



ÉCOLE
POLYTECHNIQUE
DE BRUXELLES

**From continuous-variable entanglement detection to
majorization lattices in quantum information theory**

Thèse présentée par Serge DESIDE

en vue de l'obtention du grade académique de Doctorat en Sciences de
l'ingénieur et technologie

Année académique 2025–2026

Sous la direction du Professeur Nicolas CERF, promoteur
Centre for Quantum Information and Communication

Jury de thèse :

Jean-Marc SPARENBERG (Université libre de Bruxelles, Président du jury)
Jérémy ROLAND (Université libre de Bruxelles, Secrétaire du jury)
Kiara HANSENNE (CEA Paris-Saclay)
John MARTIN (Université de Liège)
Alexssandre de OLIVEIRA JUNIOR (Technical University of Denmark)
Nicolas CERF (Université libre de Bruxelles)



Typesetting. The typographical design of this thesis is partly based on the [TeXtured L^AT_EX template](#) developed by Jonáš Dujava.

Abstract

Quantum entanglement is one of the defining facets of quantum physics and a central resource in quantum information theory, underlying essential protocols such as superdense coding and quantum teleportation. Therefore, understanding how entanglement can be detected and manipulated constitutes a fundamental challenge in the field.

The first part of this thesis concerns entanglement detection in continuous-variable quantum systems. We focus on states of the electromagnetic field and study separability criteria based on moments of the operators obtained by applying partial transposition or the realignment map to the density operator. We first extend criteria based on moments of the partially transposed density operator to the continuous-variable setting and analyze their detection capabilities for several representative quantum-optical states, both Gaussian and non-Gaussian. For the strongly non-Gaussian class of NOON states, we study in detail the effect of various experimental imperfections on entanglement detection. Then, we introduce a moment-based entanglement criterion derived from the realignment map and specialize it to continuous-variable systems. We study this criterion for Gaussian states, entangled cat states, and NOON states, and compare its detection capability with that of the partial-transposition moment criteria. For both approaches, we derive measurement schemes based on the multicopy technique, using collective measurements on several identical and independent copies of the state, in order to access the moments of interest.

The second part of this thesis concerns entanglement manipulation and focuses on the theory of majorization, more specifically the structure of the majorization lattice. Majorization provides the standard framework for analyzing deterministic transformations between bipartite pure entangled states under local operations and classical communication. Since majorization does not define a total order, not all probability vectors are comparable. Nevertheless, probability vectors of fixed dimension endowed with majorization form a lattice, whose two fundamental operations are the meet (greatest lower bound) and the join (least upper bound). We exploit this lattice structure to study probabilistic transformations between bipartite pure entangled states. In particular, we construct two probabilistic conversion protocols based respectively on the meet and join, which we call the thrifty and greedy protocols. We prove the optimality of these protocols and study their relation to the standard optimal probabilistic conversion protocol. While the greedy protocol is closely related to the standard construction, the thrifty protocol allows the entanglement resource to be better preserved on average. We finally investigate the geometry of the majorization lattice using the total variation distance. We introduce the notion of a majorization diamond, consisting of all probability distributions lying between two fixed distributions in the majorization order, by analogy with the causal diamond in relativistic spacetime. From this construction, we derive an improved bound for the Shannon entropy that goes beyond the Schur-concavity property. We also derive a continuity bound for the Shannon entropy within a fixed majorization diamond, and extend our result to the von Neumann entropy of quantum states.

Acknowledgments

Tout d'abord, j'aimerais sincèrement te remercier, Nicolas. D'emblée, ton cours de BA3 a significativement contribué à affermir mon goût pour la mécanique quantique. Tu m'as ensuite permis de mettre en pratique ce formalisme lors de mon mémoire, avant de me proposer de prolonger dans une thèse les questionnements qui y avaient été esquissés. J'ai toujours trouvé les discussions scientifiques avec toi stimulantes. Je te suis profondément reconnaissant de m'avoir donné la possibilité et les moyens d'entamer ce doctorat au QuIC. Merci pour la liberté et la confiance que tu m'as accordées.

Michael, merci d'avoir été disponible tout au long de ma thèse pour discuter de projets liés à la majorization. Je te remercie de m'avoir invité à travailler avec toi à Copenhague alors que je venais de commencer ma thèse. Au final, je suis content que nous ayons réussi à associer la lattice de majorization à ton sujet privilégié des bornes de continuité !

Just after I had finished my first article, Tobi, you approached me and asked whether I would be interested in working with you on these p_n -PPT criteria. Thank you for that opportunity. I am glad to have had the chance to collaborate with you during my PhD thesis.

C'est notamment grâce à toi, Alex, que j'ai pu me replonger avec plaisir dans l'étude de la lattice de majorization. Les discussions avec toi sur le sujet ont été très motivantes et inspirantes. Je suis convaincu que tu te débrouilleras bien dans cette aventure qu'est la thèse ! N'hésite pas à me contacter si tu souhaites remonter les marches de la gare de Lozère un jour.

Célia, tu as d'abord supervisé mon mémoire, puis nous avons appris à mieux nous connaître lors de cette conférence à Copenhague, dont je me souviendrai longtemps. Durant ma première année de thèse, tu as été présente pour moi, tant sur le plan scientifique que moral, et je t'en remercie. Nous avons ensuite passé l'agrégation, non sans difficultés, mais nous en sommes ressorti-es indemnes (je crois).

Tu as été l'assistant de mon premier TP (tu avais bien fait la pub du QuIC pour que je me le rappelle) puis mon superviseur lors de mon mémoire. J'ai donc pu commencer et finir en beauté mes études en polytech grâce à toi, Matthieu. Merci !

Thank you Vojta for having worked with me on these multicopy realignment criteria. I'll remember our discussions on Brussels' architectural history.

Léo. Tout d'abord, j'aimerais te remercier pour le soutien amical que tu m'as apporté durant ces trois années (comme tu le sais, nous n'étions malheureusement pas amis durant notre première année). Je n'aurais pu espérer meilleure compagnie dans les *bureaux du fond*. Merci pour ces échanges permanents qui ont été déterminants afin que je réalise (notamment) qu'il est plus qu'urgent d'inclure les autres animaux dans nos cercles de considération morale. J'espère que nous aurons encore de nombreuses occasions d'affiner nos jugements moraux et, surtout, politiques.

Arne, ou plutôt devrais-je dire Simba, merci pour la bonne ambiance durant ces quatre années. J'attends avec impatience tes prochaines énigmes mathématiques.

Just before you left Brussels, Yuxin, we became close friends. I hope we'll have the opportunity to meet again someday, even though our lives have taken different directions.

Thank you also to all the other members of QuIC whom I had the chance to get to know: Arthur, Benoît, Clément, Enea, Guy, Hamed, Jérémie, Julian, Julien, Joseph, Lefteris, Lin-Qing, Marco, Marin, Nicola, Ognyan, Pascale, Ravi, Timothée, and Zixuan.

Je souhaite également remercier les pédagogues et enseignant-es qui m'ont transmis leur goût des mathématiques et de la physique, en particulier Boris Meynsbrughen et Jean-Marc Sparenberg.

Je remercie le F.R.S.-FNRS d'avoir financé ces quatre années de recherche.

I would also like to thank the members of the jury for dedicating their time and effort to reading this thesis.

Mon cher Jeremy, depuis le premier jour où nos regards se sont croisés, je l'ai su : tu serais mon compagnon de route. J'ai énormément de chance de te connaître et de pouvoir compter sur ton amitié. Nous avons traversé de nombreuses épreuves, et nous en vivrons beaucoup d'autres, j'en suis convaincu. Un jour, nous réanimons le collège Rainbow.

Merci beaucoup pour ton soutien et l'amitié que tu me portes, Louise. Nous avons partagé plein de choses, et j'espère que nous continuerons dans cette voie. À quand notre prochain voyage ?

Pistache, tu es le plus beau et le plus gentil des chats. Je t'aime beaucoup, vraiment. J'espère qu'on se reverra souvent. Machita, tu es un chouette chat. J'espère qu'on se reverra de temps en temps.

Je souhaite remercier mes ami-es de m'avoir accompagné durant ces dernières années, en particulier Alix, Gabriel, Ludéric, Mohamed, Nicolas, Raphael et Tania.

Merci à mes parents pour la confiance, la liberté et la bonté de l'amour qu'ils m'ont prodigué. Et merci à Marion, Mathieu et Mayinga pour la vie que nous avons un temps partagée.

Merci aux personnes qui m'ont accompagné et encouragé au cours de ces derniers mois de thèse : Jeremy, Léo, Louise, Maman et Rosa.

Enfin, merci à toutes les personnes dont les noms ne sont pas cités mais qui ont compté pour moi.

List of publications

The present thesis is based on the following publications and manuscripts

Published in peer-reviewed scientific journals

- [62] Serge Deside, Matthieu Arnhem, Célia Griffet, and Nicolas J. Cerf. “Probabilistic pure state conversion on the majorization lattice”. *Physical Review Research* 6, 023156, May 2024.

This article is the subject of Chapter 8.

- [63] Serge Deside, Tobias Haas, and Nicolas J. Cerf. “Detecting strongly non-Gaussian entanglement”. *Physical Review Research* 8, 033064, July 2026.

This article is the subject of Chapter 5.

Manuscripts in final stages of preparation

- [64] Serge Deside, Michael G. Jabbour, and Nicolas J. Cerf. “Geometric aspects of majorization: Relating distances and entropies in the majorization lattice”.

This article is the subject of Chapter 9.

- [65] Serge Deside, Vojtěch Kala, and Nicolas J. Cerf. “Measurable realignment criteria for entanglement in continuous-variable quantum systems”.

This article is the subject of Chapter 6.

Manuscript not discussed in the present thesis

- [223] Alexander Stévin, Michael G. Jabbour, Serge Deside, Nicolas J. Cerf, “Majorization precursors to supermodularity and subadditivity on the majorization lattice”, May 2026. [arXiv:2605.30331 \[cs.IT\]](https://arxiv.org/abs/2605.30331).

Contents

Abstract	i
Acknowledgments	iii
List of publications	v
1 Introduction	1

I Preliminaries

2 Quantum entanglement	9
2.1 What is entanglement?	10
2.2 Manipulating entangled states.	12
2.2.1 Local operations and classical communication	12
2.2.2 Deterministic transformations of bipartite pure states	13
2.2.3 Probabilistic transformations of bipartite pure states.	16
2.2.4 Asymptotic transformations	17
2.3 Detecting discrete-variable entanglement	19
2.3.1 Majorization-based criteria	19
2.3.2 Positive-partial-transpose criterion	20
2.3.3 Realignment criterion	21
2.4 Detecting continuous-variable entanglement	22
2.4.1 A glimpse at quantum optics	23
2.4.2 Second-moment-based criteria	24
2.4.3 Entropy-based criterion.	26
3 Quantum optics	29
3.1 Quantizing the electromagnetic field	29
3.2 Quantum-optical states	32
3.2.1 Number states and the Fock basis	32
3.2.2 Gaussian states.	33
3.2.2.1 Coherent states	34
3.2.2.2 Thermal states	35
3.2.2.3 Squeezed states	36
3.2.2.4 Two-mode squeezed vacuum states	36

3.2.3 Non-Gaussian states	39
3.2.3.1 Single-mode cat states	39
3.2.3.2 Two-mode cat states	39
3.2.3.3 NOON states	40
3.3 Quantum-optical operations.	41
3.3.1 Gaussian unitaries	41
3.3.1.1 Symplectic formalism.	42
3.3.1.2 Single-mode operations	43
3.3.1.3 Two-mode operations	46
3.3.1.4 General Gaussian unitary operation	48
3.3.2 Measurements	49
3.3.2.1 Homodyning and heterodyning.	49
3.3.2.2 Photon-number-resolving detection	50

II Entanglement detection

4 Multicopy technique	55
4.1 Description of the method	55
4.2 Applications of the method	57
4.2.1 Purity of a quantum state	57
4.2.2 Higher-order moments	60
4.2.3 Matrices of mode-operator moments	62
5 Detecting strongly non-Gaussian entanglement	67
5.1 Introduction	67
5.2 PT-moment-based criteria for continuous variables	68
5.2.1 Infinite hierarchies of criteria	68
5.2.1.1 From the Stieltjes moment problem	68
5.2.1.2 From Descartes' rule of signs	69
5.2.2 Third-order criteria	69
5.3 Implementation	71
5.3.1 Multicopy method	71
5.3.2 Measuring PT moments	72
5.3.3 Optical circuits for low-order moments	74
5.3.3.1 Purity p_2	75
5.3.3.2 Third-order PT moment p_3	75
5.4 Benchmarks.	76
5.4.1 Comparison	76
5.4.1.1 Shchukin–Vogel criteria	76

5.4.1.2	Criteria based on non-local variables	78
5.4.2	Gaussian entanglement	78
5.4.2.1	Symplectic formalism and Simon's criterion	78
5.4.2.2	p_n -PPT criteria for Gaussian states	79
5.4.2.3	Optimal PT-moment-based criterion for Gaussian entanglement	80
5.4.3	Strongly non-Gaussian entanglement from coherent states	80
5.4.3.1	Mixed entangled cat states	80
5.4.3.2	High Harmonic Generation (HHG)	81
5.4.4	Strongly non-Gaussian entanglement from Fock states	82
5.4.4.1	NOON states	82
5.4.4.2	Other pure-state families	84
5.5	Application	85
5.5.1	Imperfections	85
5.5.1.1	Losses	85
5.5.1.2	Noisy copies	86
5.5.1.3	Finite statistics	86
5.5.2	NOON states – a case study	87
5.5.2.1	Analysis of individual errors	87
5.5.2.2	Full simulation	88
5.6	Discussion	90
6	Measurable realignment criteria for continuous variables	93
6.1	Preliminaries	93
6.2	Beyond the weak realignment criterion	96
6.3	Implementation	101
6.4	Benchmarks	105
6.4.1	Gaussian states	105
6.4.2	Non-Gaussian states	109
6.4.2.1	Schrödinger cat states	109
6.4.2.2	Lossy NOON states	110
6.5	Discussion	112

III Majorization lattice

7	Lattice structure of majorization	117
7.1	Theory of majorization	117
7.1.1	Definition with inequalities	117
7.1.2	Equivalent formulations	120
7.1.3	Lorenz curves	122

7.1.4	The Shannon entropy: a weaker notion of disorder	122
7.2	A brief look at the theory of lattices	124
7.3	Theory of the majorization lattice	127
7.3.1	General definitions	127
7.3.2	Constructing the meet and the join	128
7.3.3	Properties	131
8	Probabilistic pure-state conversion	137
8.1	Introduction	137
8.2	From the majorization lattice to the entanglement lattice	139
8.3	Probabilistic state-conversion protocol	140
8.4	Greedy and thrifty protocols on the majorization lattice	143
8.5	Generalization to a collection of initial or final states.	151
8.6	Conclusion	155
9	Geometric aspects of majorization	157
9.1	Introduction	157
9.2	Mathematical preliminaries	159
9.3	Geometry of the majorization lattice	160
9.4	Characterization of the majorization diamond	166
9.5	Beyond the Schur-concavity of the Shannon entropy in the majorization lattice	170
9.6	Continuity bound for the Shannon entropy	174
9.7	Continuity bound for the von Neumann entropy	186
9.8	Discussion and perspectives	187
10	Conclusion and perspectives	191
	Declaration on the use of Artificial Intelligence	195

IV Appendices

A	Appendix to Chapter 5	199
A.1	Gaussian states	199
A.1.1	Two-mode squeezed state at finite temperature.	199
A.1.2	Higher-order criteria	199
A.2	Mixed entangled cat states	200
A.3	High Harmonic Generation (HHG)	201

A.4	NOON states	201
A.4.1	Mode-operator moments	201
A.4.2	Lossy NOON states	202
A.4.3	Particle-number output distributions	203
A.4.4	Statistical error models	204
B	Appendix to Chapter 6	209
B.1	Properties of w_n and r_n	209
B.2	Properties of the operator Schmidt decomposition.	211
B.3	Second-order bound	212
B.4	Higher-order inequalities	213
B.5	Werner state.	217
B.6	General measurement techniques for w_n and r_{2n}	217
B.6.1	Multicopy observable for w_n	219
B.6.2	Multicopy observable for r_{2n}	221
B.6.3	Interferometric implementation	222
B.6.3.1	Measurement of w_n	222
B.6.3.2	Measurement of r_{2n}	224
C	Appendix to Chapter 9	225
C.1	Majorization minimum of the ball centered on a given vector	225
C.2	Further tools and results for the proofs	226
	Bibliography	235

1 Introduction

Together with general relativity, quantum physics constitutes one of the major breakthroughs that marked the beginning of modern physics. Since its formulation in the early twentieth century, quantum theory has provided the most successful framework for describing microscopic phenomena, that is, in regimes where the relevant physical quantities reach quantum scales set by Planck's reduced constant, $\hbar \approx 10^{-34} \text{ m}^2\text{kg/s}$. Importantly, quantum physics has also had a major technological impact. In particular, the development of semiconductor physics and the invention of the transistor [15] played a central role in the growth of modern electronics and computing.

Towards the end of the twentieth century, quantum physics began to be regarded not only as the basis for building classical information-processing devices, but also as a possible resource for information processing itself [66, 83]. This led to theoretical protocols for quantum computation [23, 67, 213], cryptography [20], and communication [21, 22]. Retrospectively, these advances marked the birth of quantum information theory [172], at the intersection of quantum physics and Shannon's information theory [49, 208]. Much of the early work in quantum information was formulated in terms of discrete quantum systems, most notably the quantum bit, or qubit [202]. However, quantum information can also be encoded in continuous degrees of freedom [156], in particular in the quadrature observables of electromagnetic-field modes. Such continuous-variable systems arise naturally in quantum optics and are especially well suited to quantum communication [28, 243]. They encompass a broad range of experimental platforms, including photonic systems [10, 70, 186, 201] and ultracold quantum gases [81, 107, 150, 152, 177, 225].

Since the early formalization of quantum physics, entanglement has been regarded as an oddity, famously associated with Einstein's description of a "spooky action at a distance" [75]. Its usefulness became clear within the framework of quantum information through protocols such as entanglement-based quantum key distribution [78], superdense coding [22], and quantum teleportation [21]. Nowadays, entanglement is regarded as an essential quantum resource [42, 183, 232], with applications ranging from quantum computing [144, 147, 148, 166, 191, 218] and simulation [7, 24, 52, 94, 229, 236] to quantum metrology [54, 98, 175, 181]. Beyond quantum information processing, entanglement also plays a central role in the study of many-body systems, including area laws and thermalization [2, 31, 36, 60, 61, 77, 112, 135, 146], as well as in questions at the interface between quantum theory and gravitation [25, 34, 220, 221].

A central theoretical and practical question then arose: how can one determine whether a given quantum state is entangled? Entanglement detection has consequently developed into a broad field, with numerous criteria adapted to different physical settings and levels of experimental information [108, 133]. A seminal result in this direction was provided by the

positive partial transpose (PPT) criterion, introduced by Peres [178] and further characterized by the Horodecki family [129]. The criterion was later extended to continuous-variable systems, where partial transposition admits a natural formulation in the phase space of canonical quadrature observables [216].

The first question addressed in this thesis concerns the efficient detection of continuous-variable entanglement beyond the Gaussian regime. Gaussian states are completely characterized by their first and second moments, which allows entanglement criteria to be formulated directly at the covariance-matrix level [72, 99, 151, 160, 216, 244]. This simplification generally disappears for non-Gaussian states. Existing criteria based on entropies [84, 85, 92, 93, 195, 200, 228, 238, 239] or higher-order mode-operator moments [163, 164, 209] can detect certain non-Gaussian states, but may become insensitive or experimentally demanding when the relevant correlations involve high orders. This motivates the search for other types of criteria that retain sensitivity to genuinely non-Gaussian correlations while remaining experimentally accessible.

In this context, the moments of the partially transposed (PT) state provide a particularly useful set of quantities. Such PT moments have appeared in replica approaches to entanglement negativity [32, 33], and were more recently proposed as experimentally accessible probes of mixed-state entanglement [80]. In finite-dimensional systems, systematic hierarchies of PT-moment conditions have subsequently been developed whose complete set recovers the PPT condition [169, 252]. Importantly, even low-order moments of the partially transposed density operator may implicitly contain information about mode-operator correlations of arbitrarily high order, which makes them particularly relevant for strongly non-Gaussian states. In parallel, the realignment criterion [193, 194] provides a complementary separability test that is not based on partial transposition. A central aim of this thesis is therefore to make these quantities experimentally accessible in continuous-variable systems and to determine whether they remain useful for strongly non-Gaussian states under realistic imperfections.

To this end, we rely on the multicopy technique [29], which provides a natural way of accessing nonlinear polynomial functions of a density operator. The method consists in jointly processing identical and independent copies of a quantum state through suitable unitary operations before measurement. It is particularly well suited to the present problem because PT moments are nonlinear functions of the state, while admitting natural realizations as expectation values of observables acting on multiple replicas. In the continuous-variable setting, multicopy observables were first used to construct witnesses of optical nonclassicality [9, 104], based on matrices of moments of the mode operators [210], and were later extended to entanglement detection [105], using the corresponding bipartite moment matrices [209]. In Chapter 5, we use this multicopy framework to extend PT-moment entanglement criteria to continuous-variable systems. In Chapter 6, we apply a related strategy to construct measurable moment-based realignment criteria for continuous-variable systems.

