97(1):012133, 2018.  
doi:10.1103/PhysRevA.97.012133.  
(page 47)
- [KMP21] I. Klep, V. Magron, and J. Povh.  
Sparse noncommutative polynomial optimization.  
*Mathematical Programming*, pages 1–41, 2021.  
doi:10.1007/s10107-020-01610-1.  
(pages 158 and 181)
- [KMV20] I. Klep, V. Magron, and J. Volčič.  
Optimization over trace polynomials.  
Preprint [arXiv:2006.12510 \[math-ph\]](https://arxiv.org/abs/2006.12510), 2020.  
(pages 158 and 181)
- [KP17] I. Kerenidis and A. Prakash.  
Quantum Recommendation Systems.  
In C. H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 49:1–49:21, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.  
doi:10.4230/LIPIcs.ITCS.2017.49.  
(page 1)
- [KS75] S. Kochen and E. P. Specker.  
The problem of hidden variables in quantum mechanics.  
In *The logico-algebraic approach to quantum mechanics*, pages 293–328. Springer, 1975.  
doi:10.1007/978-94-010-1795-4\_17.  
(pages 2, 3, 5, 35, 40, and 172)

- [KV12] T. Kiesel and W. Vogel.  
Universal nonclassicality witnesses for harmonic oscillators.  
*Physical Review A*, 85(6):062106, 2012.  
[doi:10.1103/PhysRevA.85.062106](https://doi.org/10.1103/PhysRevA.85.062106).  
(page 87)
- [KŻ04] A. Kenfack and K. Życzkowski.  
Negativity of the Wigner function as an indicator of non-classicality.  
*Journal of Optics B: Quantum and Semiclassical Optics*, 6(10):396, 2004.  
[doi:10.1088/1464-4266/6/10/003](https://doi.org/10.1088/1464-4266/6/10/003).  
(pages 4, 20, and 109)
- [KZG<sup>+</sup>09] G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt, and C. F. Roos.  
State-independent experimental test of quantum contextuality.  
*Nature*, 460(7254):494–497, 2009.  
[doi:10.1038/nature08172](https://doi.org/10.1038/nature08172).  
(pages 3 and 172)
- [Las01] J.-B. Lasserre.  
Global optimization with polynomials and the problem of moments.  
*SIAM Journal on optimization*, 11(3):796–817, 2001.  
[doi:10.1137/S1052623400366802](https://doi.org/10.1137/S1052623400366802).  
(page 28)
- [Las09] J.-B. Lasserre.  
Moments, positive polynomials and their applications.  
In *Series on Optimization and its Applications*, volume 1. Imperial College Press, 2009.  
[doi:10.1142/p665](https://doi.org/10.1142/p665).  
(pages 6, 28, 29, 30, 33, 65, 66, 140, and 180)
- [Las11] J. B. Lasserre.  
A new look at nonnegativity on closed sets and polynomial optimization.  
*SIAM Journal on Optimization*, 21(3):864–885, 2011.  
[doi:10.1137/100806990](https://doi.org/10.1137/100806990).  
(pages 65 and 158)
- [Lau09] M. Laurent.  
*Sums of Squares, Moment Matrices and Optimization Over Polynomials*, pages 157–270.  
Springer New York, New York, NY, 2009.  
[doi:10.1007/978-0-387-09686-5\\_7](https://doi.org/10.1007/978-0-387-09686-5_7).  
(pages 30 and 65)
- [LB99] S. Lloyd and S. L. Braunstein.  
Quantum computation over continuous variables.  
In *Quantum Information with Continuous Variables*, pages 9–17. Springer, 1999.  
[doi:10.1007/978-94-015-1258-9\\_2](https://doi.org/10.1007/978-94-015-1258-9_2).  
(pages 3 and 4)
- [Lee91] C. T. Lee.  
Measure of the nonclassicality of nonclassical states.  
*Physical Review A*, 44(5):R2775, 1991.  
[doi:10.1103/PhysRevA.44.R2775](https://doi.org/10.1103/PhysRevA.44.R2775).  
(page 3)
- [Leo10] U. Leonhardt.  
*Essential Quantum Optics*.  
Cambridge University Press, Cambridge, UK, 1st edition, 2010.  
[doi:10.1017/CBO9780511806117](https://doi.org/10.1017/CBO9780511806117).  
(page 3)
- [LFKB<sup>+</sup>17] A. Laversanne-Finot, A. Ketterer, M. R. Barros, S. P. Walborn, T. Coudreau, A. Keller, and P. Milman.

- General conditions for maximal violation of non-contextuality in discrete and continuous variables.  
*Journal of Physics A: Mathematical and Theoretical*, 50(15):155304, 2017.  
[doi:10.1088/1751-8121/aa6016](https://doi.org/10.1088/1751-8121/aa6016).  
(page 47)
- [LKCH00] M. Lewenstein, B. Kraus, J. I. Cirac, and P. Horodecki.  
Optimization of entanglement witnesses.  
*Physical Review A*, 62(5):052310, 2000.  
[doi:10.1103/PhysRevA.62.052310](https://doi.org/10.1103/PhysRevA.62.052310).  