The second question addressed in this thesis concerns the information carried by the lattice structure of majorization beyond the majorization relation itself. We first examine this question in the context of entanglement manipulation. Deterministic single-copy entanglement transformations of bipartite pure states under local operations and classical communication (LOCC) have been shown to be characterized by the so-called majorization relation [161, 171]. This framework was soon extended to probabilistic state conversions [233]. More generally, majorization provides a strong notion of disorder that has proved particularly useful in a variety of quantum-information problems [38, 91, 97, 123, 185, 217]. Since majorization does not define a total order, some probability vectors are incomparable. Remarkably, the set of probability vectors of fixed dimension nevertheless forms a lattice under majorization: every pair possesses a greatest lower bound and a least upper bound, referred to as the meet and join, respectively [46]. This structure has been used to reformulate problems ranging from quantum thermodynamics [59, 149] and quantum coherence [251] to quantum steering [250] and quantum entanglement manipulation [26, 199].

In Chapter 8, we exploit these notions of meet and join to construct two alternative optimal probabilistic conversion protocols between incomparable bipartite pure states, which we call the thrifty and greedy protocols. While these protocols reach the same maximal success probability as Vidal's protocol [233], the lattice structure makes it possible to distinguish them through the entanglement retained when the transformation fails. In particular, the thrifty protocol leaves a more entangled residual state than the greedy one and therefore better preserves the entanglement resource on average.

Finally, in Chapter 9, we further characterize the geometry of the majorization lattice by introducing the notion of a majorization diamond, drawing an analogy with causal diamonds formed by past and future light cones in special relativity [139]. A majorization diamond consists of the probability distributions lying between fixed lower and upper bounds in the majorization order. We then connect this construction to continuity bounds. Such bounds control the variation of entropic quantities between nearby probability distributions or quantum states and have numerous applications in quantum information, including the estimation of quantum-channel capacities [12, 16, 212, 247]. In particular, we use the additional majorization constraints to derive new continuity bounds for the Shannon and von Neumann entropies.

This thesis is organized as follows:

The first part is devoted to introducing the preliminary notions required for the subsequent chapters. In Chapter 2, we review the concept of entanglement, beginning with its definition and the LOCC framework used to describe entanglement manipulation. We then review deterministic and probabilistic single-copy conversion protocols for bipartite pure states, as well as asymptotic transformations. Finally, we discuss entanglement certification in discrete-

variable systems and its extension to the continuous-variable setting. Chapter 3 is dedicated to introducing the quantum optics concepts needed in this thesis. We first introduce the quantization of the electromagnetic field and the corresponding operator description in Fock space. We then present representative Gaussian and non-Gaussian quantum-optical states, for both single-mode systems, such as coherent and Fock states, and two-mode systems, such as NOON states and the two-mode squeezed vacuum. Finally, we discuss the main quantum-optical operations considered in this thesis, both Gaussian and non-Gaussian, including unitary transformations and measurements.

The second part is devoted to our investigation of entanglement certification in continuous-variable systems. We begin in Chapter 4 by reviewing the multicopy technique. After motivating the approach and explaining how it works, we present several examples, including the measurement of the purity and higher-order moments of the density operator. We close the chapter by discussing recent applications of multicopy measurements in quantum optics, in particular for witnessing nonclassicality and entanglement. In Chapter 5, we then study the resulting PT-moment criteria in detail, comparing the different low-order conditions and benchmarking their detection capabilities against standard approaches for Gaussian states, coherent superpositions and mixtures of Gaussian states, and strongly non-Gaussian states in Fock space. We finally investigate their robustness to losses, imperfect state copies, and finite statistics, with particular emphasis on NOON states. Chapter 6 develops the corresponding realignment-based approach beyond its lowest-order form. We analyze the resulting criteria for Gaussian states, two-mode entangled cat states, and NOON states, and discuss their continuous-variable measurement schemes.

The third part revolves around majorization and our original contributions to its use in quantum information. In Chapter 7, we define the majorization relation and present its main equivalent formulations, together with its visual representation in terms of Lorenz curves. We then introduce the general concept of a lattice before specializing it to the majorization order and deriving the meet and join of probability vectors. In Chapter 8, we use these lattice elements to identify the optimal common resource and optimal common product states associated with bipartite pure states, before constructing the corresponding probabilistic conversion protocols. We subsequently extend this framework to situations involving several possible initial or final states. Finally, in Chapter 9, we study the interplay between the total variation distance and the meet and join, characterize the geometry of the resulting majorization diamonds, and derive entropic inequalities that go beyond standard Schur-concavity. We then apply these geometric results to obtain continuity bounds for the Shannon entropy and, subsequently, for the von Neumann entropy.

The fourth part consists of appendices that complement the mathematical derivations presented in Chapters 5, 6, and 9.

Contributions of the thesis:

The original contributions of this thesis are presented in Chapters 5 and 6, which concern entanglement detection, and in Chapters 8 and 9, which concern the majorization lattice.

Part I

Preliminaries

2 Quantum entanglement

Entanglement is among the most fundamental and, at the same time, most puzzling features of quantum physics. One reason behind this is that entangled systems can display correlations between measurements that go beyond our classical understanding of physical reality. These correlations were at the heart of the seminal 1935 paper by Einstein, Podolsky and Rosen (EPR), in which they questioned whether quantum mechanics provides a complete description of physical reality or not [76]. Their argument initiated a long debate on whether the unusual correlations predicted by quantum mechanics reflected a limitation of the theory or a genuinely non-classical feature exhibited by Nature. Almost thirty years later, in 1964, Bell showed that the predictions of quantum mechanics are incompatible with those of local hidden-variable theories, thereby opening the way to experimental tests of this incompatibility [18]. Subsequently, experiments supported the quantum-mechanical predictions [11, 87], while more recent tests strengthened this conclusion by overcoming the main experimental limitations affecting earlier experiments [100, 119, 207].

In quantum information theory, however, entanglement is not merely regarded as an anomaly, but rather as a resource enabling certain tasks. Indeed, several fundamental protocols, such as quantum teleportation [21] and dense coding [22], rely on the use of maximally entangled states. It is therefore essential to understand what entanglement is, how entangled states can be manipulated, and how one can determine whether a given quantum state is entangled or not.

In this chapter, we first introduce the mathematical definition of entanglement in Section 2.1. Then, in Section 2.2, we present an overview of the manipulation of entangled states under physically meaningful constraints. Finally, we discuss methods for certifying entanglement in discrete-variable systems in Section 2.3, before turning to continuous-variable systems in Section 2.4. If not specified, this introductory chapter is mainly based on Refs. [133, 172].

2.1 What is entanglement?

Consider a bipartite¹ quantum system shared by two parties, traditionally called Alice (A) and Bob (B). The corresponding Hilbert space is written as

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B, \quad (2.1)$$

where $\mathcal{H}_A$ and $\mathcal{H}_B$ describe Alice's and Bob's subsystems, respectively, and $\otimes$ denotes the tensor product. In what follows, we assume that the Hilbert spaces are of finite dimension, unless otherwise specified. A general state of the joint system is represented by a density operator $\hat{\rho}_{AB}$ acting on $\mathcal{H}_{AB}$.

The simplest joint states are those for which Alice's and Bob's systems are prepared independently, which we describe with the following definition.

Definition 2.1 (Product state). A bipartite state $\hat{\rho}_{AB}$ is called a *product state* if it can be written as

$$\hat{\rho}_{AB} = \hat{\rho}_A \otimes \hat{\rho}_B, \quad (2.2)$$

for some states $\hat{\rho}_A$ and $\hat{\rho}_B$ of the two subsystems.

A product state contains no correlations between Alice's and Bob's subsystems. However, not every correlated state is entangled. Indeed, Alice and Bob may produce correlations by using shared classical randomness. Suppose, for example, that they toss a shared fair coin and prepare $|0\rangle_A |0\rangle_B$ when the outcome is heads, and $|1\rangle_A |1\rangle_B$ when it is tails. The resulting state is the mixed state

$$\hat{\rho}_{AB}^{\text{cc}} = \frac{1}{2} |0\rangle_A \langle 0|_A \otimes |0\rangle_B \langle 0|_B + \frac{1}{2} |1\rangle_A \langle 1|_A \otimes |1\rangle_B \langle 1|_B. \quad (2.3)$$

This state is not a product state. Indeed, if Alice and Bob both measure in the computational basis $\{|0\rangle, |1\rangle\}$, their outcomes are perfectly correlated. Nevertheless, these correlations have a purely classical explanation, namely the value of the shared coin toss. The distinction between classical correlations and entanglement is captured by the notion of separability.

¹Entanglement has also been described for systems shared between more than two parties, often termed *multipartite entanglement*, see, e.g., Ref. [132], for a concise review.

Definition 2.2 (Separable and entangled states). A bipartite state $\hat{\rho}_{AB}$ is called *separable* if it can be written as

$$\hat{\rho}_{AB} = \sum_i p_i \hat{\rho}_A^{(i)} \otimes \hat{\rho}_B^{(i)}, \quad \text{with } p_i \geq 0, \text{ and } \sum_i p_i = 1, \quad (2.4)$$

and where $\hat{\rho}_A^{(i)}$ and $\hat{\rho}_B^{(i)}$ are local quantum states of the two subsystems. A state that is not separable is called *entangled*.

A separable state may therefore contain correlations, as in Eq. (2.3), but these correlations can always be understood as follows: A classical variable i is sampled with probability p_i , and, conditioned on this value, Alice and Bob locally prepare the states $\hat{\rho}_A^{(i)}$ and $\hat{\rho}_B^{(i)}$. Entanglement refers precisely to correlations that cannot be generated in this way.

For pure states, the distinction between separability and entanglement takes a particularly simple form.

Proposition 2.1. A bipartite pure state $|\psi\rangle_{AB}$ is separable if and only if it is a product state, that is, if and only if there exist vectors $|\phi\rangle_A$ and $|\chi\rangle_B$ such that

$$|\psi\rangle_{AB} = |\phi\rangle_A \otimes |\chi\rangle_B. \quad (2.5)$$

Otherwise, $|\psi\rangle_{AB}$ is entangled.

A useful characterization of bipartite pure-state entanglement is provided by the Schmidt decomposition.

Proposition 2.2 (Schmidt decomposition). For every bipartite pure state $|\psi\rangle_{AB} \in \mathcal{H}_{AB}$, there exist orthonormal sets $\{|u_i\rangle_A\}_{i=1}^r$ and $\{|v_i\rangle_B\}_{i=1}^r$ such that

$$|\psi\rangle_{AB} = \sum_{i=1}^r \sqrt{\lambda_i} |u_i\rangle_A \otimes |v_i\rangle_B, \quad (2.6)$$

where

$$\lambda_i > 0, \quad \sum_{i=1}^r \lambda_i = 1. \quad (2.7)$$

The numbers λ_i are called the Schmidt coefficients, and r is called the Schmidt rank.

The Schmidt decomposition directly identifies whether a pure state is entangled or not. A pure state is separable if and only if its Schmidt rank is equal to one. Equivalently, it is entangled if and only if at least two Schmidt coefficients are non-zero.

As a first important example, consider the two-qubit singlet state

$$|\psi^-\rangle = \frac{1}{\sqrt{2}} (|0\rangle_A |1\rangle_B - |1\rangle_A |0\rangle_B). \quad (2.8)$$

Up to absorbing the minus sign into one of the local basis vectors, this is a Schmidt decomposition with

$$\lambda_1 = \lambda_2 = \frac{1}{2}. \quad (2.9)$$

Hence, the singlet state is entangled. Its entanglement is also reflected in its reduced states. Tracing out Bob's subsystem gives

$$\hat{\rho}_A = \text{Tr}_B (|\psi^-\rangle \langle \psi^-|) = \frac{\hat{1}_A}{2}. \quad (2.10)$$

Thus, although the joint state is pure, Alice's reduced state is maximally mixed. This is impossible for a pure product state, for which both reduced states are pure.

2.2 Manipulating entangled states

We turn now to the question of what can be done with entangled states. The idea is to treat entanglement as a resource shared between distant parties [42]. In this resource-theoretic picture, separable states are defined as free states, since they contain no entanglement, which naturally leads to the definition of free operations as operations that do not create entanglement from separable states. In this context, widely considered free operations, deemed relevant for this thesis, are the local operations supplemented with classical communication, usually abbreviated as LOCC.

In this section, we first introduce LOCC transformations. Then, we discuss deterministic and probabilistic transformations between single copies of bipartite pure states. Finally, we turn to the asymptotic regime, where one studies conversion rates rather than only the possibility of a single transformation.

2.2.1 Local operations and classical communication

Consider a bipartite system shared by two distant parties, Alice and Bob, with Hilbert space $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$. A local operation is any operation that Alice or Bob can perform in their own laboratory, including quantum operations on their subsystem and classical processing of local outcomes. For instance, Alice may apply a unitary transformation, perform a measurement, or more generally apply a quantum instrument on $\mathcal{H}_A$, while Bob may do the same on $\mathcal{H}_B$. What is not allowed is a joint quantum operation acting directly on $\mathcal{H}_{AB}$.

With the help of classical communication, the parties are allowed to coordinate their local operations. Alice may perform a measurement and send the outcome to Bob, who then chooses a local operation depending on the message received. The protocol may continue for several rounds, with each operation depending on the previous rounds.

Thus, LOCC protocols are built from local quantum operations and classical messages, and do not involve the transmission of a quantum system between the parties, nor any global operation coupling Alice's and Bob's subsystems. This restriction is what makes entanglement a useful resource. If Alice and Bob start from a separable state, LOCC cannot create entanglement between them. Local operations only modify local states, while classical communication only correlates the parties through classical information. Hence, LOCC transformations preserve the set of separable states [43].

We write

$$\hat{\rho}_{AB} \xrightarrow{\text{LOCC}} \hat{\sigma}_{AB} \quad (2.11)$$

when there exists a deterministic LOCC protocol transforming the initial state $\hat{\rho}_{AB}$ into the final state $\hat{\sigma}_{AB}$. We also consider probabilistic transformations. In that case, the desired target state is obtained only for some outcomes of the protocol. We write

$$\hat{\rho}_{AB} \xrightarrow[p]{\text{LOCC}} \hat{\sigma}_{AB} \quad (2.12)$$

when there exists an LOCC protocol producing $\hat{\sigma}_{AB}$ with probability p . If it fails, the parties generally obtain another residual state.

2.2.2 Deterministic transformations of bipartite pure states

We first consider deterministic transformations between single copies of bipartite pure states. Let Alice and Bob initially share a state $|\psi\rangle_{AB}$ and suppose that they want to convert it into a target state $|\phi\rangle_{AB}$ using only LOCC. We write

$$|\psi\rangle_{AB} \xrightarrow{\text{LOCC}} |\phi\rangle_{AB} \quad (2.13)$$

when the transformation can be achieved with certainty.

Let

$$|\psi\rangle_{AB} = \sum_{i=1}^d \sqrt{\lambda_{\psi}^{(i)}} |i\rangle_A |i\rangle_B, \quad |\phi\rangle_{AB} = \sum_{i=1}^d \sqrt{\lambda_{\phi}^{(i)}} |i\rangle_A |i\rangle_B \quad (2.14)$$

be Schmidt decompositions of the two states. We denote the corresponding Schmidt vectors by

$$\lambda_{\psi} = \left(\lambda_{\psi}^{(1)}, \dots, \lambda_{\psi}^{(d)} \right)^T, \quad \lambda_{\phi} = \left(\lambda_{\phi}^{(1)}, \dots, \lambda_{\phi}^{(d)} \right)^T, \quad (2.15)$$

and assume that their components are ordered non-increasingly.

We briefly present now the notion of majorization [161] (see Chapter 7 for a more in-depth overview of majorization theory). For two probability vectors $\mathbf{x}$ and $\mathbf{y}$ of at most d non-zero components and arranged non-increasingly, we write $\mathbf{x} < \mathbf{y}$, and say $\mathbf{x}$ is majorized by $\mathbf{y}$, if

$$\sum_{i=l}^d x_i \geq \sum_{i=l}^d y_i, \quad \forall l = 2, \dots, d, \quad (2.16)$$

with equality for $l = 1$. Intuitively, $\mathbf{x} < \mathbf{y}$ means that $\mathbf{x}$ is more uniform than $\mathbf{y}$. The deterministic transformation of bipartite pure states is characterized by Nielsen's theorem.

Theorem 2.1 (Nielsen [171]). *Let $|\psi\rangle_{AB}$ and $|\phi\rangle_{AB}$ be two bipartite pure states with Schmidt vectors λ_ψ and λ_ϕ , respectively. Then,*

$$|\psi\rangle_{AB} \xrightarrow{\text{LOCC}} |\phi\rangle_{AB} \iff \lambda_\psi < \lambda_\phi. \quad (2.17)$$

Thus, deterministic LOCC convertibility reduces to a finite set of inequalities between Schmidt coefficients. This is closely related to the notion of *entanglement monotones* [234]. An entanglement monotone is a quantity that cannot increase, on average, under LOCC. Such quantities measure, in this operational sense, the amount of entanglement contained in a state. For bipartite pure states, a natural entanglement monotone is the entropy of either reduced state. If we write

$$\hat{\rho}_A = \text{Tr}_B(|\psi\rangle_{AB}\langle\psi|_{AB}), \quad (2.18)$$

then the entropy of entanglement of $|\psi\rangle$, written $E(\psi)$, is defined as

$$E(\psi) := S(\hat{\rho}_A) = - \sum_{i=1}^d \lambda_\psi^{(i)} \log \lambda_\psi^{(i)}, \quad (2.19)$$

where $S(\hat{\rho}_A)$ denotes the von Neumann entropy of state $\hat{\rho}_A$, which is equal to the Shannon entropy of its eigenvalues. Equivalently, one may use $S(\hat{\rho}_B)$, since the two reduced states of a bipartite pure state have the same non-zero eigenvalues (which can be easily deduced from the Schmidt decomposition). The entropy of entanglement vanishes for product states, i.e., the states described by a peaked vector $(1, 0, \dots, 0)^T$ as Schmidt vector, and is maximal for maximally entangled states, i.e., the states described by a uniform vector $(\frac{1}{d}, \dots, \frac{1}{d})^T$ as Schmidt vector.

The direction of Nielsen's theorem is therefore consistent with the resource interpretation. If $\lambda_\psi < \lambda_\phi$, then in particular $E(\psi) \geq E(\phi)$ by Schur-concavity of the Shannon entropy (see Eq. (7.25) in Chapter 7). Hence, deterministic LOCC transformations can only go from a more entangled pure state to a less entangled one.

Obviously, there are also monotones directly associated with the majorization conditions. For a bipartite pure state $|\psi\rangle_{AB}$, define

$$E_l(\psi) := \sum_{i=1}^d \lambda_{\psi}^{(i)}, \quad l = 1, \dots, d. \quad (2.20)$$

For a deterministic LOCC transformation $|\psi\rangle_{AB} \xrightarrow{\text{LOCC}} |\phi\rangle_{AB}$, Nielsen's theorem can be reformulated as

$$E_l(\psi) \geq E_l(\phi), \quad \forall l = 1, \dots, d. \quad (2.21)$$

Thus, the quantities E_l are also entanglement monotones. They are simply another way of writing the cumulative-sum conditions appearing in the majorization relation of Eq. (2.16). In this sense, they contain more information than the entropy of entanglement alone since they form a necessary and sufficient set of d conditions for bipartite pure-state deterministic conversions, whereas $E(\psi) \geq E(\phi)$ is only a necessary condition.

However, since majorization is only a preorder relation, not all pairs of pure states are comparable. There may exist two Schmidt vectors such that neither

$$\lambda_{\psi} < \lambda_{\phi} \quad \text{nor} \quad \lambda_{\phi} < \lambda_{\psi} \quad (2.22)$$

holds. The corresponding states are said to be incomparable under deterministic LOCC. This incomparability can sometimes be overcome with the help of a catalyst [142]. Namely, a transformation that is impossible by deterministic LOCC alone may become possible if Alice and Bob use an additional entangled state $|c\rangle_{A'B'}$ that is nonetheless returned unchanged at the end of the protocol, i.e.,

$$|\psi\rangle_{AB} \otimes |c\rangle_{A'B'} \xrightarrow{\text{LOCC}} |\phi\rangle_{AB} \otimes |c\rangle_{A'B'}, \quad (2.23)$$

even though

$$|\psi\rangle_{AB} \not\xrightarrow{\text{LOCC}} |\phi\rangle_{AB}. \quad (2.24)$$

In essence, the catalyst assists the transformation without being consumed. If λ_c denotes the Schmidt vector of the catalyst, the condition becomes

$$\lambda_{\psi} \otimes \lambda_c < \lambda_{\phi} \otimes \lambda_c, \quad (2.25)$$

while

$$\lambda_{\psi} \not\leq \lambda_{\phi}. \quad (2.26)$$

Here, the tensor product of Schmidt vectors is obtained by taking all pairwise products of their components and ordering the resulting vector non-increasingly.

A standard example is

$$\lambda_\psi = (0.4, 0.4, 0.1, 0.1)^T, \quad \lambda_\phi = (0.5, 0.25, 0.25, 0)^T, \quad (2.27)$$

which are incomparable. However, with the catalyst

$$\lambda_c = (0.6, 0.4)^T, \quad (2.28)$$

one checks that

$$\lambda_\psi \otimes \lambda_c < \lambda_\phi \otimes \lambda_c. \quad (2.29)$$

Thus, the transformation from $|\psi\rangle_{AB}$ to $|\phi\rangle_{AB}$ becomes possible in the presence of the catalyst, even though it is impossible without it.

2.2.3 Probabilistic transformations of bipartite pure states

We remain in the single-copy pure-state setting, but now allow the transformation to succeed only with some probability. In fact, a transformation that is impossible deterministically may still be possible probabilistically.