(page 87)
- [LLS<sup>+</sup>11] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wieśniak, and A. Zeilinger.  
Experimental non-classicality of an indivisible quantum system.  
*Nature*, 474(7352):490–493, 2011.  
[doi:10.1038/nature10119](https://doi.org/10.1038/nature10119).  
(page 172)
- [LR09] A. I. Lvovsky and M. G. Raymer.  
Continuous-variable optical quantum-state tomography.  
*Reviews of Modern Physics*, 81(1):299, 2009.  
[doi:10.1103/RevModPhys.81.299](https://doi.org/10.1103/RevModPhys.81.299).  
(pages 87 and 93)
- [Lue97] D. G. Luenberger.  
*Optimization by vector space methods*.  
John Wiley & Sons, 1997.  
(page 60)
- [Man13] S. Mansfield.  
*The Mathematical Structure of Non-locality and Contextuality*.  
PhD thesis, University of Oxford, 2013.  
(page 73)
- [Man16] S. Mansfield.  
Reality of the quantum state: Towards a stronger  $\psi$ -ontology theorem.  
*Physical Review A*, 94:042124, Oct 2016.  
[doi:10.1103/PhysRevA.94.042124](https://doi.org/10.1103/PhysRevA.94.042124).  
(page 1)
- [Man17] S. Mansfield.  
Consequences and applications of the completeness of Hardy’s nonlocality.  
*Physical Review A*, 95(2):022122, 2017.  
[doi:10.1103/PhysRevA.95.022122](https://doi.org/10.1103/PhysRevA.95.022122).  
(pages 47 and 160)
- [Man20] S. Mansfield.  
Contextuality is topological.  
*Quantum Views*, 4:31, Feb. 2020.  
[doi:10.22331/qv-2020-02-03-31](https://doi.org/10.22331/qv-2020-02-03-31).  
(page 37)
- [MB14] S. Mansfield and R. S. Barbosa.  
Extendability in the sheaf-theoretic approach: Construction of Bell models from Kochen–Specker models.  
Preprint [arXiv:1402.4827 \[quant-ph\]](https://arxiv.org/abs/1402.4827), 2014.  
(page 39)
- [ME12] A. Mari and J. Eisert.  
Positive Wigner functions render classical simulation of quantum computation efficient.  
*Physical Review Letters*, 109:230503, Dec 2012.  
[doi:10.1103/PhysRevLett.109.230503](https://doi.org/10.1103/PhysRevLett.109.230503).  
(pages 4, 19, and 159)

- [MF12] S. Mansfield and T. Fritz.  
Hardy’s non-locality paradox and possibilistic conditions for non-locality.  
*Foundations of Physics*, 42(5):709–719, 2012.  
[doi:10.1007/s10701-012-9640-1](https://doi.org/10.1007/s10701-012-9640-1).  
(pages 47 and 72)
- [MK18] S. Mansfield and E. Kashefi.  
Quantum advantage from sequential-transformation contextuality.  
*Physical Review Letters*, 121(23):230401, 2018.  
[doi:10.1103/PhysRevLett.121.230401](https://doi.org/10.1103/PhysRevLett.121.230401).  
(pages 7, 54, 73, 160, 162, 172, 174, 175, and 177)
- [MKC09] A. Mandilara, E. Karpov, and N. Cerf.  
Extending Hudson’s theorem to mixed quantum states.  
*Physical Review A*, 79(6):062302, 2009.  
[doi:10.1103/PhysRevA.79.062302](https://doi.org/10.1103/PhysRevA.79.062302).  
(page 4)
- [MKN<sup>+</sup>11] A. Mari, K. Kieling, B. M. Nielsen, E. Polzik, and J. Eisert.  
Directly estimating nonclassicality.  
*Physical Review Letters*, 106(1):010403, 2011.  
[doi:10.1103/physrevlett.106.010403](https://doi.org/10.1103/physrevlett.106.010403).  
(pages 20 and 87)
- [MM92] S. MacLane and I. Moerdijk.  
*Sheaves in geometry and logic: A first introduction to topos theory*.  
Universitext. Springer, 1992.  
[doi:10.1007/978-1-4612-0927-0](https://doi.org/10.1007/978-1-4612-0927-0).  
(page 36)
- [Moy49] J. E. Moyal.  
Quantum mechanics as a statistical theory.  
In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 45, pages 99–124. Cambridge University Press, 1949.  
[doi:10.1017/S0305004100000487](https://doi.org/10.1017/S0305004100000487).  
(page 3)
- [MPP11] G. McKeown, M. G. Paris, and M. Paternostro.  
Testing quantum contextuality of continuous-variable states.  
*Physical Review A*, 83(6):062105, 2011.  
[doi:10.1103/PhysRevA.83.062105](https://doi.org/10.1103/PhysRevA.83.062105).  
(page 47)
- [Nak10] M. Nakata.  
A numerical evaluation of highly accurate multiple-precision arithmetic version of semidefinite programming solver: Sdpa-gmp,-qd and-dd.  