We write

$$|\psi\rangle_{AB} \xrightarrow[p]{\text{LOCC}} |\phi\rangle_{AB} \quad (2.30)$$

when an LOCC protocol outputs $|\phi\rangle_{AB}$ with probability p . If the protocol fails, another residual state is produced. Such a probabilistic transformation may increase the entropy of entanglement, Eq. (2.19), in the successful branch, but not on average once all outcomes are taken into account.

For bipartite pure states, the optimal probability of conversion is given by Vidal's theorem [233]. The theorem is expressed precisely in terms of the monotones E_l defined in Eq. (2.20). Let

$$|\psi\rangle_{AB} = \sum_{i=1}^d \sqrt{\lambda_\psi^{(i)}} |i\rangle_A |i\rangle_B, \quad |\phi\rangle_{AB} = \sum_{i=1}^d \sqrt{\lambda_\phi^{(i)}} |i\rangle_A |i\rangle_B \quad (2.31)$$

be Schmidt decompositions, with Schmidt vectors λ_ψ and λ_ϕ ordered non-increasingly.

Theorem 2.2 (Vidal [233]). *Let $|\psi\rangle_{AB}$ and $|\phi\rangle_{AB}$ be two bipartite pure states. The maximal probability of converting $|\psi\rangle_{AB}$ into $|\phi\rangle_{AB}$ by LOCC is*

$$p_{\max} = \min_{l \in \{1, \dots, d\}} \frac{E_l(\psi)}{E_l(\phi)}. \quad (2.32)$$

Equivalently, the transformation is possible with probability p if and only if

$$E_l(\psi) \geq p E_l(\phi), \quad \forall l \in \{1, \dots, d\}. \quad (2.33)$$

Nielsen's theorem is recovered as the case $p_{\max} = 1$. Indeed, when $\lambda_\psi < \lambda_\phi$, one has $E_l(\psi) \geq E_l(\phi)$ for every $l = 1, \dots, d$, and Eq. (2.32) gives a deterministic conversion since $E_1(\psi) = E_1(\phi) = 1$. For $p < 1$, the inequalities in Eq. (2.33) express the fact that the monotones may increase in the successful branch, but only with a limited probability of success.

As a simple example, consider the partially entangled two-qubit state

$$|\psi\rangle_{AB} = \sqrt{a}|00\rangle + \sqrt{1-a}|11\rangle, \quad a \geq \frac{1}{2}, \quad (2.34)$$

and suppose that Alice and Bob want to obtain the maximally entangled state

$$|\Phi^+\rangle_{AB} = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (2.35)$$

The Schmidt vectors are

$$\lambda_\psi = (a, 1-a)^T, \quad \lambda_{\Phi^+} = \left(\frac{1}{2}, \frac{1}{2}\right)^T. \quad (2.36)$$

Using Eq. (2.32), we obtain

$$p_{\max} = \min\left\{1, \frac{1-a}{1/2}\right\} = 2(1-a). \quad (2.37)$$

Thus, a partially entangled two-qubit state can be converted into a maximally entangled state at most with probability $2(1-a)$. This probability vanishes when $a = 1$, as expected for a separable state, and is equal to 1 when $a = 1/2$, since in that case $|\psi\rangle_{AB} = |\Phi^+\rangle_{AB}$.

2.2.4 Asymptotic transformations

In the asymptotic regime, instead of asking if one copy of a state can be transformed into one copy of another, many copies of the same state are given to Alice and Bob. Thus, we study transformations of the form

$$|\psi\rangle_{AB}^{\otimes n} \xrightarrow{\text{LOCC}} |\phi\rangle_{AB}^{\otimes m}, \quad (2.38)$$

in the limit where n and m become large. In this setting, the relevant quantity is no longer only the possibility of a single transformation, but the rate

$$R = \frac{m}{n} \quad (2.39)$$

at which copies of $|\phi\rangle_{AB}$ can be produced from copies of $|\psi\rangle_{AB}$, up to an error that vanishes in the limit $n \rightarrow \infty$.

A central task in this regime is the so-called *entanglement distillation* [19]. Starting from many copies of an entangled state, Alice and Bob try to extract maximally entangled states by LOCC. If $|\Phi_2\rangle_{AB}$ denotes a maximally entangled two-qubit state, one considers transformations of the form

$$|\psi\rangle_{AB}^{\otimes n} \xrightarrow{\text{LOCC}} |\Phi_2\rangle_{AB}^{\otimes m}. \quad (2.40)$$

The maximal asymptotic value of m/n is called the distillable entanglement, which we denote by $E_D(\psi)$. Since the target state is a two-qubit maximally entangled state $|\Phi_2\rangle_{AB}$ (also called a Bell state), this quantity is measured in ebits, that is, in Bell states extracted per copy of $|\psi\rangle_{AB}$.

The reverse task is the so-called *entanglement dilution* [19]. In this case, Alice and Bob start from maximally entangled states and use LOCC to prepare many copies of a target state $|\psi\rangle_{AB}$, i.e.,

$$|\Phi_2\rangle_{AB}^{\otimes m} \xrightarrow{\text{LOCC}} |\psi\rangle_{AB}^{\otimes n}. \quad (2.41)$$

The minimal asymptotic value of m/n required for this preparation is called the entanglement cost, which we denote by $E_C(\psi)$. It measures how many ebits are needed, per copy, to create the state $|\psi\rangle_{AB}$ asymptotically.

For bipartite pure states, distillation and dilution are asymptotically reversible. Considering base-2 logarithms, we have

$$E_D(\psi) = E_C(\psi) = E(\psi) := -\sum_i \lambda_\psi^{(i)} \log \lambda_\psi^{(i)}, \quad (2.42)$$

where λ_ψ is the Schmidt vector of state $|\psi\rangle$ and E is the entropy of entanglement defined in Eq. (2.19). Thus, in the asymptotic regime, bipartite pure-state entanglement becomes reversible, i.e., many copies of $|\psi\rangle_{AB}$ can be converted into Bell pairs at rate $E(\psi)$, and the same number of ebits per copy is sufficient to reconstruct $|\psi\rangle_{AB}$ by dilution.

Combining distillation and dilution also gives the optimal asymptotic conversion rate between two bipartite pure states as

$$R(\psi \rightarrow \phi) = \frac{E(\psi)}{E(\phi)}. \quad (2.43)$$

We assume $E(\phi) > 0$, otherwise $|\phi\rangle_{AB}$ is a product state, which can be prepared locally any number of times. When the target state is a Bell state, $E(\Phi_2) = 1$, so the rate reduces to $E(\psi)$. This is in contrast with the single-copy setting, where the full majorization relation between Schmidt vectors is required, see Theorem 2.1. In the asymptotic limit, the entropy of entanglement alone determines the reversible conversion rate between bipartite pure states.

For mixed states, the same operational tasks can be defined, but the situation is more involved. The distillable entanglement $E_D(\hat{\rho}_{AB})$ is the optimal rate at which Bell pairs can be extracted from many copies of $\hat{\rho}_{AB}$ by LOCC, while the entanglement cost $E_C(\hat{\rho}_{AB})$ is the optimal rate of Bell pairs needed to create many copies of $\hat{\rho}_{AB}$. In general, these two quantities do not coincide [128], and we have

$$E_D(\hat{\rho}_{AB}) \leq E_C(\hat{\rho}_{AB}). \quad (2.44)$$

Thus, the asymptotic reversibility of entanglement is a special property of bipartite pure states, and does not hold for mixed states in general.

2.3 Detecting discrete-variable entanglement

Let us now consider the problem of certifying entanglement in finite-dimensional systems. In the case of bipartite pure states, the criterion is quite simple and was already established after defining the Schmidt decomposition in Proposition 2.2. It reads as follows: a pure state is separable if and only if its Schmidt rank, i.e., its number of non-zero Schmidt coefficients, is equal to one. Equivalently, it is entangled if and only if at least two Schmidt coefficients are non-zero.

For mixed states, the situation is much more involved. Given a bipartite state $\hat{\rho}_{AB}$, determining whether there exists a decomposition of the form of Eq. (2.4), that is, a separable mixture of $\hat{\rho}_{AB}$, is known to be an NP-hard problem [96, 111]. One therefore relies on separability criteria, sometimes also called entanglement criteria, which can certify entanglement from specific properties of the state under investigation. Many criteria have been developed, and they are usually sufficient conditions for entanglement rather than complete characterizations of the whole set of entangled states.

In what follows, we present some of the standard criteria used in finite-dimensional systems. We first discuss spectral criteria based on majorization. Then, we present two standard criteria for entanglement detection: the positive partial transpose (PPT) criterion and the realignment criterion.

2.3.1 Majorization-based criteria

The majorization-based approach is motivated by the fact that separable states are more disordered globally than locally [173], where disorder is understood through the majorization relation. The notion of majorization can be extended to density matrices through their spectra.

We write $\hat{\rho} < \hat{\sigma}$ whenever

$$\lambda(\hat{\rho}) < \lambda(\hat{\sigma}), \quad (2.45)$$

where $\lambda(\bullet)$ denotes the vector of eigenvalues of the corresponding state $\bullet$, arranged in non-increasing order. When the two density matrices act on Hilbert spaces of different dimensions, the shorter spectrum is completed with zeros.

Theorem 2.3 (Nielsen and Kempe [173]). *Let $\hat{\rho}_{AB}$ be a bipartite state, and denote by $\hat{\rho}_A$ and $\hat{\rho}_B$ its reduced states. If $\hat{\rho}_{AB}$ is separable, then*

$$\hat{\rho}_{AB} < \hat{\rho}_A \quad \text{and} \quad \hat{\rho}_{AB} < \hat{\rho}_B. \quad (2.46)$$

Equivalently, if $\hat{\rho}_{AB} \not\prec \hat{\rho}_A$ or $\hat{\rho}_{AB} \not\prec \hat{\rho}_B$, then $\hat{\rho}_{AB}$ is entangled. Notice that, for pure states, this criterion becomes necessary and sufficient for entanglement detection. Indeed, the global spectrum is then the peaked vector $(1, 0, \dots, 0)^T$, which is majorized by a reduced spectrum only if the latter is also pure. Hence, for pure states, Theorem 2.3 is satisfied only for product pure states, and is violated for all entangled pure states.

Soon after, Hiroshima showed the following stronger statement.

Theorem 2.4 (Hiroshima [123]). *Let $\hat{\rho}_{AB}$ be a bipartite state, and denote by $\hat{\rho}_A$ and $\hat{\rho}_B$ its reduced states. If $\hat{\rho}_{AB}$ is not distillable, then*

$$\hat{\rho}_{AB} < \hat{\rho}_A \quad \text{and} \quad \hat{\rho}_{AB} < \hat{\rho}_B. \quad (2.47)$$

Since separable states are not distillable², this result strengthens the Nielsen–Kempe criterion. In particular, if $\hat{\rho}_{AB} \not\prec \hat{\rho}_A$ or $\hat{\rho}_{AB} \not\prec \hat{\rho}_B$, then the state is not only entangled, but also distillable (and we will see that not all entangled states are distillable in the next subsection).

2.3.2 Positive-partial-transpose criterion

One of the first separability criteria is rather simple, and is based on the partial transposition map [129, 178]. Let us fix a product basis $\{|i\rangle_A \otimes |j\rangle_B\}$ of the global Hilbert space $\mathcal{H}_{AB}$. The partial transpose with respect to Bob's subsystem T_B is defined by

$$(|i\rangle \langle k| \otimes |j\rangle \langle l|)^{T_B} = |i\rangle \langle k| \otimes |l\rangle \langle j|. \quad (2.48)$$

²The notion of distillability is closely related to the entanglement distillation discussed in Section 2.2.4. A state is said to be distillable if many copies of it can be used to distill Bell states by LOCC, in the asymptotic limit. A separable state is obviously not distillable since entanglement can never be created from it using LOCC, even in the limit of infinite copies.

Thus, it acts as the identity on Alice's subsystem and as the transposition map on Bob's subsystem. Equivalently, one may apply the partial transposition on Alice's subsystem, since the two choices lead to the same positivity criterion. The Peres–Horodecki positive-partial-transpose (PPT) criterion can then be stated as follows (we write $\hat{A} \geq 0$ to indicate that $\hat{A}$ is positive semidefinite).

Theorem 2.5 (PPT criterion [129, 178]). *Let $\hat{\rho}_{AB}$ be a bipartite state. If $\hat{\rho}_{AB}$ is separable, then*

$$\hat{\rho}_{AB}^{T_B} \geq 0. \quad (2.49)$$

Proving this theorem is quite straightforward from Definition 2.2 of separable states, and the fact that the transposition map is a positive map. Equivalently, if $\hat{\rho}_{AB}^{T_B}$ is not positive semidefinite, i.e., if it has at least one negative eigenvalue, then $\hat{\rho}_{AB}$ is entangled. Such states are called NPT states, meaning that their partial transpose is not positive semidefinite. In contrast, states satisfying $\hat{\rho}_{AB}^{T_B} \geq 0$ are called PPT states.

The PPT criterion is necessary for separability in any dimension. Moreover, it is necessary and sufficient in dimensions 2×2 and 2×3 [129]. In higher dimensions, however, there exist entangled states with positive partial transpose. These states are called PPT entangled states. Such states exhibit a weaker form of entanglement in the sense that they are not distillable by LOCC [127]. For this reason, they are also called bound entangled states.

As for the majorization criterion, the PPT criterion becomes necessary and sufficient for pure-state entanglement. Indeed, let

$$|\psi\rangle_{AB} = \sum_i \sqrt{\lambda_i} |i\rangle_A |i\rangle_B \quad (2.50)$$

be the Schmidt decomposition of a bipartite pure state. The partial transpose of $|\psi\rangle_{AB}$ has eigenvalues λ_i and, for each pair $i < j$, the two eigenvalues $\pm \sqrt{\lambda_i \lambda_j}$. Hence, it is positive semidefinite if and only if at most one Schmidt coefficient is non-zero. Therefore, for pure states, the PPT criterion detects exactly all entangled states.

2.3.3 Realignment criterion

Another standard separability criterion is the realignment criterion, also known as the computable cross norm criterion [192, 194]. It is based on a reshuffling of the matrix elements of the density operator. Let us fix a product basis $\{|i\rangle_A \otimes |j\rangle_B\}$ of the global Hilbert space $\mathcal{H}_{AB}$. The realignment map R is defined on product matrix elements by

$$R(|i\rangle\langle k| \otimes |j\rangle\langle l|) = |i\rangle\langle j| \otimes |k\rangle\langle l|. \quad (2.51)$$

Upon defining the trace norm of an operator $\hat{X}$ as

$$\|\hat{X}\|_1 := \text{Tr} \left[\sqrt{\hat{X}^\dagger \hat{X}} \right], \quad (2.52)$$

the realignment criterion can be stated as follows.

Theorem 2.6 (Realignment criterion [192]). *Let $\hat{\rho}_{AB}$ be a bipartite state. If $\hat{\rho}_{AB}$ is separable, then*

$$\|R(\hat{\rho}_{AB})\|_1 \leq 1. \quad (2.53)$$

Equivalently, if $\|R(\hat{\rho}_{AB})\|_1 > 1$, then $\hat{\rho}_{AB}$ is entangled. This criterion is in some sense complementary to the PPT criterion, since it can detect some PPT entangled states.

For pure states, the criterion is again necessary and sufficient. Indeed, if

$$|\psi\rangle_{AB} = \sum_i \sqrt{\lambda_i} |i\rangle_A |i\rangle_B \quad (2.54)$$

is the Schmidt decomposition of $|\psi\rangle_{AB}$, then

$$\|R(|\psi\rangle_{AB} \langle\psi|_{AB})\|_1 = \left(\sum_i \sqrt{\lambda_i} \right)^2. \quad (2.55)$$

This quantity is equal to 1 if and only if only one Schmidt coefficient is non-zero, and is strictly greater than 1 if there is more than one non-zero Schmidt coefficient.

2.4 Detecting continuous-variable entanglement

Continuous-variable systems are ubiquitous in quantum optics (see Chapter 3 for a more detailed overview), where a quantum state may be encoded in continuous degrees of freedom such as the canonical quadratures $\hat{x}$ and $\hat{p}$, which obey the canonical commutation relation

$$[\hat{x}, \hat{p}] = i\hbar \hat{1}. \quad (2.56)$$

As we will see in this section, Gaussian states play a central role in continuous-variable quantum information [243]. Their importance comes from the fact that they are completely characterized by the first and second moments of the canonical quadratures. Moreover, as first moments can be modified by local displacements, they do not affect entanglement properties. Hence, for Gaussian states, many separability criteria can be expressed directly in terms of the covariance matrix.

We first introduce some basic notions of quantum optics that will be useful in the rest of the section. Then, we discuss standard continuous-variable separability criteria based on moments up to order two, before turning to criteria based on entropic quantities.

2.4.1 A glimpse at quantum optics

This section relies primarily on Ref. [243]. For simplicity of notation, we work in natural units, i.e., $\hbar = 1$, so that

$$[\hat{x}, \hat{p}] = i\hat{1}. \quad (2.57)$$

A quantum state $\hat{\rho}$ can be represented in phase space by its Wigner function. For an N -mode state, it is defined as

$$W(\mathbf{x}, \mathbf{p}) := \frac{1}{(2\pi)^N} \int_{\mathbb{R}^N} d^N \mathbf{y} e^{-i\mathbf{p} \cdot \mathbf{y}} \left\langle \mathbf{x} + \frac{\mathbf{y}}{2} \left| \hat{\rho} \right| \mathbf{x} - \frac{\mathbf{y}}{2} \right\rangle, \quad (2.58)$$

where $\mathbf{x} = (x_1, \dots, x_N)^T$ and $\mathbf{p} = (p_1, \dots, p_N)^T$ are the vectors of phase-space coordinates associated with the quadratures $\hat{x}_i$ and $\hat{p}_i$. The Wigner function is a quasi-probability distribution, i.e., it is real-valued (but it may take negative values!), normalized,

$$\int_{\mathbb{R}^N} d^N \mathbf{x} \int_{\mathbb{R}^N} d^N \mathbf{p} W(\mathbf{x}, \mathbf{p}) = 1, \quad (2.59)$$

and its marginals reproduce the position and momentum probability distributions,

$$\int_{\mathbb{R}^N} d^N \mathbf{p} W(\mathbf{x}, \mathbf{p}) = \langle \mathbf{x} | \hat{\rho} | \mathbf{x} \rangle, \quad \int_{\mathbb{R}^N} d^N \mathbf{x} W(\mathbf{x}, \mathbf{p}) = \langle \mathbf{p} | \hat{\rho} | \mathbf{p} \rangle. \quad (2.60)$$

The Wigner representation is useful in that it translates many questions about continuous-variable quantum states into questions about functions on phase space. In particular, Gaussian states admit a particularly simple description in phase space, since they are completely characterized by a finite number of parameters. More precisely, a Gaussian state is a state whose Wigner function is Gaussian. Its complete information is contained in its first and second moments. Upon defining the vector of quadrature operators as

$$\hat{\mathbf{r}} := (\hat{x}_1, \hat{p}_1, \dots, \hat{x}_N, \hat{p}_N)^T, \quad (2.61)$$

the displacement vector and covariance matrix are then defined by

$$\mathbf{d} := \langle \hat{\mathbf{r}} \rangle = \text{Tr} [\hat{\rho} \hat{\mathbf{r}}], \quad (2.62)$$

and

$$\gamma_{ij} := \frac{1}{2} \langle \{\hat{r}_i, \hat{r}_j\} \rangle - \langle \hat{r}_i \rangle \langle \hat{r}_j \rangle, \quad (2.63)$$

where $\{\cdot, \cdot\}$ denotes the anti-commutator.

With the above ordering, the canonical commutation relations for an N -mode system (see Eq. (2.57) for the 1-mode case) can be written

$$[\hat{r}_k, \hat{r}_l] = i\Omega_{kl}\hat{\mathbb{I}}, \quad (2.64)$$

where the symplectic form Ω depends on the ordering choice of Eq. (2.61), and is expressed as

$$\Omega := \bigoplus_{k=1}^N \omega, \quad \omega := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad (2.65)$$

where $\oplus$ denotes the direct sum. The uncertainty relation can be expressed as follows in phase space, and constitutes an important physicality condition that covariance matrices of states have to obey,

$$\gamma + \frac{i}{2}\Omega \geq 0, \quad (2.66)$$

where the inequality ≥ 0 means that the matrix is positive semidefinite. Let us denote by $\mathbf{r} := (x_1, p_1, \dots, x_N, p_N)^T$, the real phase-space vector associated with the operator vector $\hat{\mathbf{r}}$ defined in Eq. (2.61). For a Gaussian state, the Wigner function takes the form

$$W_G(\mathbf{r}) = \frac{1}{(2\pi)^N \sqrt{\det \gamma}} e^{-\frac{1}{2}(\mathbf{r} - \langle \hat{\mathbf{r}} \rangle)^T \gamma^{-1} (\mathbf{r} - \langle \hat{\mathbf{r}} \rangle)}. \quad (2.67)$$

Using the overlap formula for Wigner functions, one obtains the following formula for the purity of a Gaussian state,

$$\text{Tr}[\hat{\rho}^2] = \frac{1}{2^N \sqrt{\det \gamma}}. \quad (2.68)$$

Thus, for Gaussian states, quantities such as the purity, and several separability criteria, as we will see in the next subsection, can be expressed directly in terms of the covariance matrix γ .

2.4.2 Second-moment-based criteria

We now discuss some seminal separability criteria based only on moments up to order two. Such criteria are especially important in continuous-variable systems, because first and second moments are often experimentally accessible through homodyne measurements [28]. Moreover, for Gaussian states, they are not merely partial information because, together with the displacement vector, the covariance matrix completely characterizes the state.

The criteria presented hereafter are due to Duan et al. [72], Simon [216], and Mancini et al. [99, 160]. They are closely related to the continuous-variable form of the PPT separability criterion (see Theorem 2.5), but expressed directly in terms of quadrature fluctuations.

In phase space, partial transposition has a simple interpretation since it corresponds to a reflection of the momentum quadrature of the transposed subsystem [216]. For a two-mode state, this reads

$$W(x_A, p_A, x_B, p_B) \xrightarrow{T_B} W(x_A, p_A, x_B, -p_B) =: W(x_A, p_A, x_B, p_B)^{T_B}. \quad (2.69)$$

Thus, at the covariance-matrix level, partial transposition amounts to changing the sign of the correlations involving p_B . We first present the criterion of Duan et al. Let us define the following operators,

$$\hat{u} := a\hat{x}_A + \frac{1}{a}\hat{x}_B, \quad \hat{v} := a\hat{p}_A - \frac{1}{a}\hat{p}_B, \quad (2.70)$$

where $a > 0$ is a real parameter.

Theorem 2.7 (Duan et al. [72]). *For any separable state $\hat{\rho}_{AB}$, and for any $a > 0$, one has*

$$\sigma_u^2 + \sigma_v^2 \geq a^2 + \frac{1}{a^2}, \quad (2.71)$$

where σ_u^2 and σ_v^2 are the variances of $\hat{u}$ and $\hat{v}$, respectively.

Hence, a violation of Eq. (2.71) certifies entanglement. In practice, one may optimize over the parameter a in order to obtain the strongest test for a given state.

Mancini et al. later proposed a (refined) product form of this criterion.

Theorem 2.8 (Mancini et al. [99, 160]). *For any separable state $\hat{\rho}_{AB}$, and for any $a > 0$, one has*

$$\sigma_u^2 \sigma_v^2 \geq \frac{1}{4} \left(a^2 + \frac{1}{a^2} \right)^2, \quad (2.72)$$

where σ_u^2 and σ_v^2 are the variances of $\hat{u}$ and $\hat{v}$, respectively.

This condition is stronger than Eq. (2.71) for fixed a . Indeed, if Eq. (2.72) holds, then

$$\sigma_u \sigma_v \geq \frac{1}{2} \left(a^2 + \frac{1}{a^2} \right), \quad (2.73)$$

and therefore, using the obvious inequality,

$$\sigma_u^2 + \sigma_v^2 \geq 2\sigma_u \sigma_v, \quad (2.74)$$

we recover Eq. (2.71).

Finally, Simon formulated the PPT criterion directly at the level of the covariance matrix.

For a two-mode state, partial transposition is implemented by

$$\gamma \xrightarrow{T_B} \Lambda_B \gamma \Lambda_B =: \gamma^{T_B}, \quad \Lambda_B := \text{diag}(1, 1, 1, -1), \quad (2.75)$$

which corresponds to the transformation $p_B \mapsto -p_B$. Notice that γ^{T_B} is a notation and should not be understood as a partial transposition applied directly on the covariance matrix, since it is an operation applied on the quantum state.

Theorem 2.9 (Simon [216]). *For any separable state $\hat{\rho}_{AB}$, the covariance matrix of the partially transposed state satisfies*

$$\gamma^{T_B} + \frac{i}{2}\Omega \geq 0. \quad (2.76)$$

Therefore, if $\gamma^{T_B} + i\Omega/2$ is not positive semidefinite, then the state is entangled.

For arbitrary continuous-variable states, the above criteria are only sufficient criteria for entanglement, i.e., their violation certifies entanglement, but their satisfaction does not imply separability in general. This is because these criteria only use second moments, whereas a non-Gaussian state may contain entanglement information encoded in higher-order correlations.

However, the situation is much simpler for Gaussian states. Since Gaussian states are completely determined by their covariance matrix and displacement vector, second-moment criteria can become complete. In particular, for two-mode Gaussian states, Simon's covariance-matrix PPT criterion is necessary and sufficient for separability. In the same setting, the optimized Duan et al. criterion and the Mancini et al. criterion are also necessary and sufficient. More generally, Simon's PPT condition remains necessary and sufficient for Gaussian states of a $1 \times N$ bipartition [244]. For multimode Gaussian states of an $M \times N$ bipartition with $M, N \geq 2$, however, PPT is no longer sufficient in general, since PPT entangled Gaussian states exist. Special classes of multimode Gaussian states, such as bisymmetric states, nevertheless admit stronger necessary and sufficient criteria [39, 205].

2.4.3 Entropy-based criterion

We close this chapter by briefly discussing another family of continuous-variable separability criteria, based on entropic quantities rather than only on second moments. We present here the criterion introduced by Walborn et al. [239].

Let f be a probability density on $\mathcal{X}$. Its Shannon differential entropy is defined as

$$h(f) := - \int_{\mathcal{X}} dx f(x) \log f(x), \quad (2.77)$$

where we use base-2 logarithms.

Let us introduce the rotated quadratures

$$\hat{r}_j = \cos \theta_j \hat{x}_j + \sin \theta_j \hat{p}_j, \quad \hat{s}_j = \cos \theta_j \hat{p}_j - \sin \theta_j \hat{x}_j, \quad (2.78)$$

for $j = 1, 2$. From Eq. (2.57), it is immediate to see that they satisfy

$$[\hat{r}_j, \hat{s}_{j'}] = i\delta_{jj'}\hat{1}. \quad (2.79)$$

We then define the EPR-like global quadratures

$$\hat{r}_\pm := \hat{r}_1 \pm \hat{r}_2, \quad \hat{s}_\pm := \hat{s}_1 \pm \hat{s}_2. \quad (2.80)$$

Notice that $\hat{r}_\pm$ commutes with $\hat{s}_\mp$, so that the observables are compatible and may be measured jointly. Let $R_\pm$ denote the probability density associated with a measurement of $\hat{r}_\pm$, and let $S_\mp$ denote the probability density associated with a measurement of $\hat{s}_\mp$. The entropic separability criterion can then be stated as follows.

Theorem 2.10 (Walborn et al. [239]). *For any separable bipartite continuous-variable state, one has*

$$h(R_\pm) + h(S_\mp) \geq \log(2\pi e). \quad (2.81)$$

Equivalently, if $h(R_\pm) + h(S_\mp) < \log(2\pi e)$, then the state is entangled. This criterion can be seen as an entropic analogue of the variance-based criteria discussed above. Instead of quantifying the spread of the quadratures only through their variances, it uses the Shannon differential entropy of their full probability distributions. It can therefore be more sensitive for non-Gaussian states, where second moments may fail in general to capture all relevant correlations.

For Gaussian probability densities, the differential entropy is completely determined by the variance [49]. More precisely, if f_G is a Gaussian probability density with variance σ_G^2 , then

$$h(f_G) = \frac{1}{2} \log(2\pi e \sigma_G^2). \quad (2.82)$$

Therefore, in the Gaussian case, Eq. (2.81) becomes

$$\sigma_{r_\pm}^2 \sigma_{s_\mp}^2 \geq 1, \quad (2.83)$$

which is precisely a condition of the same form as the Mancini et al. criterion (see Eq. (2.72)).

In the remainder of this thesis, Theorems 2.1 and 2.2 will be central to the study of bipartite pure-state entanglement conversions within the majorization lattice framework in Chapter 8. Theorem 2.5 will support the partially transposed moment-based criteria introduced in Chapter 5 for detecting entanglement in continuous-variable states, while Theorem 2.6 will be crucial to the development of other experimentally accessible criteria for continuous-variable systems in Chapter 6.

3 Quantum optics

Since the dawn of quantum theory, light has no longer been viewed solely as a classical electromagnetic wave, but also as a quantum object [4, 240]. In quantum mechanics, the quanta of light are called photons and can be understood as excitations of the electromagnetic field. They play an important role in many quantum-information protocols, where they can be used to encode, transmit, and process information [28, 243].

In this chapter, we review the quantization of the free electromagnetic field in Section 3.1. We then present the main quantum-optical states in Section 3.2, including both Gaussian and non-Gaussian states. Finally, in Section 3.3, we discuss the corresponding operations, together with the most important measurements used in quantum optics.

3.1 Quantizing the electromagnetic field

This section follows the approach of Ref. [4]¹. Consider a plane wave of the electromagnetic field, with frequency ω_k and wave vector $\mathbf{k}$, propagating in free space. The electric and magnetic fields can be written as

$$\mathbf{E}(\mathbf{r}, t) = \epsilon E_0 e^{i\mathbf{k} \cdot \mathbf{r} - i\omega_k t} + \text{c.c.}, \quad \mathbf{B}(\mathbf{r}, t) = \frac{\mathbf{k} \times \epsilon}{\omega_k} E_0 e^{i\mathbf{k} \cdot \mathbf{r} - i\omega_k t} + \text{c.c.}, \quad (3.1)$$

where $\omega_k = c|\mathbf{k}|$, c is the speed of light in vacuum, and ϵ is a unit transverse polarization vector satisfying $\mathbf{k} \cdot \epsilon = 0$. The quantity E_0 is the complex amplitude of the electric field and has units of electric field. The corresponding magnetic-field amplitude is fixed by Maxwell's equations and is given by

$$\mathbf{B}_0 = \frac{\mathbf{k} \times \epsilon}{\omega_k} E_0. \quad (3.2)$$

For a unit transverse polarization vector, this gives $|\mathbf{B}_0| = |E_0|/c$.

We adopt the Coulomb gauge, and set the scalar potential to zero. The electric and magnetic fields are related to the vector potential by

$$\mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{A}(\mathbf{r}, t)}{\partial t}, \quad \mathbf{B}(\mathbf{r}, t) = \nabla \times \mathbf{A}(\mathbf{r}, t). \quad (3.3)$$

¹Importantly, we have converted the CGS units used by the author into SI units.

The vector potential can then be expanded over transverse modes as

$$\mathbf{A}(\mathbf{r}, t) = \sum_{\mathbf{k}, s} \frac{A_{\mathbf{k}, s}}{\sqrt{V}} \boldsymbol{\epsilon}_{\mathbf{k}, s} e^{i\mathbf{k} \cdot \mathbf{r} - i\omega_k t} + \text{c.c.}, \quad (3.4)$$

where V is a finite quantization volume, introduced to discretize the modes. The index $s = 1, 2$ labels a choice of two independent transverse polarization vectors associated with each wave vector $\mathbf{k}$, which satisfy $\mathbf{k} \cdot \boldsymbol{\epsilon}_{\mathbf{k}, s} = 0$, and the summation goes over all allowed values of $\mathbf{k}$.

Upon quantization, the mode amplitudes are promoted to operators. We introduce the annihilation and creation operators $\hat{a}_{\mathbf{k}, s}$ and $\hat{a}_{\mathbf{k}, s}^\dagger$, satisfying

$$[\hat{a}_{\mathbf{k}, s}, \hat{a}_{\mathbf{k}', s'}^\dagger] = \delta_{\mathbf{k}, \mathbf{k}'} \delta_{s, s'} \hat{1}, \quad [\hat{a}_{\mathbf{k}, s}, \hat{a}_{\mathbf{k}', s'}] = [\hat{a}_{\mathbf{k}, s}^\dagger, \hat{a}_{\mathbf{k}', s'}^\dagger] = 0. \quad (3.5)$$

The quantized vector potential is then

$$\hat{\mathbf{A}}(\mathbf{r}, t) = \sum_{\mathbf{k}, s} \sqrt{\frac{\hbar}{2\epsilon_0 \omega_k V}} \left(\boldsymbol{\epsilon}_{\mathbf{k}, s} \hat{a}_{\mathbf{k}, s} e^{i\mathbf{k} \cdot \mathbf{r} - i\omega_k t} + \boldsymbol{\epsilon}_{\mathbf{k}, s}^* \hat{a}_{\mathbf{k}, s}^\dagger e^{-i\mathbf{k} \cdot \mathbf{r} + i\omega_k t} \right). \quad (3.6)$$

Consequently, using Eq. (3.3),

$$\begin{aligned} \hat{\mathbf{E}}(\mathbf{r}, t) &= i \sum_{\mathbf{k}, s} \sqrt{\frac{\hbar \omega_k}{2\epsilon_0 V}} \left(\boldsymbol{\epsilon}_{\mathbf{k}, s} \hat{a}_{\mathbf{k}, s} e^{i\mathbf{k} \cdot \mathbf{r} - i\omega_k t} - \boldsymbol{\epsilon}_{\mathbf{k}, s}^* \hat{a}_{\mathbf{k}, s}^\dagger e^{-i\mathbf{k} \cdot \mathbf{r} + i\omega_k t} \right), \\ \hat{\mathbf{B}}(\mathbf{r}, t) &= i \sum_{\mathbf{k}, s} \sqrt{\frac{\hbar}{2\epsilon_0 \omega_k V}} \left((\mathbf{k} \times \boldsymbol{\epsilon}_{\mathbf{k}, s}) \hat{a}_{\mathbf{k}, s} e^{i\mathbf{k} \cdot \mathbf{r} - i\omega_k t} - (\mathbf{k} \times \boldsymbol{\epsilon}_{\mathbf{k}, s}^*) \hat{a}_{\mathbf{k}, s}^\dagger e^{-i\mathbf{k} \cdot \mathbf{r} + i\omega_k t} \right). \end{aligned} \quad (3.7)$$

The Hamiltonian of the free electromagnetic field is its total energy,

$$\hat{H} = \frac{1}{2} \int d^3 \mathbf{r} \left(\epsilon_0 \hat{\mathbf{E}}^2 + \frac{1}{\mu_0} \hat{\mathbf{B}}^2 \right), \quad (3.8)$$

which becomes

$$\hat{H} = \sum_{\mathbf{k}, s} \hbar \omega_k \left(\hat{a}_{\mathbf{k}, s}^\dagger \hat{a}_{\mathbf{k}, s} + \frac{\hat{1}}{2} \right). \quad (3.9)$$

Thus, each transverse mode of the electromagnetic field is equivalent to a quantum harmonic oscillator. In the rest of this thesis, we choose natural units, so that $\hbar = 1$.

Let us now drop the indices $\mathbf{k}, s$ and consider a single mode of the electromagnetic field. The corresponding annihilation and creation operators, $\hat{a}$ and $\hat{a}^\dagger$, satisfy the commutation relation

$$[\hat{a}, \hat{a}^\dagger] = \hat{1}. \quad (3.10)$$

We can write the number operator as

$$\hat{n} = \hat{a}^\dagger \hat{a}. \quad (3.11)$$

Its eigenstates $|n\rangle$ satisfy

$$\hat{n}|n\rangle = n|n\rangle, \quad (3.12)$$

and form a discrete basis called the Fock basis, whose elements are called Fock states. In this single-mode description, a photon corresponds to one quantum of excitation of the field mode. Hence, the Fock state $|n\rangle$ contains exactly n photons. The annihilation and creation operators act on number states as

$$\hat{a}|n\rangle = \sqrt{n}|n-1\rangle, \quad \hat{a}^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle, \quad (3.13)$$

with $\hat{a}|0\rangle = 0$. The (dimensionless) position and momentum field quadratures introduced in Section 2.4 are related to the annihilation and creation operators by

$$\hat{x} = \frac{\hat{a} + \hat{a}^\dagger}{\sqrt{2}}, \quad \hat{p} = \frac{\hat{a} - \hat{a}^\dagger}{i\sqrt{2}}, \quad (3.14)$$

so that

$$[\hat{x}, \hat{p}] = i\hat{1}, \quad \hat{a} = \frac{\hat{x} + i\hat{p}}{\sqrt{2}}, \quad \hat{a}^\dagger = \frac{\hat{x} - i\hat{p}}{\sqrt{2}}. \quad (3.15)$$

Their eigenstates $|x\rangle$ and $|p\rangle$ obey

$$\hat{x}|x\rangle = x|x\rangle, \quad \hat{p}|p\rangle = p|p\rangle. \quad (3.16)$$

These are generalized, non-normalizable eigenstates and therefore do not represent physical states with finite energy. When dealing with $N \in \mathbb{N}_{>0}$ modes, we write the annihilation and creation operators as $\hat{a}_j$ and $\hat{a}_j^\dagger$, with $j = 1, \dots, N$. They satisfy the canonical commutation relations

$$[\hat{a}_j, \hat{a}_k^\dagger] = \delta_{jk}\hat{1}, \quad [\hat{a}_j, \hat{a}_k] = [\hat{a}_j^\dagger, \hat{a}_k^\dagger] = 0, \quad j, k = 1, \dots, N. \quad (3.17)$$

Equivalently, introducing the quadrature vector

$$\hat{\mathbf{r}} := (\hat{x}_1, \hat{p}_1, \dots, \hat{x}_N, \hat{p}_N)^T, \quad (3.18)$$

the quadrature commutation relations can be written compactly as

$$[\hat{r}_j, \hat{r}_k] = i\Omega_{jk}\hat{1}, \quad j, k = 1, \dots, 2N, \quad (3.19)$$

where Ω is the N -mode symplectic form defined by

$$\Omega := \bigoplus_{j=1}^N \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \quad (3.20)$$

3.2 Quantum-optical states

We present here the main families of quantum states encountered in quantum optics. Since many of them are naturally decomposed in the number basis, also called the Fock basis, we first recall this basis and its physical interpretation. We then distinguish between Gaussian and non-Gaussian states, which play different roles in continuous-variable quantum optics.

3.2.1 Number states and the Fock basis

As discussed around Eq. (3.12), the number states (or Fock states) $|n\rangle$, with $n \in \mathbb{N}_{\geq 0}$, form an orthonormal basis of the Hilbert space of a single mode. The Fock state $|0\rangle$ is often referred to as the vacuum state. Using the action of the creation operator $\hat{a}^\dagger$ on number states, see Eq. (3.13), any Fock state can be obtained from the vacuum as [4]

$$|n\rangle = \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}} |0\rangle. \quad (3.21)$$

Fock states have a zero displacement vector (see Eq. (2.62) for its definition), $\mathbf{d}_n = \mathbf{0}$. They are also invariant under phase-space rotations up to a global phase, since $e^{-i\theta\hat{n}}|n\rangle = e^{-i\theta n}|n\rangle$.

To build intuition, we give the Wigner function of the Fock state $|n\rangle$, recalling the definition of the Wigner function from Eq. (2.58) in the previous chapter. It is given by [4]²

$$W_n(x, p) = \frac{(-1)^n}{\pi} e^{-(x^2+p^2)} L_n(2(x^2+p^2)), \quad (3.22)$$

where L_n is the Laguerre polynomial of order n . This expression depends only on $x^2 + p^2$, which reflects the rotational symmetry of Fock states in phase space. For $n = 0$, the Wigner function reduces to

$$W_0(x, p) = \frac{1}{\pi} e^{-(x^2+p^2)}, \quad (3.23)$$

which is Gaussian. In contrast, all Fock states with $n \geq 1$ are non-Gaussian, and their Wigner functions have negative regions in phase space. Figure 3.1 displays the Wigner functions of the vacuum state $|0\rangle$, the Fock state $|1\rangle$ containing one photon, and the Fock state $|4\rangle$ containing four photons.

²Notice that in Ref. [4], the author is always considering $W(\alpha)$, which we translate into $W(x, p)$ by taking into account $\alpha = (x + ip)/\sqrt{2}$.

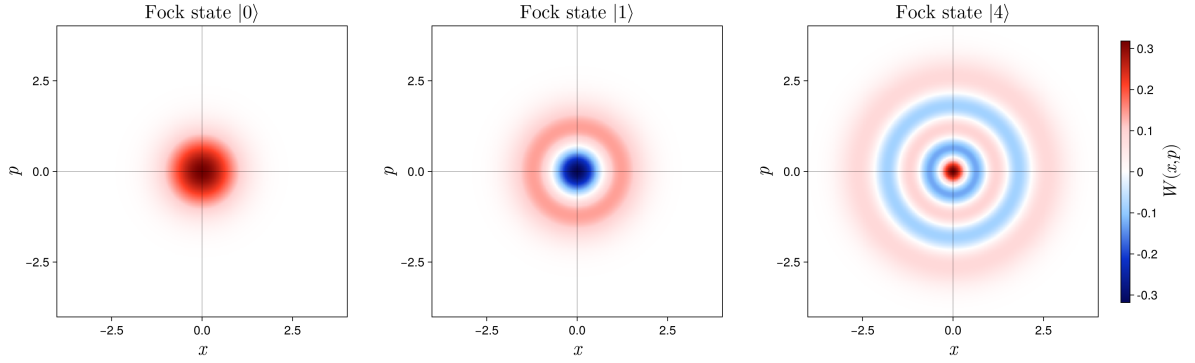


Figure 3.1: Wigner functions in the phase-space plane (x, p) for the Fock states $|0\rangle$, $|1\rangle$, and $|4\rangle$. Blue regions correspond to negative values of the Wigner function, while red regions correspond to positive values. Thus, we observe that the vacuum state $|0\rangle$ has a non-negative (Gaussian) Wigner function, while the Wigner functions of states $|1\rangle$ and $|4\rangle$ present negativities, preventing them from being Gaussian.

3.2.2 Gaussian states

Many important quantum states in quantum optics are Gaussian [243], therefore we discuss some of them here. A Gaussian state admits a Gaussian Wigner function, in the form of Eq. (2.67). Such a state is completely characterized by its first and second moments, namely its displacement vector, defined in Eq. (2.62), and its covariance matrix, defined in Eq. (2.63). We recall these expressions in the single-mode case now.