In *2010 IEEE International Symposium on Computer-Aided Control System Design*, pages 29–34. IEEE, 2010.  
[doi:10.1109/cacsd.2010.5612693](https://doi.org/10.1109/cacsd.2010.5612693).  
(pages 28 and 99)
- [NC11] M. A. Nielsen and I. L. Chuang.  
*Quantum Computation and Quantum Information: 10th Anniversary Edition*.  
Cambridge University Press, New York, NY, USA, 10th edition, 2011.  
[doi:10.1017/CBO9780511976667](https://doi.org/10.1017/CBO9780511976667).  
(pages 11, 13, 20, and 91)
- [NFAV15] M. Navascués, A. Feix, M. Araújo, and T. Vértesi.  
Characterizing finite-dimensional quantum behavior.  
*Physical Review A*, 92(4):042117, 2015.  
[doi:10.1103/PhysRevA.92.042117](https://doi.org/10.1103/PhysRevA.92.042117).

- (pages 34, 167, 171, and 181)
- [NFC09] J. Niset, J. Fiurášek, and N. J. Cerf.  
No-go theorem for Gaussian quantum error correction.  
*Physical Review Letters*, 102(12):120501, 2009.  
[doi:10.1103/PhysRevLett.102.120501](https://doi.org/10.1103/PhysRevLett.102.120501).  
(page 4)
- [Nie13] J. Nie.  
Certifying convergence of Lasserre’s hierarchy via flat truncation.  
*Mathematical Programming*, 142(1):485–510, 2013.  
[doi:10.1007/s10107-012-0589-9](https://doi.org/10.1007/s10107-012-0589-9).  
(page 139)
- [Nik30] O. Nikodym.  
Sur une généralisation des intégrales de M. J. Radon.  
*Fundamenta Mathematicae*, 15(1):131–179, 1930.  
[doi:10.4064/fm-15-1-131-179](https://doi.org/10.4064/fm-15-1-131-179).  
(pages 84 and 103)
- [NPA07] M. Navascués, S. Pironio, and A. Acín.  
Bounding the set of quantum correlations.  
*Physical Review Letters*, 98:010401, Jan 2007.  
[doi:10.1103/PhysRevLett.98.010401](https://doi.org/10.1103/PhysRevLett.98.010401).  
(page 181)
- [NPA08] M. Navascués, S. Pironio, and A. Acín.  
A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations.  
*New Journal of Physics*, 10(7):073013, 2008.  
[doi:10.1088/1367-2630/10/7/073013](https://doi.org/10.1088/1367-2630/10/7/073013).  
(pages 33, 158, and 167)
- [OE12] M. Ohliger and J. Eisert.  
Efficient measurement-based quantum computing with continuous-variable systems.  
*Physical Review A*, 85:062318, Jun 2012.  
[doi:10.1103/PhysRevA.85.062318](https://doi.org/10.1103/PhysRevA.85.062318).  
(page 3)
- [OGa17] A. L. Oestereich and E. F. Galvão.  
Reliable computation from contextual correlations.  
*Physical Review A*, 96:062305, Dec 2017.  
[doi:10.1103/PhysRevA.96.062305](https://doi.org/10.1103/PhysRevA.96.062305).  
(page 160)
- [Orb11] P. Orbanz.  
Probability theory ii, 2011.  
(page 62)
- [ORBR17] C. Okay, S. Roberts, S. D. Bartlett, and R. Raussendorf.  
Topological proofs of contextuality in quantum mechanics.  
*Quantum Information and Computation*, 17(13–14):1135–1166, 2017.  
[doi:10.26421/QIC17.13-14](https://doi.org/10.26421/QIC17.13-14).  
(page 73)
- [OTBLG06] A. Ourjoumtsev, R. Tualle-Broui, J. Laurat, and P. Grangier.  
Generating optical Schrödinger kittens for quantum information processing.  
*Science*, 312(5770):83–86, 2006.  
[doi:10.1126/science.1122858](https://doi.org/10.1126/science.1122858).  
(page 94)
- [OTR18] C. Okay, E. Tyhurst, and R. Raussendorf.

- The cohomological and the resource-theoretic perspective on quantum contextuality: common ground through the contextual fraction.  
Preprint [arXiv:1806.04657 \[quant-ph\]](https://arxiv.org/abs/1806.04657), 2018.  
(page 73)
- [Par67] K. Parthasarathy.  
*Probability Measures on Metric Spaces*.  
Elsevier, 1967.  
[doi:10.1016/C2013-0-08107-8](https://doi.org/10.1016/C2013-0-08107-8).  
(page 61)
- [Par03] P. A. Parrilo.  
Semidefinite programming relaxations for semialgebraic problems.  
*Mathematical programming*, 96(2):293–320, 2003.  
[doi:10.1007/s10107-003-0387-5](https://doi.org/10.1007/s10107-003-0387-5).  
(pages 28 and 65)
- [PBR12] M. F. Pusey, J. Barrett, and T. Rudolph.  
On the reality of the quantum state.  
*Nature Physics*, 8(6):475–478, 2012.  