For a single-mode Gaussian state, with $\hat{r} = (\hat{x}, \hat{p})^T$, the displacement vector is

$$\mathbf{d} := (\langle \hat{x} \rangle, \langle \hat{p} \rangle)^T, \quad (3.24)$$

and the covariance matrix is

$$\gamma := \begin{pmatrix} \sigma_x^2 & \sigma_{xp} \\ \sigma_{xp} & \sigma_p^2 \end{pmatrix} = \begin{pmatrix} \langle \hat{x}^2 \rangle - \langle \hat{x} \rangle^2 & \frac{1}{2} \langle \hat{x} \hat{p} + \hat{p} \hat{x} \rangle - \langle \hat{x} \rangle \langle \hat{p} \rangle \\ \frac{1}{2} \langle \hat{x} \hat{p} + \hat{p} \hat{x} \rangle - \langle \hat{x} \rangle \langle \hat{p} \rangle & \langle \hat{p}^2 \rangle - \langle \hat{p} \rangle^2 \end{pmatrix}, \quad (3.25)$$

where σ_x^2 and σ_p^2 are the variances of $\hat{x}$ and $\hat{p}$, respectively, and σ_{xp} is their covariance. In the case of single-mode Gaussian states, the uncertainty relation of Eq. (2.66) can be expressed compactly as

$$\det \gamma = \sigma_x^2 \sigma_p^2 - \sigma_{xp}^2 \geq \frac{1}{4}. \quad (3.26)$$

In particular, if $\sigma_{xp} = 0$, it reduces to

$$\sigma_x^2 \sigma_p^2 \geq \frac{1}{4}. \quad (3.27)$$

The Wigner function of a general single-mode Gaussian state is given by

$$W_G(\mathbf{r}) = \frac{1}{2\pi\sqrt{\det\gamma}} e^{-\frac{1}{2}(\mathbf{r}-\mathbf{d})^T\gamma^{-1}(\mathbf{r}-\mathbf{d})}. \quad (3.28)$$

We now discuss some typical Gaussian states. Coherent states and squeezed vacuum states are single-mode pure Gaussian states, thermal states are single-mode Gaussian states which are mixed (except in the vacuum limit), and two-mode squeezed vacuum states are an important example of pure two-mode entangled Gaussian states.

3.2.2.1 Coherent states

When considering an ideal single-mode laser source, its emitted light can often be approximated by a coherent state. Such a state is an eigenstate of the annihilation operator $\hat{a}$, namely

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle, \quad (3.29)$$

where $\alpha \in \mathbb{C}$ is the eigenvalue associated with the coherent state $|\alpha\rangle$. The coherent state can be expanded in the Fock basis as [4]

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (3.30)$$

The photon-number distribution is therefore Poissonian,

$$P(n) = |\langle n|\alpha\rangle|^2 = e^{-|\alpha|^2} \frac{|\alpha|^{2n}}{n!}, \quad (3.31)$$

with average photon number $\langle \hat{n} \rangle = |\alpha|^2$.

The Wigner function of a coherent state $|\alpha\rangle$ is [4]

$$W_\alpha(x, p) = \frac{1}{\pi} e^{-(x-\sqrt{2}\operatorname{Re}(\alpha))^2 - (p-\sqrt{2}\operatorname{Im}(\alpha))^2}, \quad (3.32)$$

and is indeed Gaussian. The left panel of Figure 3.2 displays the Wigner function of a coherent state. Notice that the vacuum state $|0\rangle$ is a coherent state with $\alpha = 0$ (compare Eq. (3.23) and Eq. (3.32) with $\alpha = 0$). Coherent states are not orthogonal. Indeed, their overlap is

$$\langle \alpha|\beta\rangle = e^{-\frac{|\alpha|^2}{2} - \frac{|\beta|^2}{2} + \alpha^* \beta}. \quad (3.33)$$

Nevertheless, they form an overcomplete set, since they satisfy the resolution of the identity

$$\frac{1}{\pi} \int_{\mathbb{C}} d^2\alpha |\alpha\rangle \langle\alpha| = \hat{\mathbb{I}}. \quad (3.34)$$

The displacement vector of a coherent state is

$$\mathbf{d}_\alpha = \sqrt{2} (\text{Re}(\alpha), \text{Im}(\alpha))^T. \quad (3.35)$$

Thus, the complex amplitude α determines the displacement of the state in phase space. The covariance matrix is

$$\gamma_\alpha = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}. \quad (3.36)$$

Hence, coherent states saturate the uncertainty relation (Eq. (3.26)), since $\det \gamma_\alpha = 1/4$.

3.2.2.2 Thermal states

Thermal states are mixed Gaussian states that describe a single mode in thermal equilibrium. They can be seen as noisy versions of the vacuum state. In the Fock basis, a thermal state can be written as [4]

$$\hat{\rho}_{\text{th}} = \sum_{n=0}^{\infty} \frac{\bar{n}^n}{(\bar{n}+1)^{n+1}} |n\rangle \langle n|, \quad (3.37)$$

where $\bar{n} := \langle \hat{n} \rangle$ is the mean photon number. The thermal state has the Wigner function [4]

$$W_{\text{th}}(x, p) = \frac{1}{\pi(2\bar{n}+1)} e^{-\frac{x^2+p^2}{2\bar{n}+1}}. \quad (3.38)$$

which is displayed in the central panel of Figure 3.2. Its displacement vector is

$$\mathbf{d}_{\text{th}} = \mathbf{0}, \quad (3.39)$$

while its covariance matrix is

$$\gamma_{\text{th}} = \begin{pmatrix} \bar{n} + \frac{1}{2} & 0 \\ 0 & \bar{n} + \frac{1}{2} \end{pmatrix}. \quad (3.40)$$

When $\bar{n} = 0$, the thermal state reduces to the vacuum state $|0\rangle \langle 0|$. For $\bar{n} > 0$, the uncertainty relation is not saturated, since

$$\det \gamma_{\text{th}} = \left(\bar{n} + \frac{1}{2} \right)^2 > \frac{1}{4}. \quad (3.41)$$

3.2.2.3 Squeezed states

Squeezed states exhibit reduced fluctuations in one quadrature at the expense of increased fluctuations in the conjugate quadrature. Here, we focus on squeezed vacuum states, which are pure Gaussian states.

We consider a single-mode squeezed vacuum state with real squeezing parameter r and squeezing angle ϕ . With the conventions used for the squeezing operator introduced later in Section 3.3.1.2, its Fock-basis expansion is [4]

$$|0, r, \phi\rangle = \frac{1}{\sqrt{\cosh r}} \sum_{n=0}^{\infty} \left(-e^{i\phi}\right)^n \frac{\sqrt{(2n)!}}{2^n n!} (\tanh r)^n |2n\rangle. \quad (3.42)$$

In the remainder of this section, we consider $\phi = 0$. We notice that only even photon-number states appear in this decomposition.

For a squeezing in the x direction, the Wigner function is [4]

$$W_r(x, p) = \frac{1}{\pi} e^{-(e^{2r} x^2 + e^{-2r} p^2)}. \quad (3.43)$$

which is displayed in the right panel of Figure 3.2. Its displacement vector is

$$d_r = 0, \quad (3.44)$$

while its covariance matrix is

$$\gamma_r = \frac{1}{2} \begin{pmatrix} e^{-2r} & 0 \\ 0 & e^{2r} \end{pmatrix}. \quad (3.45)$$

Thus, for $r > 0$, the $\hat{x}$ quadrature is squeezed while the $\hat{p}$ quadrature is anti-squeezed. Since $\det \gamma_r = 1/4$, squeezed vacuum states saturate the uncertainty relation. More generally, one may also consider displaced squeezed states $|\alpha, r, \phi\rangle$ for which the displacement vector is non-zero while the covariance matrix retains the same squeezed form.

3.2.2.4 Two-mode squeezed vacuum states

An important family of pure Gaussian states, entangled for $r > 0$, is given by the two-mode squeezed vacuum (TMSV) state, which arises for instance from nonlinear optical interactions such as parametric down-conversion. In the Fock basis, it can be written as [4]

$$|\text{TMSV}(r, \phi)\rangle = \frac{1}{\cosh r} \sum_{n=0}^{\infty} \left(e^{i\phi}\right)^n (\tanh r)^n |n\rangle_A |n\rangle_B, \quad (3.46)$$

where $r \geq 0$ is the squeezing parameter and ϕ is the squeezing angle.

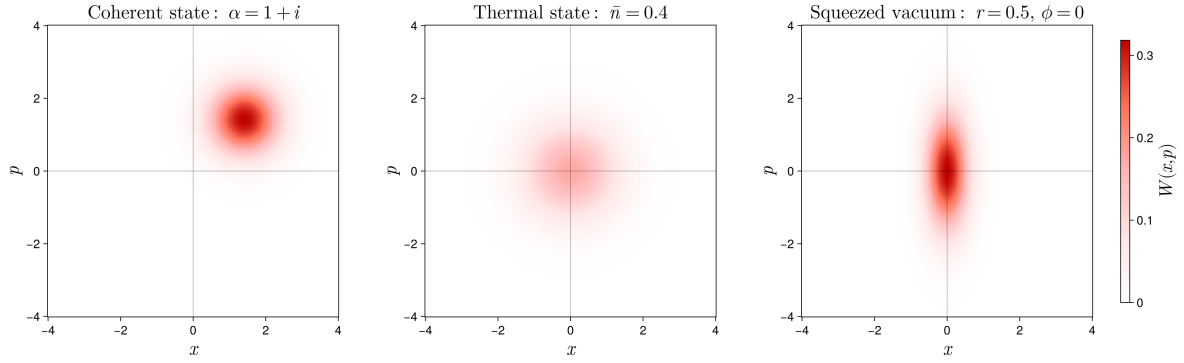


Figure 3.2: Wigner functions in the phase-space plane (x, p) for a coherent state with $\alpha = 1 + i$, a thermal state with mean photon number $\bar{n} = 0.4$, and a squeezed vacuum state with squeezing parameter $r = 0.5$ and squeezing angle $\phi = 0$. The coherent state is displaced along the x and p quadratures by $\sqrt{2}$ values, the thermal state is centered at the origin, and the squeezed vacuum state is squeezed along the x direction.

Equivalently, defining the complex parameter $\lambda = e^{i\phi} \tanh r$, this can be written as

$$|\text{TMSV}(\lambda)\rangle = \sqrt{1 - |\lambda|^2} \sum_{n=0}^{\infty} \lambda^n |n\rangle_A |n\rangle_B, \quad (3.47)$$

with $|\lambda| < 1$. In the common choice $\phi = 0$, which we consider in the remainder of this section, $\lambda = \tanh r$ is real, and this reduces to

$$|\text{TMSV}(\lambda)\rangle = \sqrt{1 - \lambda^2} \sum_{n=0}^{\infty} \lambda^n |n\rangle_A |n\rangle_B. \quad (3.48)$$

This state displays perfect photon-number correlations between the two modes. It can be understood as a continuous-variable analogue of a maximally entangled state, which would be recovered only in the unphysical limit $r \rightarrow \infty$.

With the quadrature ordering $\hat{\mathbf{r}} = (\hat{x}_A, \hat{p}_A, \hat{x}_B, \hat{p}_B)^T$, its Wigner function is [4]

$$W_{\text{TMSV}}(\mathbf{r}) = \frac{1}{\pi^2} e^{-\cosh(2r)(x_A^2 + p_A^2 + x_B^2 + p_B^2) + 2\sinh(2r)(x_A x_B - p_A p_B)}. \quad (3.49)$$

Its displacement vector is

$$\mathbf{d}_{\text{TMSV}} = \mathbf{0}, \quad (3.50)$$

while its covariance matrix is

$$\gamma_{\text{TMSV}} = \frac{1}{2} \begin{pmatrix} \cosh(2r) & 0 & \sinh(2r) & 0 \\ 0 & \cosh(2r) & 0 & -\sinh(2r) \\ \sinh(2r) & 0 & \cosh(2r) & 0 \\ 0 & -\sinh(2r) & 0 & \cosh(2r) \end{pmatrix}. \quad (3.51)$$

Using the hyperbolic identity $\cosh^2 a - \sinh^2 a = 1$, we find $\det \gamma_{\text{TMSV}} = 1/16$, which means that the (two-mode) uncertainty relation is saturated. Tracing out either mode gives a thermal state. Indeed, the reduced state of mode A is

$$\hat{\rho}_A = \text{Tr}_B [|\text{TMSV}(r)\rangle\langle\text{TMSV}(r)|] = \frac{1}{\cosh^2 r} \sum_{n=0}^{\infty} (\tanh^2 r)^n |n\rangle_A \langle n|_A, \quad (3.52)$$

and similarly for $\hat{\rho}_B$. This is a thermal state with mean photon number $\bar{n} = \sinh^2 r$, so that Eq. (3.37) is recovered.

The entanglement of the TMSV state can be checked at the level of second moments using Simon's criterion (see Theorem 2.9). As discussed in the beginning of Section 2.4.2, partial transposition with respect to mode B corresponds in phase space to the transformation $p_B \mapsto -p_B$. In terms of the covariance matrix, this amounts to

$$\gamma_{\text{TMSV}}^{T_B} := \Lambda_B \gamma_{\text{TMSV}} \Lambda_B, \quad \Lambda_B := \text{diag}(1, 1, 1, -1). \quad (3.53)$$

From Simon's criterion, for a separable two-mode Gaussian state, the partially transposed covariance matrix must satisfy

$$\gamma^{T_B} + \frac{i}{2} \Omega \geq 0. \quad (3.54)$$

Equivalently, the symplectic eigenvalues of γ^{T_B} must be at least $1/2$. These quantities are the positive numbers appearing in the Williamson normal form of γ^{T_B} , obtained by diagonalization with a symplectic matrix; they will be discussed in more detail in Section 3.3.1.1. For the TMSV state, one obtains

$$\tilde{v}_- = \frac{e^{-2r}}{2}, \quad (3.55)$$

which is smaller than $1/2$ for every $r > 0$. Simon's criterion is therefore violated, and the TMSV state is entangled for every non-zero squeezing parameter. The same conclusion can be obtained from the entropy-based separability criterion of Walborn (see Theorem 2.10). Defining the EPR-like quadratures

$$\hat{r}_- = \hat{x}_A - \hat{x}_B, \quad \hat{s}_+ = \hat{p}_A + \hat{p}_B, \quad (3.56)$$

one finds, for the TMSV state,

$$\sigma_{r_-}^2 = \sigma_{s_+}^2 = e^{-2r}. \quad (3.57)$$

Since these variables are Gaussian, their differential entropies satisfy Eq. (2.82). Thus,

$$h(R_-) + h(S_+) = \log(2\pi e) - \frac{2r}{\ln 2}. \quad (3.58)$$

This is smaller than $\log(2\pi e)$ for every $r > 0$, and therefore violates the separability bound.

3.2.3 Non-Gaussian states

Beyond the Gaussian regime, non-Gaussian states play an important role in several quantum information tasks. We discuss some examples here, ranging from coherent-state superpositions (cat states) to Fock-basis entangled states (NOON states). Since these states are non-Gaussian, their first and second moments do not fully characterize them. We also recall that the Fock states discussed in Section 3.2.1 are non-Gaussian for photon numbers $n \in \mathbb{N}_{>0}$.

3.2.3.1 Single-mode cat states

In analogy with Schrödinger's thought experiment, one can define superpositions of two coherent states, which intuitively become distinct for large $|\alpha|$. The even $|\mathcal{C}_+(\alpha)\rangle$ and odd $|\mathcal{C}_-(\alpha)\rangle$ cat states are defined as [4]

$$|\mathcal{C}_\pm(\alpha)\rangle = \mathcal{N}_\pm (|\alpha\rangle \pm |-\alpha\rangle), \quad (3.59)$$

where we take $\alpha \in \mathbb{R}$ ($\alpha \neq 0$ for odd cats) for simplicity. The normalization constant is

$$\mathcal{N}_\pm = \frac{1}{\sqrt{2(1 \pm e^{-2\alpha^2})}}. \quad (3.60)$$

Their Wigner functions are [4]

$$W_\pm(x, p) = \frac{2\mathcal{N}_\pm^2}{\pi} e^{-x^2 - p^2} \left[e^{-2\alpha^2} \cosh(2\sqrt{2}\alpha x) \pm \cos(2\sqrt{2}\alpha p) \right]. \quad (3.61)$$

The first term corresponds to the two Gaussian peaks associated with $|\alpha\rangle$ and $|-\alpha\rangle$, while the oscillating term is the interference contribution. These interference fringes are responsible for the non-Gaussian character of cat states and for the negative regions of their Wigner functions.

3.2.3.2 Two-mode cat states

Single-mode cat states can be generalized to two modes by considering coherent superpositions of two-mode coherent states. A useful family is given by the partially dephased entangled cat states [241, 253]

$$\hat{\rho}_\pm(z) = \frac{1}{\mathcal{M}_\pm(z)} \left[|\alpha, \beta\rangle \langle \alpha, \beta| + |-\alpha, -\beta\rangle \langle -\alpha, -\beta| \pm (1-z) (|\alpha, \beta\rangle \langle -\alpha, -\beta| + |-\alpha, -\beta\rangle \langle \alpha, \beta|) \right], \quad (3.62)$$

where $|\alpha, \beta\rangle := |\alpha\rangle_A |\beta\rangle_B$, $\alpha, \beta \in \mathbb{C}$, and $z \in [0, 1]$ is a dephasing parameter. The normalization constant is

$$\mathcal{M}_\pm(z) = 2 \left[1 \pm (1-z) e^{-2(|\alpha|^2 + |\beta|^2)} \right]. \quad (3.63)$$

For $z = 0$, the state reduces to a pure two-mode cat state of the form

$$|\psi_{\pm}(0)\rangle = \frac{1}{\sqrt{\mathcal{M}_{\pm}(0)}} (|\alpha, \beta\rangle \pm |-\alpha, -\beta\rangle), \quad (3.64)$$

which is entangled provided that both α and β are non-zero. For $z = 1$, the coherences between the two components vanish, and the state becomes an incoherent separable mixture of two product coherent states, i.e.,

$$\hat{\rho}_{\pm}(1) = \frac{1}{2} (|\alpha, \beta\rangle \langle \alpha, \beta| + |-\alpha, -\beta\rangle \langle -\alpha, -\beta|). \quad (3.65)$$

3.2.3.3 NOON states

Another important family of non-Gaussian states is given by finite superpositions of Fock states, known as NOON states [196]. A generalized NOON state with total photon number $N \in \mathbb{N}_{>0}$ can be written as

$$|\psi_N\rangle = \alpha |N\rangle_A |0\rangle_B + \beta |0\rangle_A |N\rangle_B, \quad (3.66)$$

where $\alpha, \beta \in \mathbb{C}$, and $|\alpha|^2 + |\beta|^2 = 1$. The balanced case,

$$|\text{NOON}\rangle = \frac{1}{\sqrt{2}} (|N\rangle_A |0\rangle_B + e^{i\phi} |0\rangle_A |N\rangle_B), \quad (3.67)$$

with a phase ϕ , is usually referred to as a NOON state.

Two specific cases are particularly well known. For $N = 1$ and $\alpha = \beta = 1/\sqrt{2}$, one obtains the single-photon path-entangled state

$$\frac{1}{\sqrt{2}} (|1\rangle_A |0\rangle_B + |0\rangle_A |1\rangle_B). \quad (3.68)$$

For $N = 2$ and $\alpha = -\beta = 1/\sqrt{2}$, one obtains

$$\frac{1}{\sqrt{2}} (|2\rangle_A |0\rangle_B - |0\rangle_A |2\rangle_B), \quad (3.69)$$

which is the state associated with the Hong–Ou–Mandel effect [126].

When $\alpha\beta \neq 0$, NOON states are entangled. They are also non-Gaussian, since they involve finite superpositions of Fock states (which are non-Gaussian for $N \in \mathbb{N}_{>0}$).

3.3 Quantum-optical operations

In quantum mechanics, closed-system evolutions are described by unitary operators. In quantum optics, many important operations, especially in the continuous-variable setting, are Gaussian, i.e., they map Gaussian states to Gaussian states and admit a simple description in phase space [243]. These include displacements, phase shifts, squeezing operations, beam splitters, and two-mode squeezing operations. Measurements, on the other hand, are not unitary operations and must be treated separately. Some of them, such as homodyne and heterodyne detection, are Gaussian measurements. Others, such as photon-number-resolving detection, are non-Gaussian, since they project onto the Fock basis.

3.3.1 Gaussian unitaries

We first recall the definition of a Gaussian unitary operation [243]. In quantum mechanics, the dynamics of a closed system is generated by a Hermitian Hamiltonian $\hat{H}$. Since we set $\hbar = 1$, the corresponding time-evolution operator is

$$\hat{U}(t) = e^{-i\hat{H}t}, \quad (3.70)$$

which is unitary, i.e., $\hat{U}^\dagger \hat{U} = \mathbb{1}$. In the following, we absorb the time dependence into the parameters of the transformation.

For an N -mode system, a Gaussian unitary is generated by a Hamiltonian that is at most quadratic in the annihilation and creation operators. We denote by

$$\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_N)^T \quad (3.71)$$

the column vector of annihilation operators, and by $\hat{\mathbf{a}}^\dagger$ its adjoint row vector. Thus, $\hat{\mathbf{a}}^{\dagger T}$ denotes the corresponding column vector of creation operators. Such a Hamiltonian can be written as

$$\hat{H} = i \left(\hat{\mathbf{a}}^\dagger \boldsymbol{\eta} + \hat{\mathbf{a}}^\dagger F \hat{\mathbf{a}} + \hat{\mathbf{a}}^\dagger G \hat{\mathbf{a}}^{\dagger T} \right) + \text{H.c.}, \quad (3.72)$$

where $\boldsymbol{\eta} \in \mathbb{C}^N$, $F, G \in \mathbb{C}^{N \times N}$, and H.c. denotes the Hermitian conjugate.