(page 1)
- [PC10] Á. R. Plastino and A. Cabello.  
State-independent quantum contextuality for continuous variables.  
*Physical Review A*, 82(2):022114, 2010.  
[doi:10.1103/PhysRevA.82.022114](https://doi.org/10.1103/PhysRevA.82.022114).  
(page 47)
- [PPK<sup>+</sup>09] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski.  
Information causality as a physical principle.  
*Nature*, 461(7267):1101–1104, 2009.  
[doi:10.1038/nature08400](https://doi.org/10.1038/nature08400).  
(page 159)
- [PR94] S. Popescu and D. Rohrlich.  
Quantum nonlocality as an axiom.  
*Foundations of Physics*, 24(3):379–385, 1994.  
[doi:10.1007/bf02058098](https://doi.org/10.1007/bf02058098).  
(pages 40, 47, 49, 160, and 168)
- [PTWP21] J. Pauwels, A. Tavakoli, E. Woodhead, and S. Pironio.  
Entanglement in prepare-and-measure scenarios: many questions, a few answers.  
Preprint [arXiv:2108.00442 \[quant-ph\]](https://arxiv.org/abs/2108.00442), 2021.  
(page 178)
- [Put93] M. Putinar.  
Positive polynomials on compact semi-algebraic sets.  
*Indiana University Mathematics Journal*, 42(3):969–984, 1993.  
URL <http://www.jstor.org/stable/24897130>.  
(page 30)
- [PWB15] H. Pashayan, J. J. Wallman, and S. D. Bartlett.  
Estimating outcome probabilities of quantum circuits using quasiprobabilities.  
*Physical Review Letters*, 115(7):070501, 2015.  
[doi:10.1103/PhysRevLett.115.070501](https://doi.org/10.1103/PhysRevLett.115.070501).  
(page 6)
- [Rau13] R. Raussendorf.  
Contextuality in measurement-based quantum computation.  
*Physical Review A*, 88(2):022322, 2013.  
[doi:10.1103/PhysRevA.88.022322](https://doi.org/10.1103/PhysRevA.88.022322).  
(page 3)

- [Rau16] R. Raussendorf.  
Cohomological framework for contextual quantum computations.  
Preprint [arXiv:1602.04155 \[quant-ph\]](https://arxiv.org/abs/1602.04155), 2016.  
(page 73)
- [RB01] R. Raussendorf and H. J. Briegel.  
A one-way quantum computer.  
*Physical Review Letters*, 86(22):5188, 2001.  
[doi:10.1103/PhysRevLett.86.5188](https://doi.org/10.1103/PhysRevLett.86.5188).  
(pages 3 and 55)
- [RBD<sup>+</sup>17] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega.  
Contextuality and Wigner-function negativity in qubit quantum computation.  
*Physical Review A*, 95(5):052334, 2017.  
[doi:10.1103/PhysRevA.95.052334](https://doi.org/10.1103/PhysRevA.95.052334).  
(pages 4 and 75)
- [RE21] R. P. Rundle and M. J. Everitt.  
Overview of the phase space formulation of quantum mechanics with application to quantum technologies.  
*Advanced Quantum Technologies*, page 2100016, 2021.  
[doi:10.1002/qute.202100016](https://doi.org/10.1002/qute.202100016).  
(page 14)
- [Ric98] T. Richter.  
Determination of photon statistics and density matrix from double homodyne detection measurements.  
*Journal of Modern Optics*, 45(8):1735–1749, 1998.  
[doi:10.1080/09500349808230666](https://doi.org/10.1080/09500349808230666).  
(page 87)
- [Rie09] F. Riesz.  
*Sur les opérations fonctionnelles linéaires*.  
Gauthier-Villars, 1909.  
(page 83)
- [Rie23] M. Riesz.  
Sur le problème des moments, troisième note.  
*Ark. Mat. Fys*, 16:1–52, 1923.  
(page 106)
- [RKM<sup>+</sup>01] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland.  
Experimental violation of a bell’s inequality with efficient detection.  
*Nature*, 409(6822):791–794, 2001.  
[doi:10.1038/35057215](https://doi.org/10.1038/35057215).  
(page 2)
- [Rou17] F. Roumen.  
Cohomology of effect algebras.  
In R. Duncan and C. Heunen, editors, *13th International Conference on Quantum Physics and Logic (QPL 2016)*, volume 236 of *Electronic Proceedings in Theoretical Computer Science*, pages 174–201. Open Publishing Association, 2017.  
[doi:10.4204/EPTCS.236.12](https://doi.org/10.4204/EPTCS.236.12).  
(page 73)
- [RS75] M. Reed and B. Simon.  
*II: Fourier Analysis, Self-Adjointness*, volume 2.  
Elsevier, 1975.  
(pages 104 and 115)
- [RT21] J. Ross and M. Toma.  
On Hodge-Riemann cohomology classes.  
Preprint [arXiv:2106.112853 \[math.AG\]](https://arxiv.org/abs/2106.112853), 2021.  
(page 180)

- [San92] B. C. Sanders.  
Entangled coherent states.  
*Physical Review A*, 45(9):6811, 1992.  
[doi:10.1103/PhysRevA.45.6811](https://doi.org/10.1103/PhysRevA.45.6811).  
(page 95)
- [SBK<sup>+</sup>09] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde.  
Preparation contextuality powers parity-oblivious multiplexing.  
*Physical Review Letters*, 102(1):010401, 2009.  
[doi:10.1103/PhysRevLett.102.010401](https://doi.org/10.1103/PhysRevLett.102.010401).  
(pages 159 and 160)
- [SC83] F. Soto and P. Claverie.  
When is the Wigner function of multidimensional systems nonnegative?  
*Journal of Mathematical Physics*, 24(1):97–100, 1983.  
[doi:10.1063/1.525607](https://doi.org/10.1063/1.525607).  
(page 4)
- [Sch47] L. Schwartz.  
Théorie des distributions et transformation de Fourier.  
In *Annales de l'Université de Grenoble*, volume 23, pages 7–24, 1947.  
(page 8)
- [Sch60] J. Schwinger.  
Unitary operator bases.  
*Proceedings of the national academy of sciences of the United States Of America*, 46(4):570, 1960.  
[doi:10.1142/9789812795694\\_0026](https://doi.org/10.1142/9789812795694_0026).  
(page 15)
- [Sch05] M. Schlosshauer.  
Decoherence, the measurement problem, and interpretations of quantum mechanics.  
*Reviews of modern physics*, 76:1267–1305, Feb 2005.  
[doi:10.1103/RevModPhys.76.1267](https://doi.org/10.1103/RevModPhys.76.1267).  
(page 1)
- [SCW<sup>+</sup>12] H.-Y. Su, J.-L. Chen, C. Wu, S. Yu, and C. Oh.  
Quantum contextuality in a one-dimensional quantum harmonic oscillator.  
*Physical Review A*, 85(5):052126, 2012.  
[doi:10.1103/PhysRevA.85.052126](https://doi.org/10.1103/PhysRevA.85.052126).  
(page 47)
- [SHD<sup>+</sup>13] X. Su, S. Hao, X. Deng, L. Ma, M. Wang, X. Jia, C. Xie, and K. Peng.  
Gate sequence for continuous variable one-way quantum computation.  
*Nature communications*, 4(1):1–9, 2013.  
[doi:10.1038/ncomms3828](https://doi.org/10.1038/ncomms3828).  
(page 79)
- [Shi86] A. Shimony.  
Events and processes in the quantum world.  
In R. Penrose and C. J. Isham, editors, *Quantum concepts in space and time*, pages 182–203. Oxford University Press, 1986.  
[doi:10.1017/CBO9781139172196.011](https://doi.org/10.1017/CBO9781139172196.011).  
(page 41)
- [Sho94] P. W. Shor.  
Algorithms for quantum computation: discrete logarithms and factoring.  
In *Proceedings 35th annual symposium on foundations of computer science*, pages 124–134. Ieee, 1994.  
[doi:10.1109/sfcs.1994.365700](https://doi.org/10.1109/sfcs.1994.365700).  
(page 1)
- [SHP19] D. Saha, P. Horodecki, and M. Pawłowski.

- State independent contextuality advances one-way communication.  
*New Journal of Physics*, 21(9):093057, 2019.  
[doi:10.1088/1367-2630/ab4149](https://doi.org/10.1088/1367-2630/ab4149).  
(pages 6, 160, 169, 174, and 178)
- [Sim17] A. W. Simmons.  
How (maximally) contextual is quantum mechanics?  
Preprint [arXiv:1712.03766 \[quant-ph\]](https://arxiv.org/abs/1712.03766), 2017.  
(page 39)
- [SKK<sup>+</sup>09] W. Son, J. Kofler, M. S. Kim, V. Vedral, and i. c. v. Brukner.  
Positive phase space transformation incompatible with classical physics.  
*Physical Review Letters*, 102:110404, Mar 2009.  
[doi:10.1103/PhysRevLett.102.110404](https://doi.org/10.1103/PhysRevLett.102.110404).  
(pages 76 and 85)
- [Slo16] W. Slofstra.  
Tsirelson’s problem and an embedding theorem for groups arising from non-local games.  
Preprint [arXiv:1606.03140 \[quant-ph\]](https://arxiv.org/abs/1606.03140), 2016.  
(page 73)
- [SMSC<sup>+</sup>15] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam.  
Strong loophole-free test of local realism.  
*Physical Review Letters*, 115:250402, Dec 2015.  
[doi:10.1103/PhysRevLett.115.250402](https://doi.org/10.1103/PhysRevLett.115.250402).  
(page 2)
- [Spe05] R. W. Spekkens.  
Contextuality for preparations, transformations, and unsharp measurements.  
*Physical Review A*, 71:052108, May 2005.  
[doi:10.1103/PhysRevA.71.052108](https://doi.org/10.1103/PhysRevA.71.052108).  
(pages 3, 40, 46, 73, 160, and 172)
- [SS12] G. Sagnol and M. Stahlberg.  
Picos, a python interface to conic optimization solvers.  