A Gaussian unitary operation is represented by a unitary operator $\hat{U}_G$. At the level of states, it acts as

$$\hat{\rho}' = \hat{U}_G \hat{\rho} \hat{U}_G^\dagger. \quad (3.73)$$

Equivalently, in the Heisenberg picture, it induces an affine Bogoliubov transformation of the mode operators,

$$\hat{U}_G^\dagger \hat{\mathbf{a}} \hat{U}_G = A \hat{\mathbf{a}} + B \hat{\mathbf{a}}^{\dagger T} + \boldsymbol{\eta}, \quad (3.74)$$

where $A, B \in \mathbb{C}^{N \times N}$.

Preservation of the canonical commutation relations of Eq. (3.17) imposes

$$AB^T = BA^T, \quad AA^\dagger = BB^\dagger + I_N, \quad (3.75)$$

where I_N is the N -dimensional identity matrix.

3.3.1.1 Symplectic formalism

In phase space, the same Gaussian unitary $\hat{U}_G$ is described by an affine symplectic transformation of the quadrature operators,

$$\hat{\mathbf{r}}' = \hat{U}_G^\dagger \hat{\mathbf{r}} \hat{U}_G = \mathcal{S}_G \hat{\mathbf{r}} + \boldsymbol{\delta}, \quad (3.76)$$

where $\hat{\mathbf{r}}$ is the quadrature vector (see Eq. (3.18)), $\boldsymbol{\delta} \in \mathbb{R}^{2N}$, and $\mathcal{S}_G$ is a real symplectic matrix, i.e.,

$$\mathcal{S}_G \Omega \mathcal{S}_G^T = \Omega. \quad (3.77)$$

Thus, $\mathcal{S}_G$ preserves the symplectic form Ω defined in Eq. (3.20). At the level of the displacement vector, one has

$$\mathbf{d}' = \mathcal{S}_G \mathbf{d} + \boldsymbol{\delta}, \quad (3.78)$$

whereas the covariance matrix transforms as

$$\gamma' = \mathcal{S}_G \gamma \mathcal{S}_G^T. \quad (3.79)$$

Williamson's theorem is a powerful tool in symplectic analysis, since it reduces the study of a covariance matrix to the study of its so-called symplectic eigenvalues.

Theorem 3.1 (Williamson's theorem [243, 246]). *Let γ be a real, positive-definite N -mode covariance matrix. Then, there exists a symplectic matrix $\mathcal{S}$ such that*

$$\gamma = \mathcal{S} \gamma^\oplus \mathcal{S}^T, \quad (3.80)$$

where

$$\gamma^\oplus = \bigoplus_{k=1}^N \nu_k I_2. \quad (3.81)$$

The positive numbers ν_k are called the symplectic eigenvalues of γ .

Equivalently, the eigenvalues of $i\Omega\gamma$ come in pairs $\pm\nu_k$. Hence, the symplectic eigenvalues of γ are the N positive eigenvalues of $i\Omega\gamma$. In terms of these quantities, the uncertainty relation of Eq. (2.66), namely

$$\gamma + \frac{i}{2}\Omega \geq 0, \quad (3.82)$$

is equivalent to

$$\nu_k \geq \frac{1}{2}, \quad \forall k = 1, \dots, N. \quad (3.83)$$

This decomposition has a direct physical interpretation. For an N -mode Gaussian state with zero displacement vector $\hat{\rho}_{0,\gamma}$, we can write

$$\hat{\rho}_{0,\gamma} = \hat{U}_{\mathcal{S}} \hat{\rho}_{0,\gamma^{\oplus}} \hat{U}_{\mathcal{S}}^{\dagger}, \quad (3.84)$$

where $\hat{U}_{\mathcal{S}}$ is the Gaussian unitary operator implementing the symplectic transformation $\mathcal{S}$. In this case, $\hat{\rho}_{0,\gamma^{\oplus}}$ is a tensor product of N thermal states. Indeed, the covariance matrix $\gamma^{\oplus}$ corresponds to the direct sum of covariance matrices of thermal states with mean photon numbers $\bar{n}_k = \nu_k - 1/2$, see Eq. (3.40). That is,

$$\gamma^{\oplus} = \bigoplus_{k=1}^N \gamma_{\text{th}}(\bar{n}_k) = \bigoplus_{k=1}^N \left(\bar{n}_k + \frac{1}{2} \right) I_2, \quad (3.85)$$

where $\gamma_{\text{th}}(\bar{n}_k)$ denotes the covariance matrix of a thermal state with mean photon number $\bar{n}_k$. Equivalently, in state space,

$$\hat{\rho}_{0,\gamma^{\oplus}} = \bigotimes_{k=1}^N \hat{\rho}_{\text{th}}(\bar{n}_k), \quad (3.86)$$

where $\hat{\rho}_{\text{th}}(\bar{n}_k)$ denotes a thermal state with mean photon number $\bar{n}_k$, see Eq. (3.37) for its Fock-basis expression.

In general, any N -mode Gaussian state with arbitrary displacement vector can be written using displacement operators as

$$\hat{\rho}_{\mathbf{d},\gamma} = \hat{D}(\mathbf{d}) \hat{U}_{\mathcal{S}} \hat{\rho}_{0,\gamma^{\oplus}} \hat{U}_{\mathcal{S}}^{\dagger} \hat{D}^{\dagger}(\mathbf{d}), \quad (3.87)$$

where $\hat{D}(\mathbf{d})$ denotes the displacement operator associated with the phase-space displacement vector $\mathbf{d}$, see Eq. (3.93) for its mathematical expression.

3.3.1.2 Single-mode operations

We now discuss the main single-mode Gaussian unitary operations in state space and in phase space.

Phase shift As a first elementary single-mode Gaussian unitary operation, a phase shift corresponds to a rotation in phase space and represents the free evolution of a harmonic oscillator. It is passive in the sense that it preserves the number of photons. Its expression is [4]

$$\hat{R}(\theta) = e^{-i\theta \hat{a}^\dagger \hat{a}} = e^{-i\theta \hat{n}}, \quad (3.88)$$

where $\theta \in \mathbb{R}$, and we used the number operator introduced in Eq. (3.11). Since $[\hat{R}(\theta), \hat{n}] = 0$, this operation indeed preserves the photon number. Its action on the annihilation and creation operators is

$$\hat{R}^\dagger(\theta) \hat{a} \hat{R}(\theta) = e^{-i\theta} \hat{a}, \quad \hat{R}^\dagger(\theta) \hat{a}^\dagger \hat{R}(\theta) = e^{i\theta} \hat{a}^\dagger. \quad (3.89)$$

Hence, for the quadratures,

$$\hat{R}^\dagger(\theta) \hat{x} \hat{R}(\theta) = \cos \theta \hat{x} + \sin \theta \hat{p}, \quad \hat{R}^\dagger(\theta) \hat{p} \hat{R}(\theta) = -\sin \theta \hat{x} + \cos \theta \hat{p}. \quad (3.90)$$

Therefore, the corresponding symplectic matrix is

$$\mathcal{S}_R(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (3.91)$$

This is indeed a phase-space rotation. There is no affine displacement term, so the first and second moments transform as

$$\mathbf{d}' = \mathcal{S}_R(\theta) \mathbf{d}, \quad \gamma' = \mathcal{S}_R(\theta) \gamma \mathcal{S}_R^T(\theta). \quad (3.92)$$

Displacement A displacement in phase space is implemented by the displacement operator [4]

$$\hat{D}(\alpha) = e^{\alpha \hat{a}^\dagger - \alpha^* \hat{a}}, \quad (3.93)$$

where $\alpha \in \mathbb{C}$. This operation is not photon-number preserving. Its inverse is another displacement, namely

$$\hat{D}^\dagger(\alpha) = \hat{D}^{-1}(\alpha) = \hat{D}(-\alpha). \quad (3.94)$$

The coherent states described in Section 3.2.2.1 are generated by displacing the vacuum state, i.e.,

$$|\alpha\rangle = \hat{D}(\alpha) |0\rangle. \quad (3.95)$$

In the Heisenberg picture, the annihilation and creation operators transform as

$$\hat{D}^\dagger(\alpha) \hat{a} \hat{D}(\alpha) = \hat{a} + \alpha, \quad \hat{D}^\dagger(\alpha) \hat{a}^\dagger \hat{D}(\alpha) = \hat{a}^\dagger + \alpha^*. \quad (3.96)$$

Equivalently, the quadratures transform as

$$\hat{D}^\dagger(\alpha)\hat{x}\hat{D}(\alpha) = \hat{x} + \sqrt{2}\text{Re}(\alpha), \quad \hat{D}^\dagger(\alpha)\hat{p}\hat{D}(\alpha) = \hat{p} + \sqrt{2}\text{Im}(\alpha). \quad (3.97)$$

Therefore, in phase space, a displacement is represented by the transformation

$$\hat{\mathbf{r}}' = \hat{\mathbf{r}} + \boldsymbol{\delta}_\alpha, \quad (3.98)$$

with the affine displacement term $\boldsymbol{\delta}_\alpha$ and symplectic matrix $\mathcal{S}_D$,

$$\boldsymbol{\delta}_\alpha = \sqrt{2}(\text{Re}(\alpha), \text{Im}(\alpha))^T, \quad \mathcal{S}_D = I_2. \quad (3.99)$$

At the level of first and second moments, one obtains

$$\mathbf{d}' = \mathbf{d} + \boldsymbol{\delta}_\alpha, \quad \boldsymbol{\gamma}' = \boldsymbol{\gamma}. \quad (3.100)$$

Hence, displacements change only the first moments of the state and leave the covariance matrix unchanged.

Squeezing Squeezing is an operation that reduces the fluctuations of one quadrature while increasing those of the conjugate quadrature. It is represented by the squeezing operator [4]

$$\hat{S}(\xi) = e^{\frac{1}{2}(\xi^* \hat{a}^2 - \xi \hat{a}^{\dagger 2})}, \quad (3.101)$$

where $\xi = r e^{i\phi}$, with $r \geq 0$ and $\phi \in \mathbb{R}$. This operation does not preserve the photon number. Its inverse is also a squeezing operator, namely

$$\hat{S}^\dagger(\xi) = \hat{S}^{-1}(\xi) = \hat{S}(-\xi). \quad (3.102)$$

The squeezed vacuum states described in Section 3.2.2.3 are obtained by applying this operator to the vacuum state,

$$|0, r, \phi\rangle = \hat{S}(r e^{i\phi})|0\rangle. \quad (3.103)$$

Its action on the annihilation and creation operators reads

$$\hat{S}^\dagger(\xi)\hat{a}\hat{S}(\xi) = \cosh(r)\hat{a} - e^{i\phi}\sinh(r)\hat{a}^\dagger, \quad \hat{S}^\dagger(\xi)\hat{a}^\dagger\hat{S}(\xi) = \cosh(r)\hat{a}^\dagger - e^{-i\phi}\sinh(r)\hat{a}. \quad (3.104)$$

For $\phi = 0$, this gives, for the quadratures,

$$\hat{S}^\dagger(r)\hat{x}\hat{S}(r) = e^{-r}\hat{x}, \quad \hat{S}^\dagger(r)\hat{p}\hat{S}(r) = e^r\hat{p}, \quad (3.105)$$

showing that the $\hat{x}$ quadrature is squeezed, while the $\hat{p}$ quadrature is anti-squeezed.

Therefore, for $\phi = 0$, the corresponding symplectic matrix is

$$\mathcal{S}_S(r) = \begin{pmatrix} e^{-r} & 0 \\ 0 & e^r \end{pmatrix}. \quad (3.106)$$

There is no affine displacement term, so the first and second moments transform as

$$\mathbf{d}' = \mathcal{S}_S(r)\mathbf{d}, \quad \gamma' = \mathcal{S}_S(r)\gamma\mathcal{S}_S^T(r). \quad (3.107)$$

General single-mode Gaussian unitary A general single-mode Gaussian unitary can be decomposed into a combination of displacements, squeezing operations, and phase shifts. Up to a global phase, it can be written as [243]

$$\hat{U}_G = \hat{D}(\alpha)\hat{R}(\theta)\hat{S}(r)\hat{R}(\varphi), \quad (3.108)$$

where $\alpha \in \mathbb{C}$ and $r, \theta, \varphi \in \mathbb{R}$. When this unitary acts on the vacuum, the rightmost rotation leaves the vacuum unchanged, since $\hat{R}(\varphi)|0\rangle = |0\rangle$. The remaining phase shift can be absorbed into the squeezing phase, because with our conventions

$$\hat{R}(\theta)\hat{S}(r)|0\rangle = \hat{S}\left(re^{-2i\theta}\right)|0\rangle. \quad (3.109)$$

Thus, a pure single-mode Gaussian state can be written as the following displaced squeezed state

$$|\alpha, r, \phi\rangle = \hat{D}(\alpha)\hat{S}(re^{i\phi})|0\rangle. \quad (3.110)$$

3.3.1.3 Two-mode operations

We review here the two standard two-mode Gaussian unitary operations, namely the beam splitter and the two-mode squeezing operation. The former conserves the total number of photons, while the latter does not. Let $\hat{b}$ and $\hat{b}^\dagger$ denote the annihilation and creation operators of the second mode B .

Beam splitter A beam splitter mixes the two modes A and B and is represented by the unitary operator [243]

$$\hat{U}_{BS}(\theta) = e^{\theta(\hat{a}^\dagger\hat{b} - \hat{a}\hat{b}^\dagger)}, \quad (3.111)$$

where $\theta \in \mathbb{R}$ is related to the transmissivity of the beam splitter by

$$T = \cos^2 \theta. \quad (3.112)$$

For $T = 1/2$, equivalently $\theta = \pi/4$, the beam splitter is called balanced.

In the Heisenberg picture, the modes transform as

$$\hat{U}_{\text{BS}}^\dagger(\theta) \hat{a} \hat{U}_{\text{BS}}(\theta) = \cos \theta \hat{a} + \sin \theta \hat{b}, \quad \hat{U}_{\text{BS}}^\dagger(\theta) \hat{b} \hat{U}_{\text{BS}}(\theta) = -\sin \theta \hat{a} + \cos \theta \hat{b}. \quad (3.113)$$

The corresponding transformation for the creation operators is obtained by taking the Hermitian conjugate. Hence, for the quadrature ordering $\hat{\mathbf{r}} = (\hat{x}_A, \hat{p}_A, \hat{x}_B, \hat{p}_B)^T$, the symplectic matrix is

$$\mathcal{S}_{\text{BS}}(\theta) = \begin{pmatrix} \cos \theta & 0 & \sin \theta & 0 \\ 0 & \cos \theta & 0 & \sin \theta \\ -\sin \theta & 0 & \cos \theta & 0 \\ 0 & -\sin \theta & 0 & \cos \theta \end{pmatrix}. \quad (3.114)$$

There is no affine displacement term. Therefore,

$$\mathbf{d}' = \mathcal{S}_{\text{BS}}(\theta) \mathbf{d}, \quad \boldsymbol{\gamma}' = \mathcal{S}_{\text{BS}}(\theta) \boldsymbol{\gamma} \mathcal{S}_{\text{BS}}^T(\theta). \quad (3.115)$$

When a balanced beam splitter is applied to the input state $|1\rangle_A |1\rangle_B$, which contains exactly one photon in each mode, the output is the well-known Hong–Ou–Mandel state [126] (which is a NOON state, see Eq. (3.69)),

$$\hat{U}_{\text{BS}}\left(\frac{\pi}{4}\right) |1\rangle_A |1\rangle_B = \frac{1}{\sqrt{2}} (|2\rangle_A |0\rangle_B - |0\rangle_A |2\rangle_B). \quad (3.116)$$

This entangled state expresses the fact that the two photons always leave through the same output mode: They are detected either both in the first output mode or both in the second one, while coincidence events are suppressed by two-photon interference.

Two-mode squeezer The two-mode squeezing operation couples the two modes, by amplifying them with photons created in pairs. For $\xi = r e^{i\phi}$, with $r \geq 0$ and $\phi \in \mathbb{R}$, it is represented by the unitary operator [4]

$$\hat{U}_{\text{TMS}}(\xi) = e^{(\xi \hat{a}^\dagger \hat{b}^\dagger - \xi^* \hat{a} \hat{b})}. \quad (3.117)$$

In the Heisenberg picture, the mode operators transform as

$$\hat{U}_{\text{TMS}}^\dagger(\xi) \hat{a} \hat{U}_{\text{TMS}}(\xi) = \cosh(r) \hat{a} + e^{i\phi} \sinh(r) \hat{b}^\dagger, \quad \hat{U}_{\text{TMS}}^\dagger(\xi) \hat{b} \hat{U}_{\text{TMS}}(\xi) = \cosh(r) \hat{b} + e^{i\phi} \sinh(r) \hat{a}^\dagger. \quad (3.118)$$

The corresponding transformation for the creation operators is obtained by taking the Hermitian conjugate.

For $\phi = 0$, and for the quadrature ordering $\hat{r} = (\hat{x}_A, \hat{p}_A, \hat{x}_B, \hat{p}_B)^T$, the symplectic matrix is

$$\mathcal{S}_{\text{TMS}}(r) = \begin{pmatrix} \cosh r & 0 & \sinh r & 0 \\ 0 & \cosh r & 0 & -\sinh r \\ \sinh r & 0 & \cosh r & 0 \\ 0 & -\sinh r & 0 & \cosh r \end{pmatrix}. \quad (3.119)$$

There is no affine displacement term. Therefore,

$$\mathbf{d}' = \mathcal{S}_{\text{TMS}}(r)\mathbf{d}, \quad \gamma' = \mathcal{S}_{\text{TMS}}(r)\gamma\mathcal{S}_{\text{TMS}}^T(r). \quad (3.120)$$

We can decompose the two-mode squeezing into a balanced beam splitter, two opposite single-mode squeezers, and the inverse balanced beam splitter, i.e., at the symplectic level,

$$\mathcal{S}_{\text{TMS}}(r) = \mathcal{S}_{\text{BS}}^T\left(\frac{\pi}{4}\right) [\mathcal{S}_S(-r) \oplus \mathcal{S}_S(r)] \mathcal{S}_{\text{BS}}\left(\frac{\pi}{4}\right). \quad (3.121)$$

The two-mode squeezed vacuum state introduced in Section 3.2.2.4 is generated by applying the two-mode squeezing unitary to the two-mode vacuum, that is,

$$|\text{TMSV}(r, \phi)\rangle = \hat{U}_{\text{TMS}}(r e^{i\phi}) |0\rangle_A |0\rangle_B = \frac{1}{\cosh r} \sum_{n=0}^{\infty} \left(e^{i\phi}\right)^n (\tanh r)^n |n\rangle_A |n\rangle_B. \quad (3.122)$$

3.3.1.4 General Gaussian unitary operation

The previous examples are the elementary building blocks of general Gaussian unitary operations. This is formalized by the Bloch–Messiah reduction.

Theorem 3.2 (Bloch–Messiah reduction [243]). *Any N -mode Gaussian unitary operation $\hat{U}_G$ can be decomposed, up to displacements and a global phase, as*

$$\hat{U}_G = \hat{U}_K \left[\bigotimes_{k=1}^N \hat{S}^{(k)}(r_k) \right] \hat{U}_L, \quad (3.123)$$

where $\hat{U}_K$ and $\hat{U}_L$ are passive Gaussian unitaries, i.e., multiport interferometers composed of phase shifts and beam splitters, and $\hat{S}^{(k)}(r_k)$ is a single-mode squeezing operation acting on mode k .

Equivalently, at the level of symplectic matrices, any N -mode symplectic matrix can be decomposed as

$$\mathcal{S}_G = K \left[\bigoplus_{k=1}^N \mathcal{S}_S(r_k) \right] L, \quad (3.124)$$

where K and L are orthogonal symplectic matrices, corresponding to passive linear-optical transformations, and $\mathcal{S}_S(r_k)$ is the symplectic matrix associated with a single-mode squeezing operation on mode k .

Thus, any N -mode Gaussian unitary can be realized using displacements, single-mode squeezers, phase shifts, and beam splitters.

3.3.2 Measurements

We now discuss some central non-unitary operations, namely Gaussian measurements, such as homodyne and heterodyne detection, and the non-Gaussian photon-number-resolving detection. Since measurements are not described by unitary operations on the measured system alone, they generally induce irreversible state changes on the measured system.

3.3.2.1 Homodyning and heterodyning

Homodyne and heterodyne detections are Gaussian measurements of continuous variables [243]. Homodyne detection gives access to one field quadrature, while heterodyne detection gives simultaneous access to both quadratures, at the cost of adding noise due to the impossibility of jointly measuring two conjugate quadratures without noise, since they are non-commuting observables.