In *Proceedings of the in 21st International Symposium on Mathematical Programming*, 2012.  
(pages 28 and 99)
- [SSM88] R. Simon, E. C. G. Sudarshan, and N. Mukunda.  
Gaussian pure states in quantum mechanics and the symplectic group.  
*Physical Review A*, 37:3028–3038, Apr 1988.  
[doi:10.1103/PhysRevA.37.3028](https://doi.org/10.1103/PhysRevA.37.3028).  
(page 76)
- [Sti55] W. F. Stinespring.  
Positive functions on  $C^*$ -algebras.  
*Proceedings of the American Mathematical Society*, 6(2):211–216, 1955.  
URL <http://www.jstor.org/stable/2032342>.  
(page 13)
- [Sze59] G. Szegő.  
Orthogonal polynomials, revised ed.  
In *American Mathematical Society Colloquium Publications*, volume 23, 1959.  
[doi:10.1090/coll/023](https://doi.org/10.1090/coll/023).  
(pages 20 and 103)
- [Tan19] E. Tang.

- A quantum-inspired classical algorithm for recommendation systems.  
*Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 217–228, June 2019.  
[doi:10.1145/3313276.3316310](https://doi.org/10.1145/3313276.3316310).  
(page 1)
- [Tao11] T. Tao.  
*An introduction to measure theory*, volume 126.  
American Mathematical Society Providence, RI, 2011.  
[doi:10.1090/gsm/126](https://doi.org/10.1090/gsm/126).  
(page 20)
- [TC19] M. Terra Cunha.  
On measures and measurements: a fibre bundle approach to contextuality.  
*Philosophical Transactions of the Royal Society A*, 377(2157):20190146, 2019.  
[doi:10.1098/rsta.2019.0146](https://doi.org/10.1098/rsta.2019.0146).  
(pages 47 and 73)
- [Ter01] B. M. Terhal.  
A family of indecomposable positive linear maps based on entangled quantum states.  
*Linear Algebra and its Applications*, 323(1-3):61–73, 2001.  
[doi:10.1016/S0024-3795\(00\)00251-2](https://doi.org/10.1016/S0024-3795(00)00251-2).  
(page 87)
- [THMB15] A. Tavakoli, A. Hameedi, B. Marques, and M. Bourennane.  
Quantum random access codes using single d-level systems.  
*Physical Review Letters*, 114(17):170502, 2015.  
[doi:10.1103/PhysRevLett.114.170502](https://doi.org/10.1103/PhysRevLett.114.170502).  
(pages 159, 162, and 171)
- [TRR19] A. Tavakoli, D. Rosset, and M.-O. Renou.  
Enabling computation of correlation bounds for finite-dimensional quantum systems via symmetrization.  
*Physical Review Letters*, 122(7):070501, 2019.  
[doi:https://doi.org/10.1103/physrevlett.122.070501](https://doi.org/10.1103/physrevlett.122.070501).  
(page 171)
- [Tyc30] A. Tychonoff.  
Über die topologische erweiterung von räumen.  
*Mathematische Annalen*, 102(1):544–561, 1930.  
[doi:10.1007/BF01782364](https://doi.org/10.1007/BF01782364).  
(page 21)
- [TZ18] R. Takagi and Q. Zhuang.  
Convex resource theory of non-Gaussianity.  
*Physical Review A*, 97(6):062337, 2018.  
[doi:10.1103/PhysRevA.97.062337](https://doi.org/10.1103/PhysRevA.97.062337).  
(page 3)
- [VB96] L. Vandenberghe and S. Boyd.  
Semidefinite programming.  
*SIAM review*, 38(1):49–95, 1996.  
[doi:10.1137/1038003](https://doi.org/10.1137/1038003).  
(page 27)
- [vDH11] W. van Dam and M. Howard.  
Noise thresholds for higher-dimensional systems using the discrete Wigner function.  
*Physical Review A*, 83(3):032310, 2011.  
[doi:10.1103/PhysRevA.83.032310](https://doi.org/10.1103/PhysRevA.83.032310).  
(pages 165 and 178)
- [VFGE12] V. Veitch, C. Ferrie, D. Gross, and J. Emerson.

- Negative quasi-probability as a resource for quantum computation.  
*New Journal of Physics*, 14(11):113011, 2012.  
[doi:10.1088/1367-2630/14/11/113011](https://doi.org/10.1088/1367-2630/14/11/113011).  
(pages 6 and 19)
- [VHC21] Z. Van Herstraeten and N. J. Cerf.  
Quantum Wigner entropy.  
Preprint [arXiv:2105.12843 \[quant-ph\]](https://arxiv.org/abs/2105.12843), May 2021.  
(page 158)
- [Vil09] C. Villani.  
*Optimal transport: old and new*, volume 338.  
Springer, 2009.  
(page 181)
- [Vor62] N. N. Vorob'ev.  
Consistent families of measures and their extensions.  
*Theory of Probability & Its Applications*, 7(2):147–163, 1962.  
[doi:10.1137/1107014](https://doi.org/10.1137/1107014).  
translated from Russian by N.Greenleaf.  
(pages 40 and 181)
- [Wal21] M. Walschaers.  
Non-Gaussian quantum states and where to find them.  
Preprint [arXiv:2104.12596 \[quant-ph\]](https://arxiv.org/abs/2104.12596), 2021.  
(pages 3, 16, and 76)
- [Wes18] L. Wester.  
*Classical and quantum structures of computation*.  
PhD thesis, University of Oxford, 2018.  
(page 177)
- [Wig32] E. Wigner.  
On the quantum correction for thermodynamic equilibrium.  
*Physical Review*, 40:749–759, Jun 1932.  
[doi:10.1103/PhysRev.40.749](https://doi.org/10.1103/PhysRev.40.749).  
(pages 3 and 14)
- [Wig95] E. P. Wigner.  
Remarks on the mind-body question.  
In *Philosophical reflections and syntheses*, pages 247–260. Springer, 1995.  
[doi:10.1007/978-3-642-78374-6\\_20](https://doi.org/10.1007/978-3-642-78374-6_20).  
(page 1)
- [WLT18] T. Weisser, J. B. Lasserre, and K.-C. Toh.  
Sparse-bsos: a bounded degree sos hierarchy for large scale polynomial optimization with sparsity.  
*Mathematical Programming Computation*, 10(1):1–32, 2018.  
[doi:10.1007/s12532-017-0121-6](https://doi.org/10.1007/s12532-017-0121-6).  
(page 181)
- [Woo87] W. K. Wootters.  
A Wigner-function formulation of finite-state quantum mechanics.  
*Annals of Physics*, 176(1):1–21, 1987.  
[doi:10.1016/0003-4916\(87\)90176-X](https://doi.org/10.1016/0003-4916(87)90176-X).  
(pages 15, 164, and 165)
- [WPGP+12] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd.  
Gaussian quantum information.  
*Reviews of Modern Physics*, 84(2):621, 2012.  
[doi:10.1103/RevModPhys.84.621](https://doi.org/10.1103/RevModPhys.84.621).  
(pages 3 and 17)

- [Wün98] A. Wünsche.  
Laguerre 2d-functions and their application in quantum optics.  
*Journal of Physics A: Mathematical and General*, 31(40):8267, 1998.  
[doi:10.1088/0305-4470/31/40/017](https://doi.org/10.1088/0305-4470/31/40/017).  
(page 18)
- [YUA<sup>+</sup>13] S. Yokoyama, R. Ukai, S. C. Armstrong, C. Sornphiphatphong, T. Kaji, S. Suzuki, J.-i. Yoshikawa, H. Yonezawa, N. C. Menicucci, and A. Furusawa.  
Ultra-large-scale continuous-variable cluster states multiplexed in the time domain.  
*Nature Photonics*, 7(12):982, 2013.  
[doi:10.1038/nphoton.2013.287](https://doi.org/10.1038/nphoton.2013.287).  
(pages 3 and 86)
- [YYK<sup>+</sup>16] J.-i. Yoshikawa, S. Yokoyama, T. Kaji, C. Sornphiphatphong, Y. Shiozawa, K. Makino, and A. Furusawa.  
Invited article: Generation of one-million-mode continuous-variable cluster state by unlimited time-domain multiplexing.  
*APL Photonics*, 1(6):060801, 2016.  
[doi:10.1063/1.4962732](https://doi.org/10.1063/1.4962732).  
(page 3)
- [Zei91] D. Zeilberger.  
The method of creative telescoping.  
*Journal of Symbolic Computation*, 11(3):195–204, 1991.  
[doi:10.1016/S0747-7171\(08\)80044-2](https://doi.org/10.1016/S0747-7171(08)80044-2).  
(page 128)
- [ZSS18] Q. Zhuang, P. W. Shor, and J. H. Shapiro.  
Resource theory of non-Gaussian operations.  
*Physical Review A*, 97(5):052317, 2018.  
[doi:10.1103/PhysRevA.97.052317](https://doi.org/10.1103/PhysRevA.97.052317).  
(page 3)
- [Żuk14] M. Żukowski.  
Temporal inequalities for sequential multi-time actions in quantum information processing.  
*Frontiers of Physics*, 9(5):629–633, 2014.  
[doi:10.1007/s11467-013-0400-2](https://doi.org/10.1007/s11467-013-0400-2).  
(page 172)
- [Zur01] W. H. Zurek.  
Sub-Planck structure in phase space and its relevance for quantum decoherence.  
*Nature*, 412(6848):712–717, 2001.  
[doi:10.1038/35089017](https://doi.org/10.1038/35089017).  
(page 95)
- [ZUZ<sup>+</sup>13] X. Zhang, M. Um, J. Zhang, S. An, Y. Wang, D.-l. Deng, C. Shen, L.-M. Duan, and K. Kim.  
State-independent experimental test of quantum contextuality with a single trapped ion.  
*Physical Review Letters*, 110:070401, Feb 2013.  
[doi:10.1103/PhysRevLett.110.070401](https://doi.org/10.1103/PhysRevLett.110.070401).  