Homodyning In the ideal description, homodyne detection corresponds to a projective measurement in the eigenbasis of a rotated quadrature [28, 243]

$$\hat{x}_\theta = \cos \theta \hat{x} + \sin \theta \hat{p}. \quad (3.125)$$

The corresponding measurement projectors are

$$\hat{\Pi}(x_\theta) = |x_\theta\rangle \langle x_\theta|, \quad (3.126)$$

where $|x_\theta\rangle$ denotes a generalized eigenstate of $\hat{x}_\theta$. For a state $\hat{\rho}$, the probability density of obtaining the outcome x_θ is

$$p(x_\theta) = \text{Tr}[\hat{\rho} \hat{\Pi}(x_\theta)]. \quad (3.127)$$

Therefore, $p(x_\theta)dx_\theta$ should be understood as the probability of obtaining an outcome within an infinitesimal interval dx_θ . Equivalently, in phase space, this probability distribution is obtained as the marginal of the Wigner function,

$$p(x_\theta) = \int dp_\theta W(x_\theta, p_\theta), \quad (3.128)$$

where p_θ is the quadrature conjugate to x_θ .

In practice, homodyne detection is implemented by mixing the input signal with a strong coherent field, called the local oscillator, on a balanced beam splitter. Two photodetectors then measure the intensities at the output ports, and the difference between these intensities gives access to the chosen quadrature. The measured quadrature is selected by controlling the phase of the local oscillator.

Heterodyning In contrast to homodyne detection, heterodyne detection gives access to both quadratures simultaneously. It is described by the coherent-state POVM [204]

$$\hat{\Pi}(\alpha) = \frac{1}{\pi} |\alpha\rangle \langle \alpha|, \quad (3.129)$$

which satisfies, see Eq. (3.34),

$$\int_{\mathbb{C}} d^2\alpha \hat{\Pi}(\alpha) = \hat{\mathbb{I}}. \quad (3.130)$$

For a state $\hat{\rho}$, the probability density of obtaining the outcome α is

$$p(\alpha) = \text{Tr} [\hat{\rho} \hat{\Pi}(\alpha)]. \quad (3.131)$$

Equivalently, this distribution is the Husimi Q function,

$$Q(\alpha) = \frac{1}{\pi} \langle \alpha | \hat{\rho} | \alpha \rangle, \quad (3.132)$$

which is another well-known phase-space distribution, distinct from the Wigner function.

Experimentally, heterodyne detection can be implemented by combining the input signal with an auxiliary vacuum mode on a balanced beam splitter. Homodyne measurements are then performed on two conjugate quadratures of the output modes, giving simultaneous estimates of the input quadratures $\hat{x}$ and $\hat{p}$ [243]. Because these quadratures are conjugate, the joint measurement comes with an added vacuum noise contribution.

3.3.2.2 Photon-number-resolving detection

Photon-number-resolving detection measures the number of photons present in a given mode [243]. Mathematically, an ideal photon-number measurement is described by projectors onto the Fock basis,

$$\hat{\Pi}_n = |n\rangle \langle n|, \quad (3.133)$$

so that, for a state $\hat{\rho}$, the probability of detecting n photons is

$$p(n) = \text{Tr} [\hat{\rho} \hat{\Pi}_n]. \quad (3.134)$$

Since Fock states are non-Gaussian for $n \in \mathbb{N}_{>0}$, photon-number-resolving detection is generally a non-Gaussian measurement.

Experimentally, photon-number resolution is challenging, especially for large photon numbers, due to finite detection efficiency, dark counts, and limited number resolution [218], although many advances in superconducting detectors have been made [40, 188].

In the remainder of this thesis, Gaussian and non-Gaussian states, as well as the corresponding operations, will be used to benchmark the criteria discussed in Chapters 5 and 6. Homodyne measurements will play a central role in Chapter 6, while photon-number-resolving detectors will be important in Chapter 5.

Part II

Entanglement detection

Multicopy technique

Introduced as a method for estimating polynomial functions of density matrix elements [29], the multicopy technique provides a way to access nonlinear functionals of a quantum state without reconstructing the full density operator. This method relies on several independently and identically prepared replicas of the state of interest obtained by repeating the same state-preparation procedure, rather than on cloning a single unknown quantum system, which would be forbidden by the no-cloning theorem [69, 248]. In practice, they may be generated by identical sources operated in parallel or by repeated preparations that are subsequently synchronized for the multicopy measurement [174].

The central idea is to turn a nonlinear function of $\hat{\rho}$ into a linear expectation value on a larger Hilbert space by considering a n -copy state $\hat{\rho}^{\otimes n}$. In quantum optics, this approach has been used to design experimentally motivated witnesses of nonclassicality [9, 104] and entanglement [105], thereby providing an alternative to full state tomography [50, 237] when only a specific nonlinear quantity is required.

In this chapter, we review the fundamentals of the multicopy technique in Section 4.1. We then discuss, in Section 4.2, examples relevant to this work, namely the purity, its extension to higher-order moments, and matrices of mode operator moments involved in witnessing nonclassicality or entanglement of quantum-optical states.

4.1 Description of the method

We now describe the multicopy method proposed in Ref. [29]. Let $\{|i\rangle\}$ be an orthonormal basis of the Hilbert space $\mathcal{H}$. The density operator can be decomposed as

$$\hat{\rho} = \sum_{i,j} \rho_{ij} |i\rangle\langle j|, \quad \text{with } \rho_{ij} := \langle i|\hat{\rho}|j\rangle. \quad (4.1)$$

The corresponding n -copy state is

$$\hat{\rho}^{\otimes n} := \underbrace{\hat{\rho} \otimes \cdots \otimes \hat{\rho}}_{n \text{ copies}}, \quad (4.2)$$

which acts on the enlarged Hilbert space

$$\mathcal{H}^{\otimes n} := \underbrace{\mathcal{H} \otimes \cdots \otimes \mathcal{H}}_{n \text{ copies}}. \quad (4.3)$$

In the product basis

$$|i_1, \dots, i_n\rangle := |i_1\rangle \otimes \cdots \otimes |i_n\rangle, \quad (4.4)$$

one has

$$\hat{\rho}^{\otimes n} = \sum_{i_1, \dots, i_n, j_1, \dots, j_n} \rho_{i_1 j_1} \cdots \rho_{i_n j_n} |i_1, \dots, i_n\rangle\langle j_1, \dots, j_n|. \quad (4.5)$$

We now consider, for simplicity, a homogeneous polynomial functional of degree n in the matrix elements of $\hat{\rho}$,

$$f(\hat{\rho}) = \sum_{i_1, \dots, i_n, j_1, \dots, j_n} c_{i_1 j_1 \dots i_n j_n} \rho_{i_1 j_1} \cdots \rho_{i_n j_n}, \quad (4.6)$$

where the coefficients $c_{i_1 j_1 \dots i_n j_n}$ are complex numbers. We introduce an operator $\hat{O}_f$ acting on $\mathcal{H}^{\otimes n}$, defined by

$$\hat{O}_f := \sum_{i_1, \dots, i_n, j_1, \dots, j_n} c_{i_1 j_1 \dots i_n j_n} |j_1, \dots, j_n\rangle \langle i_1, \dots, i_n|. \quad (4.7)$$

Then, for each term in the expansion of $\hat{\rho}^{\otimes n}$,

$$\text{Tr} [\hat{O}_f |i_1, \dots, i_n\rangle \langle j_1, \dots, j_n|] = \langle j_1, \dots, j_n | \hat{O}_f |i_1, \dots, i_n\rangle = c_{i_1 j_1 \dots i_n j_n}. \quad (4.8)$$

It follows that

$$\text{Tr} [\hat{O}_f \hat{\rho}^{\otimes n}] = \sum_{i_1, \dots, i_n, j_1, \dots, j_n} c_{i_1 j_1 \dots i_n j_n} \rho_{i_1 j_1} \cdots \rho_{i_n j_n} = f(\hat{\rho}). \quad (4.9)$$

Thus, a nonlinear polynomial functional of $\hat{\rho}$ can be obtained as the expectation value of a linear operator acting on n independently prepared replicas of the state. Terms of degree smaller than n can be treated similarly, since they can be embedded into the n -copy formalism through multiplication by suitable powers of $\text{Tr} \hat{\rho} = 1$.

In general, the operator $\hat{O}_f$ need not be Hermitian. If it is not Hermitian, it does not correspond directly to an observable. In that case, one may decompose it into its Hermitian and anti-Hermitian parts,

$$\hat{O}_f^{(R)} := \frac{\hat{O}_f + \hat{O}_f^\dagger}{2}, \quad \hat{O}_f^{(I)} := \frac{\hat{O}_f - \hat{O}_f^\dagger}{2i}. \quad (4.10)$$

Both $\hat{O}_f^{(R)}$ and $\hat{O}_f^{(I)}$ are Hermitian operators, and they give respectively the real and imaginary parts of $f(\hat{\rho})$,

$$\text{Re } f(\hat{\rho}) = \text{Tr} [\hat{O}_f^{(R)} \hat{\rho}^{\otimes n}], \quad \text{Im } f(\hat{\rho}) = \text{Tr} [\hat{O}_f^{(I)} \hat{\rho}^{\otimes n}]. \quad (4.11)$$

Therefore, even when the formal operator $\hat{O}_f$ is not itself an observable, the complex quantity $f(\hat{\rho})$ can be reconstructed from the expectation values of two Hermitian multicopy observables.

This whole construction extends straightforwardly to multipartite states, where each replica is a copy of the full multipartite state. This extension will be useful when considering entanglement detection.

4.2 Applications of the method

Let us now consider two relevant examples where the multicopy method has been used, notably in quantum optics, namely, the purity, as well as higher-order density matrix moments.

4.2.1 Purity of a quantum state

The purity of a quantum state $\hat{\rho}$ is defined as

$$\mathcal{P}(\hat{\rho}) := \text{Tr} [\hat{\rho}^2]. \quad (4.12)$$

Using the decomposition of Eq. (4.1), it can be expressed as the polynomial

$$\mathcal{P}(\hat{\rho}) = \sum_{i,j} \rho_{ij} \rho_{ji}. \quad (4.13)$$

This is a homogeneous polynomial of degree 2 in the matrix elements of $\hat{\rho}$. Therefore, according to the multicopy construction presented above, it can be obtained from two independently prepared replicas of the state [29]. In the notation introduced above, the only non-zero coefficients are

$$c_{i_1 j_1 i_2 j_2} = \delta_{i_1 j_2} \delta_{i_2 j_1}. \quad (4.14)$$

The corresponding operator (see Eq. (4.7)) is then

$$\begin{aligned} \hat{O}_{\mathcal{P}} &= \sum_{i_1, j_1, i_2, j_2} c_{i_1 j_1 i_2 j_2} |j_1, j_2\rangle \langle i_1, i_2| \\ &= \sum_{i,j} |j, i\rangle \langle i, j|. \end{aligned} \quad (4.15)$$

This operator is the swap operator $\hat{S}$ acting between the two replicas, i.e.,

$$\hat{S} := \sum_{i,j} |j, i\rangle \langle i, j|, \quad (4.16)$$

which exchanges the two copies of the state. Hence,

$$\mathcal{P}(\hat{\rho}) = \text{Tr} [\hat{S} (\hat{\rho} \otimes \hat{\rho})]. \quad (4.17)$$

This identity shows that the purity, although nonlinear in $\hat{\rho}$, is the expectation value of a linear observable acting on two copies.

In quantum optics, the expectation value of the swap operator can be accessed by interfering the two replicas on a balanced beam splitter [104, 135]; see Section 3.3.1.3 for its definition. For two single-mode replicas, the beam splitter transforms the input modes as

$$\hat{a}'_1 = \frac{\hat{a}_1 + \hat{a}_2}{\sqrt{2}}, \quad \hat{a}'_2 = \frac{-\hat{a}_1 + \hat{a}_2}{\sqrt{2}}. \quad (4.18)$$

These output modes correspond respectively to the symmetric and antisymmetric combinations of the two input modes. Indeed, in terms of the output modes,

$$\hat{S}^\dagger \hat{a}'_1 \hat{S} = \hat{a}'_1, \quad \hat{S}^\dagger \hat{a}'_2 \hat{S} = -\hat{a}'_2. \quad (4.19)$$

Thus, the symmetric output mode remains unchanged under the exchange of the two replicas, whereas the antisymmetric output mode acquires a minus sign. The latter transformation in Eq. (4.19) is precisely the action of the photon-number parity operator (which can also be viewed as a phase-shift unitary with phase equal to π , as introduced in Section 3.3.1.2). It is defined as

$$\hat{\Pi} := (-1)^{\hat{n}} = e^{i\pi\hat{n}}, \quad (4.20)$$

and satisfies

$$\hat{\Pi}^\dagger \hat{a} \hat{\Pi} = -\hat{a}. \quad (4.21)$$

An output state containing m_1 photons in the symmetric mode and m_2 photons in the antisymmetric mode can be written as

$$|m_1\rangle_{\text{out sym}} \otimes |m_2\rangle_{\text{out antisym}} = \frac{(\hat{a}'_1)^\dagger{}^{m_1} (\hat{a}'_2)^\dagger{}^{m_2}}{\sqrt{m_1! m_2!}} (|0\rangle_{\text{out sym}} \otimes |0\rangle_{\text{out antisym}}). \quad (4.22)$$

Since each creation operator $\hat{a}'_2^\dagger$ acquires a minus sign under the swap, we obtain

$$\hat{S}(|m_1\rangle_{\text{out sym}} \otimes |m_2\rangle_{\text{out antisym}}) = (-1)^{m_2} (|m_1\rangle_{\text{out sym}} \otimes |m_2\rangle_{\text{out antisym}}). \quad (4.23)$$

The eigenvalue of the swap is therefore determined by the parity of the photon number in the antisymmetric output mode. In this output basis, the swap operator is thus represented by the operator $(\hat{1} \otimes \hat{\Pi})$. Equivalently, if $\hat{U}_{\text{BS}}^{1/2}$ denotes the balanced beam-splitter unitary, transforming this output-basis operator back to the input basis gives

$$\hat{S} = (\hat{U}_{\text{BS}}^{1/2})^\dagger (\hat{1} \otimes \hat{\Pi}) \hat{U}_{\text{BS}}^{1/2}, \quad (4.24)$$

which means that measuring the parity of the second mode after the beam splitter is equal to measuring the swap operator on the two input replicas.

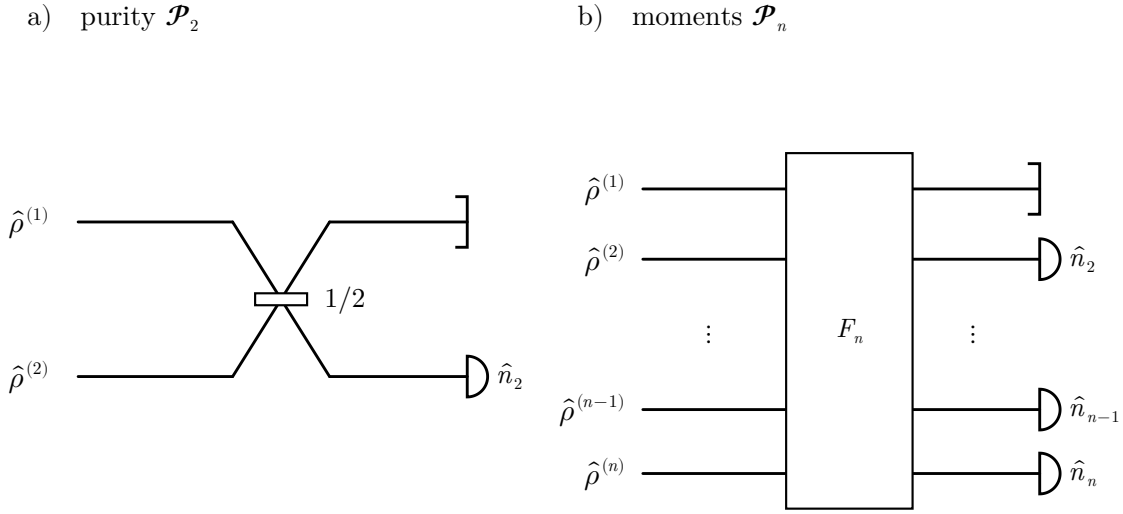


Figure 4.1: **a)** Multicopy measurement of the purity $\mathcal{P}_2 = \text{Tr}[\hat{\rho}^2]$. Two independently prepared replicas of the state interfere on a balanced beam splitter. The particle number $\hat{n}_2$ is measured in the second output mode, whereas the first output mode is not detected. **b)** Multicopy measurement of the n th moment $\mathcal{P}_n = \text{Tr}[\hat{\rho}^n]$. The n replicas are sent through an n -mode Fourier interferometer F_n , after which particle-number measurements are performed on output modes 2 to n (the first output mode does not need to be measured).

Consequently, using Eq. (4.17) and the cyclicity of the trace, one finds

$$\begin{aligned} \mathcal{P}(\hat{\rho}) &= \text{Tr} \left[(\hat{\mathbb{I}} \otimes \hat{\Pi}) \hat{U}_{\text{BS}}^{1/2} (\hat{\rho} \otimes \hat{\rho}) (\hat{U}_{\text{BS}}^{1/2})^\dagger \right] \\ &=: \langle \hat{\mathbb{I}} \otimes \hat{\Pi} \rangle_{\hat{U}_{\text{BS}}^{1/2} (\hat{\rho} \otimes \hat{\rho}) (\hat{U}_{\text{BS}}^{1/2})^\dagger}. \end{aligned} \quad (4.25)$$

Thus, the purity is obtained by preparing many pairs of identical replicas, interfering each pair on a balanced beam splitter, measuring the photon-number parity of the antisymmetric output mode, and averaging the recorded parity values over all runs. This is illustrated in the left panel of Figure 4.1.

For multimode states, the same idea applies by swapping the two full replicas, which can be implemented by applying balanced beam splitters pairwise between corresponding modes of the two copies, followed by suitable parity measurements.

4.2.2 Higher-order moments

Let us now consider the higher-order moments of the density matrix,

$$\mathcal{P}_n(\hat{\rho}) := \text{Tr}[\hat{\rho}^n], \quad (4.26)$$

where $n \in \mathbb{N}_{>1}$. For $n = 2$, we recover the purity, $\mathcal{P}_2(\hat{\rho}) = \mathcal{P}(\hat{\rho})$. These moments are, for instance, involved in the definition of the Rényi entropies,

$$S_n(\hat{\rho}) := \frac{1}{1-n} \log(\text{Tr}[\hat{\rho}^n]) = \frac{1}{1-n} \log(\mathcal{P}_n(\hat{\rho})). \quad (4.27)$$

Thus, estimating $\mathcal{P}_n(\hat{\rho})$ gives access to the Rényi entropy of integer order $n \geq 2$.

Using the decomposition of Eq. (4.1), one finds

$$\text{Tr}[\hat{\rho}^n] = \sum_{i_1, \dots, i_n} \rho_{i_1 i_2} \rho_{i_2 i_3} \cdots \rho_{i_n i_1}. \quad (4.28)$$

Equivalently, in the notation of the general multicopy construction, this corresponds to the coefficients

$$c_{i_1 j_1 \dots i_n j_n} = \delta_{j_1 i_2} \delta_{j_2 i_3} \cdots \delta_{j_{n-1} i_n} \delta_{j_n i_1}. \quad (4.29)$$

The associated multicopy operator is therefore

$$\begin{aligned} \hat{O}_{\mathcal{P}_n} &= \sum_{i_1, \dots, i_n, j_1, \dots, j_n} c_{i_1 j_1 \dots i_n j_n} |j_1, \dots, j_n\rangle \langle i_1, \dots, i_n| \\ &= \sum_{i_1, \dots, i_n} |i_2, i_3, \dots, i_n, i_1\rangle \langle i_1, i_2, \dots, i_n|. \end{aligned} \quad (4.30)$$

This operator is the (left) cyclic-shift operator, which we denote here by $\hat{V}_n$. It acts on basis states as

$$\hat{V}_n |i_1, i_2, \dots, i_n\rangle = |i_2, \dots, i_n, i_1\rangle. \quad (4.31)$$

Hence,

$$\mathcal{P}_n(\hat{\rho}) = \text{Tr}[\hat{V}_n \hat{\rho}^{\otimes n}]. \quad (4.32)$$

The case $n = 2$ reduces to the swap operator discussed above. For $n > 2$, however, $\hat{V}_n$ is unitary but not Hermitian. Its adjoint is the cyclic shift in the opposite direction,

$$\hat{V}_n^\dagger |i_1, i_2, \dots, i_n\rangle = |i_n, i_1, \dots, i_{n-1}\rangle. \quad (4.33)$$

Thus, $\hat{V}_n$ is not itself an observable.

Since $\text{Tr}[\hat{\rho}^n]$ is real due to $\hat{\rho}$ being Hermitian, the imaginary part of the expectation value of the cyclic shift vanishes. We may therefore consider its Hermitian part,

$$\hat{V}_n^{(R)} = \frac{\hat{V}_n + \hat{V}_n^\dagger}{2}, \quad (4.34)$$

so that

$$\mathcal{P}_n(\hat{\rho}) = \text{Tr}[\hat{V}_n^{(R)} \hat{\rho}^{\otimes n}]. \quad (4.35)$$

In quantum optics, the cyclic-shift operator $\hat{V}_n$ can be diagonalized using an n -mode Fourier interferometer [53]. With the convention

$$\hat{a}'_k = \frac{1}{\sqrt{n}} \sum_{\ell=1}^n e^{\frac{2\pi i}{n}(k-1)(\ell-1)} \hat{a}_\ell, \quad (4.36)$$

the corresponding passive unitary $\hat{U}_F$ is defined by

$$\hat{U}_F^\dagger \hat{a}_k \hat{U}_F = \hat{a}'_k. \quad (4.37)$$

Since the Fourier matrix is unitary, we find

$$\sum_{k=1}^n \hat{a}_k^\dagger \hat{a}'_k = \sum_{k=1}^n \hat{a}_k^\dagger \hat{a}_k, \quad (4.38)$$

that is, the total number of photons is conserved. It can therefore be implemented using only beam splitters and phase shifters [243].