(page 3)
- [ZWD<sup>+</sup>20] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, et al.  
Quantum computational advantage using photons.  
*Science*, 370(6523):1460–1463, 2020.  
[doi:10.1017/cbo9780511622748.004](https://doi.org/10.1017/cbo9780511622748.004).  
(page 182)

# ALPHABETICAL INDEX

- $\sigma$ -algebra, **20**, 52, 53, 77
- Bell inequality, **2**, 181
  - CV, 55, 63, 64
  - DV, **44**
- Born rule, **12**, **14**, 80, 169
- Characteristic function, **18**
- CHSH model, **37**, **38**, 160
- Coherent state, **17**
- Compatibility condition, **37**, 41
- Cone, **23**, 24–27, 57, 58, 60, 111, 114
- Contextual fraction
  - CV, 56, 63, 65, 71, 72, 86, 180, 182
  - DV, **43**, 177
- Contextuality, **2**
  - CV, 3, 46, **49**, 73, 75, 76, 85, 180, 182
  - DV, **39**, 86, 159, 172, 181
  - sequential, 160, 172, 173, 175–178
- Displacement operator, 90
  - CV, **17**, 18, 77, 78, 89, 99, 141
  - DV, **16**, 164, 166
- Empirical model
  - CV, **48**, 56, 73, 80, 83, 84
  - DV, **37**, 173, 174, 177
- EPR argument, **2**, 3, 172
- Extendability, 181
  - CV, **49**, 51, 52, 76, 80, 83
  - DV, **39**, 42
- FAB theorem
  - CV, 47, **51**, 72, 180
  - DV, **42**
- Fidelity, **13**, 88, 89, 93, 94, 96, 100, 108–110, 141, 142, 156–158
- Fourier transform, **19**, 85
- Hidden-variable model
  - CV, **50**, 83, 84
  - deterministic, **50**, 51, 84
  - factorisable, **51**, 51, 52, 84
  - DV, **41**, 160, 172–176
    - deterministic, **42**, 42
    - factorisable, **42**, 42
- Homodyne detection, **79**, 86–89, 93
- Hudson’s theorem, **4**, **19**
- Ladder operators, **17**
- Lagrangian
  - Grassmannian, **77**, 77, 80
  - subspace, **77**, 77, 80, 82
- Lagrangian method, 58, 59, 69
- Laguerre
  - function, **20**, 103, 104, 106, 109, 126, 140, 155
  - polynomial, **20**, 103
- Lasserre hierarchy, **28**, 65, 158, 180
- Linear program, **25**, 44, 56, 58, 61, 63, 65–68, 88, 89, 100, 102, 109–113, 116, 121, 127, 139, 143, 144, 158, 180
- Localising matrix, **32**, 69
- Markov kernel, **23**, 50, 74, 83
- Measurable
  - function, **22**, 50, 52, 81
  - space, **21**, 48, 50–52, 73, 77
    - finite product of, **21**
    - infinite product of, **21**
- Measure, **22**, 49–52, 56, 61–63, 83, 84, 126
- Measurement scenario
  - CV, **48**, 56, 77, 84, 85
  - DV, **35**
- Moment matrix, **32**, 69, 70, 104
- Moment sequence, **31**, 65, 103, 106, 107, 126, 138
- NV hierarchy, **34**, 171
- Parity operator, 18, 164
  - CV, **18**, 91
  - DV, **16**, 164
- Pauli gates
  - CV, **17**
  - DV, **15**
- Pauli measurement, 75–77, 85, 182
- Phase-space

- CV, **16**, 75, 77  
DV, **15**, 163, 164  
POVM, **14**, 79, 89, 90, 141  
PR box, 37, 40, 47, 160, 168  
Push-forward, **22**, 81, 84, 85  
Putinar's Positivellensatz, **30**, 72  
PVM, **14**, 33, 165, 168
- Quadrature, 3, **17**, 76–78
- Radon-Nikodym theorem, 84, 103, 110  
Riesz  
  functional, **31**, 106, 107  
  Haviland theorem, **31**, 106, 107, 152  
  Markov-Kakutani representation theorem, 26, 56  
  representation theorem, 83
- Semidefinite program, **27**, 65–71, 88, 89, 92, 93, 99, 100, 102, 116–124, 127, 146–150, 155, 156, 158
- Sheaf  
  event, 36, 48, 49, 82  
  restriction, 37, 48, 52, 82
- Strong duality, **25**  
  of linear programs, 26, 59, 111, 124, 125, 138, 156  
  of semidefinite programs, 28, 120, 121, 126, 154
- Sum-of-squares polynomial, **29**, 68, 70, 144, 145
- Symplectic form, **19**, 77, 80
- Total variation distance, **22**, 22, 60, 91  
Trace distance, **13**, 20, 91
- Wigner function, 3, **18**, 19, 20, 75, 76, 80, 83–85, 87, 88, 91, 158, 160, 164, 181  
Wigner negativity, 4, 75, 76, 83, 85, 87, 89–94, 96, 99, 108, 142, 158, 159, 165, 167, 178, 180, 182