After the Fourier interferometer, the cyclic shift is mapped onto an operator diagonal in the output photon-number basis,

$$\hat{V}_n = \hat{U}_F^\dagger e^{-\frac{2\pi i}{n} \sum_{k=1}^n (k-1) \hat{n}_k} \hat{U}_F, \quad (4.39)$$

where $\hat{n}_k$ denotes the photon-number operator of the k th output mode. Since the moment $\mathcal{P}_n(\hat{\rho}) = \text{Tr}[\hat{\rho}^n]$ is real, it can be estimated by averaging the real part of this phase factor over many experimental runs,

$$\mathcal{P}_n(\hat{\rho}) = \left\langle \cos \left[\frac{2\pi}{n} \sum_{k=1}^n (k-1) n_k \right] \right\rangle_{\hat{U}_F \hat{\rho}^{\otimes n} \hat{U}_F^\dagger}. \quad (4.40)$$

Thus, $\mathcal{P}_n(\hat{\rho})$ can be estimated by preparing n independent replicas of the state, sending them through an n -mode Fourier interferometer, measuring the output photon numbers, and averaging the corresponding cosine weights over sufficiently many runs. This is illustrated in the right panel of Figure 4.1.

4.2.3 Matrices of mode-operator moments

Let us now describe how the multicopy technique can be used to address experimentally relevant questions for certifying nonclassicality and entanglement in quantum optics.

Nonclassicality In Ref. [210], Shchukin and Vogel proposed a hierarchy of necessary and sufficient conditions for deciding whether a quantum state of light is classical or nonclassical. Here, classicality is understood in the optical sense, i.e., a state is classical if its Glauber–Sudarshan P -representation can be interpreted as a positive probability measure over coherent states,

$$\hat{\rho} = \int d^2\alpha P(\alpha) |\alpha\rangle\langle\alpha|. \quad (4.41)$$

If no such positive P -representation exists, the state is said to be nonclassical.

The Shchukin–Vogel classicality criterion is based on an infinite matrix of normally ordered moments. Introducing a sequence of mode-operator monomials,

$$\hat{f} := (\hat{f}_1, \hat{f}_2, \hat{f}_3, \dots) := (\hat{1}, \hat{a}, \hat{a}^\dagger, \hat{a}^2, \hat{a}^\dagger \hat{a}, \hat{a}^{\dagger 2}, \dots), \quad (4.42)$$

one defines

$$M_{ij}(\hat{\rho}) := \langle : \hat{f}_i^\dagger \hat{f}_j : \rangle = \text{Tr} \left[: \hat{f}_i^\dagger \hat{f}_j : \hat{\rho} \right], \quad (4.43)$$

where $: \dots :$ denotes normal ordering, which means creation operators are placed to the left of annihilation operators. A state is classical if and only if the full infinite matrix $M(\hat{\rho})$ is positive semidefinite. Therefore, by Sylvester’s criterion, the negativity of any principal minor of $M(\hat{\rho})$ certifies nonclassicality. Conversely, if none of the tested minors is negative, no general conclusion can be drawn.

We denote by $d_{i_1 \dots i_r}$ the principal minor obtained by keeping the rows and columns indexed by $i_1, \dots, i_r$,

$$d_{i_1 \dots i_r} := \det \left[M_{i_\mu i_\nu}(\hat{\rho}) \right]_{\mu, \nu=1}^r. \quad (4.44)$$

Since each entry $M_{ij}(\hat{\rho})$ is linear in $\hat{\rho}$, the determinant $d_{i_1 \dots i_r}$ is a polynomial functional of degree r . It can therefore be accessed, in principle, by a multicopy observable acting on r independently prepared replicas of the state. In Ref. [9], it was shown that several such minors, including d_{12} , d_{14} , d_{23} , d_{15} , d_{123} , and d_{1235} , admit interferometric implementations based on passive linear optics and photon-number-resolving detection.

Let us focus on the minor d_{23} , which reads

$$d_{23} := \det \begin{pmatrix} \langle \hat{a}^\dagger \hat{a} \rangle & \langle \hat{a}^{\dagger 2} \rangle \\ \langle \hat{a}^2 \rangle & \langle \hat{a}^\dagger \hat{a} \rangle \end{pmatrix} = \langle \hat{a}^\dagger \hat{a} \rangle^2 - |\langle \hat{a}^2 \rangle|^2. \quad (4.45)$$

This is a second-degree polynomial functional of $\hat{\rho}$, and can therefore be written as a two-copy expectation value. Introducing two replicas with annihilation operators $\hat{a}_1$ and $\hat{a}_2$, one obtains

$$d_{23} = \text{Tr} [\hat{O}_{d_{23}} \hat{\rho}^{\otimes 2}], \quad (4.46)$$

with

$$\hat{O}_{d_{23}} = \hat{n}_1 \hat{n}_2 - \frac{1}{2} (\hat{a}_1^2 \hat{a}_2^{\dagger 2} + \hat{a}_1^{\dagger 2} \hat{a}_2^2), \quad (4.47)$$

where $\hat{n}_k = \hat{a}_k^\dagger \hat{a}_k$.

This observable can be diagonalized by a simple two-mode linear-optical circuit. After applying a $\pi/2$ phase shift on the second replica, denoted by $\hat{R}_2(\pi/2)$ and explicitly given in Section 3.3.1.2, followed by a balanced beam splitter $\hat{U}_{\text{BS}}^{1/2}$, explicitly given in Section 3.3.1.3, one has

$$\hat{O}_{d_{23}} = \hat{R}_2^\dagger(\pi/2) (\hat{U}_{\text{BS}}^{1/2})^\dagger \hat{B}_{23} \hat{U}_{\text{BS}}^{1/2} \hat{R}_2(\pi/2), \quad (4.48)$$

where $\hat{B}_{23}$ is diagonal in the output photon-number basis,

$$\hat{B}_{23} = \frac{1}{2} \left[(\hat{n}_{a'_1} - \hat{n}_{a'_2})^2 - (\hat{n}_{a'_1} + \hat{n}_{a'_2}) \right]. \quad (4.49)$$

Thus, experimentally, one prepares two replicas, sends them through this circuit, measures the output photon numbers $n_{a'_1}$ and $n_{a'_2}$ with photon-number-resolving detectors, and averages the value

$$B_{23}(n_{a'_1}, n_{a'_2}) = \frac{1}{2} \left[(n_{a'_1} - n_{a'_2})^2 - (n_{a'_1} + n_{a'_2}) \right], \quad (4.50)$$

over many experimental runs.

This witness detects, for example, the nonclassicality of squeezed vacuum states, introduced in Section 3.2.2.3. Consider two identical replicas of the squeezed vacuum

$$|0, r, \phi\rangle = \frac{1}{\sqrt{\cosh r}} \sum_{n=0}^{\infty} (-e^{i\phi})^n \frac{\sqrt{(2n)!}}{2^n n!} (\tanh r)^n |2n\rangle. \quad (4.51)$$

After the $\pi/2$ phase shift on the second replica and the balanced beam splitter, the two output modes are in a two-mode squeezed vacuum state (see Section 3.2.2.4),

$$|\psi_{\text{out}}\rangle = \sqrt{1 - \lambda^2} \sum_{n=0}^{\infty} \lambda^n |n\rangle_{a'_1} |n\rangle_{a'_2}, \quad \lambda = \tanh r. \quad (4.52)$$

Hence, the output photon numbers are perfectly correlated, $n_{a'_1} = n_{a'_2} = n$. For each outcome, the measured value is therefore

$$B_{23}(n, n) = \frac{1}{2} [(n - n)^2 - (n + n)] = -n. \quad (4.53)$$

Since the probability of obtaining n photons in each output mode is

$$p_n = (1 - \lambda^2) \lambda^{2n}, \quad (4.54)$$

the average value is

$$\begin{aligned} d_{23} &= \langle B_{23} \rangle \\ &= - \sum_{n=0}^{\infty} n(1 - \lambda^2) \lambda^{2n} \\ &= - \frac{\lambda^2}{1 - \lambda^2} \\ &= - \sinh^2 r. \end{aligned} \quad (4.55)$$

Thus, $d_{23} < 0$ for any $r > 0$, which certifies the nonclassicality of the squeezed vacuum.

Entanglement Soon after proposing their nonclassicality witness, Shchukin and Vogel introduced a similar hierarchy of moment conditions for bipartite continuous-variable entanglement [209]. The idea is to apply the same positivity logic not to the state $\hat{\rho}$ itself, but to its partial transpose $\hat{\rho}^{T_B}$. Recall the Peres–Horodecki PPT criterion from Theorem 2.5: if $\hat{\rho}$ is separable, then $\hat{\rho}^{T_B}$ must be positive semidefinite. Therefore, any violation of this positivity condition certifies entanglement.

Let $\hat{a}$ and $\hat{b}$ denote the annihilation operators of the two modes. Following Shchukin and Vogel, we consider operators $\hat{f}$ admitting a normally ordered expansion in the two-mode operators, and introduce a sequence of monomials

$$\hat{f} := (\hat{f}_1, \hat{f}_2, \hat{f}_3, \dots) := (\hat{1}, \hat{a}, \hat{a}^\dagger, \hat{b}, \hat{b}^\dagger, \dots), \quad (4.56)$$

where each $\hat{f}_i$ is of the form $\hat{a}^{\dagger n} \hat{a}^m \hat{b}^{\dagger k} \hat{b}^l$. For any finite linear combination $\hat{f} = \sum_i c_i \hat{f}_i$, positivity of $\hat{\rho}^{T_B}$ implies

$$\text{Tr} [\hat{f}^\dagger \hat{f} \hat{\rho}^{T_B}] \geq 0. \quad (4.57)$$

This defines the partially transposed moment matrix

$$M_{ij}^{T_B}(\hat{\rho}) := \text{Tr} [\hat{f}_i^\dagger \hat{f}_j \hat{\rho}^{T_B}]. \quad (4.58)$$

The entries of this matrix can then be expressed in terms of experimentally accessible moments of the original state $\hat{\rho}$. The Shchukin–Vogel criterion states that $\hat{\rho}^{TB} \geq 0$ is equivalent to the positivity of the full infinite moment matrix $M^{TB}(\hat{\rho})$. By Sylvester’s criterion, this is equivalent to the nonnegativity of all its principal minors [105]. We denote these minors by

$$d_{i_1 \dots i_r} := \det \left[M_{i_\mu i_\nu}^{TB}(\hat{\rho}) \right]_{\mu, \nu=1}^r. \quad (4.59)$$

Thus, if one finds $d_{i_1 \dots i_r} < 0$, the partial transpose is not positive and the state is necessarily entangled. As in the nonclassicality case, each entry of $M^{TB}(\hat{\rho})$ is linear in $\hat{\rho}$, so $d_{i_1 \dots i_r}$ is a polynomial functional of degree r . It can therefore be associated with a multicopy observable acting on r independently prepared replicas of the full two-mode state.

Following the ordering above, the minor d_{124} is obtained by keeping the monomials $\hat{f}_1 = \hat{\mathbb{I}}$, $\hat{f}_2 = \hat{a}$, and $\hat{f}_4 = \hat{b}$. After rewriting the partially transposed moments as ordinary moments of $\hat{\rho}$, this minor reads

$$d_{124} = \det \begin{pmatrix} 1 & \langle \hat{a} \rangle & \langle \hat{b}^\dagger \rangle \\ \langle \hat{a}^\dagger \rangle & \langle \hat{a}^\dagger \hat{a} \rangle & \langle \hat{a}^\dagger \hat{b}^\dagger \rangle \\ \langle \hat{b} \rangle & \langle \hat{a} \hat{b} \rangle & \langle \hat{b}^\dagger \hat{b} \rangle \end{pmatrix}. \quad (4.60)$$

This determinant is a third-degree polynomial functional of $\hat{\rho}$, and thus corresponds to a three-copy multicopy observable. In quantum optics, Ref. [105] shows that this three-copy observable can be accessed with passive linear optics and photon-number-resolving detection. More precisely, the measurement can be decomposed into three passive six-mode circuits, denoted by $\hat{U}_{D,j}$ with $j = 1, 2, 3$, acting on the three replicas of Alice’s and Bob’s modes. After each circuit, one measures the same photon-number observable

$$\hat{C} = \frac{1}{2} \left(\hat{n}_{a'_2} \hat{n}_{b'_3} + \hat{n}_{a'_3} \hat{n}_{b'_2} \right), \quad (4.61)$$

which is diagonal in the output photon-number basis. The determinant is then obtained as the linear combination

$$d_{124} = \langle \hat{C} \rangle_1 - \langle \hat{C} \rangle_2 + \langle \hat{C} \rangle_3, \quad (4.62)$$

where

$$\langle \hat{C} \rangle_j = \text{Tr} \left[\hat{C} \hat{U}_{D,j} \hat{\rho}^{\otimes 3} \hat{U}_{D,j}^\dagger \right]. \quad (4.63)$$

Thus, the experimental procedure consists of preparing three independent replicas of the same two-mode state, applying the required passive interferometers, and averaging suitable products of output photon numbers.

For the two-mode squeezed vacuum introduced in Section 3.2.2.4, i.e.,

$$|\text{TMSV}(\lambda)\rangle = \sqrt{1-\lambda^2} \sum_{n=0}^{\infty} \lambda^n |n\rangle_A |n\rangle_B, \quad (4.64)$$

with $0 \leq \lambda < 1$, one has

$$\langle \hat{a} \rangle = \langle \hat{b} \rangle = 0, \quad \langle \hat{a}^\dagger \hat{a} \rangle = \langle \hat{b}^\dagger \hat{b} \rangle = \frac{\lambda^2}{1-\lambda^2}, \quad (4.65)$$

and

$$\langle \hat{a} \hat{b} \rangle = \frac{\lambda}{1-\lambda^2}. \quad (4.66)$$

The determinant therefore reduces to

$$\begin{aligned} d_{124} &= \langle \hat{a}^\dagger \hat{a} \rangle \langle \hat{b}^\dagger \hat{b} \rangle - |\langle \hat{a} \hat{b} \rangle|^2 \\ &= \frac{\lambda^4}{(1-\lambda^2)^2} - \frac{\lambda^2}{(1-\lambda^2)^2} \\ &= -\frac{\lambda^2}{1-\lambda^2}. \end{aligned} \quad (4.67)$$

Hence, $d_{124} < 0$ for any $\lambda > 0$, which certifies the entanglement of the two-mode squeezed vacuum.

In Ref. [105], various multicopy measurements of selected Shchukin–Vogel entanglement witnesses were proposed. In particular, the minor d_{124} was shown to provide a stronger test than the Duan et al. criterion [72] presented in Theorem 2.7. Other minors target genuinely non-Gaussian features. Indeed, d_{149} can certify entanglement in some odd two-mode cat states (see Section 3.2.3.2 for their definition), while $d_{1,9,13}$ is useful for detecting the entanglement of NOON states (see Section 3.2.3.3 for their definition).

In the remainder of this thesis, we focus on continuous-variable entanglement. The multicopy technique is used in Chapter 5 to measure moments of the partially transposed state, and in Chapter 6 to measure moments of the realigned state.

5 Detecting strongly non-Gaussian entanglement

The content of this chapter is based on the following published work: Serge Deside, Tobias Haas, and Nicolas J. Cerf. “Detecting strongly non-Gaussian entanglement”. Physical Review Research 8, 033064, July 2026 [63].

In this chapter, we develop continuous-variable entanglement criteria based on moments of the partially transposed state and establish their experimental accessibility through multicopy interferometry. We benchmark the resulting criteria on Gaussian and non-Gaussian states and show that they are particularly effective in the strongly non-Gaussian regime, which includes the entire class of NOON states. We further assess their performance under realistic experimental imperfections through analytical and numerical studies of losses, noise between state copies, and finite statistics.

5.1 Introduction

As discussed in Chapter 2, efficiently certifying entanglement becomes particularly challenging in continuous-variable systems beyond the Gaussian regime. While covariance-matrix criteria are particularly effective for Gaussian states, the detection of strongly non-Gaussian entanglement¹ generally requires information beyond second moments. States involving high-order Fock-space coherences, such as NOON states [71, 196], therefore provide a demanding benchmark for experimentally accessible criteria.

In this context, criteria implied by the seminal positive partial transpose (PPT) criterion (see Theorem 2.5) and based on partially transposed (PT) moments $p_n := \text{Tr}[(\hat{\rho}^{T_B})^n]$ have recently been developed for discrete-variable systems [80, 169, 252], together with randomized-measurement schemes for their estimation. Here, we address both their extension to continuous-variable systems and the development of an experimentally accessible method for measuring the corresponding moments. The measurement schemes considered in this chapter rely on particle-number-resolving detection, which is available in both photonic [40, 162, 188] and ultracold-atom [14, 106, 134, 211] platforms.

¹Here, the term strongly non-Gaussian refers qualitatively to (bipartite) states lying well outside the convex hull of Gaussian states.

The chapter is organized as follows. In Section 5.2, we extend the two sets of PT-moment-based PPT criteria to continuous variables, with particular emphasis on experimentally relevant third-order conditions. In Section 5.3, we derive their multicopy measurement schemes and work out passive linear-optics implementations for the second- and third-order moments. In Section 5.4, we compare the resulting criteria with standard approaches for Gaussian states, coherent superpositions and mixtures of Gaussian states, and strongly non-Gaussian states in Fock space. In Section 5.5, we investigate losses, imperfect copies, and finite statistics, and perform a simulated experiment combining these imperfections. We conclude in Section 5.6.

5.2 PT-moment-based criteria for continuous variables

We start with extending the p_n -PPT criteria reported in [80, 169, 252] to continuous variables, with an emphasis on criteria involving the first two nontrivial PT moments.

5.2.1 Infinite hierarchies of criteria

5.2.1.1 From the Stieltjes moment problem

The first set of criteria is motivated by the question of whether a given set of PT moments $(p_1, \dots, p_n)$, defined via $p_n = \text{Tr}[(\hat{\rho}^{T_B})^n]$, is compatible with a positive semidefinite partially transposed state $\hat{\rho}^{T_B}$ [169, 252]. Corresponding to a quantum version of the classical Stieltjes moment problem, compatibility can be certified upon inspecting positivity conditions formulated in terms of Hankel matrices. Given the Hankel matrix over PT moments of increasing order along rows and columns

$$P_n = \begin{pmatrix} p_1 & p_2 & p_3 & \cdots & p_{\frac{n+1}{2}} \\ p_2 & p_3 & p_4 & \cdots & \\ p_3 & p_4 & p_5 & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ p_{\frac{n+1}{2}} & & & & p_n \end{pmatrix}, \quad n = 1, 3, 5, \dots, \quad (5.1)$$

the p_n -PPT criterion states [169, 252]

$$\hat{\rho} \text{ separable} \implies \hat{\rho}^{T_B} \geq 0 \iff P_n \geq 0, \quad \forall \text{ odd } n, \quad (5.2)$$

such that violation of the latter inequality for some n flags entanglement. Note that the necessary and sufficient part for PPT when considering *all* moments is discussed in Lemma 4 from Ref. [252].

Since $p_1 = 1$ and p_2 equals the state's purity $\text{Tr}[\hat{\rho}^2]$, we need to exploit at least the third-order PT moment p_3 to flag entanglement. Sylvester's criterion implies

$$P_{n+2} \geq 0 \implies P_n \geq 0, \quad (5.3)$$

showing that higher-order criteria are stronger (and larger) sets of conditions on the positivity of $\hat{\rho}^{T_B}$. Although the set of criteria from Eq. (5.2) has been benchmarked and refined solely for discrete-variable systems, it readily extends to continuous variables since the Stieltjes-type moment problem is generally defined over non-negative numbers and, thus, does not constrain the dimension of the spectrum of the state.

5.2.1.2 From Descartes' rule of signs

The second set of criteria tests for PPT by linking the eigenvalues of $\hat{\rho}^{T_B}$, more precisely, their elementary symmetric polynomials, to PT moments via Newton's identities [169], i.e.,

$$n e_n = \sum_{j=1}^n (-1)^{j-1} e_{n-j} p_j, \quad n = 1, 2, \dots, \quad (5.4)$$

where e_n denotes the n th elementary symmetric polynomial evaluated at the eigenvalues $\{\lambda_1, \dots\}$ of $\hat{\rho}^{T_B}$, defined as

$$e_n(\lambda_1, \dots) := \sum_{1 \leq j_1 < \dots < j_n} \lambda_{j_1} \cdots \lambda_{j_n}, \quad (5.5)$$

with $e_0(\lambda_1, \dots) = e_1(\lambda_1, \dots) = 1$. Then,

$$\hat{\rho} \text{ separable} \implies e_n \geq 0, \quad \forall n \in \mathbb{N}, \quad (5.6)$$

from which conditions on the PT moments can be obtained via Eq. (5.4). Analogously to the criteria from Eq. (5.2), as a consequence of Descartes' rule of signs [169], the set of conditions of Eq. (5.6) becomes equivalent to the non-negativity of $\hat{\rho}^{T_B}$ when $n \rightarrow \infty$ (see Ref. [51]).

While the criteria from Eq. (5.6) have been formulated only for finite-dimensional Hilbert spaces, their extension to continuous variables is straightforward: Newton's identities hold equally for elementary symmetric functions, which formally generalize elementary symmetric polynomials in the limit of (countably) infinitely many eigenvalues, and so does Descartes' rule of signs when applied to the characteristic polynomial of an infinite-dimensional matrix.

5.2.2 Third-order criteria

Albeit the apparent outperformance of higher-order criteria, PT moments decrease exponentially when increasing n , thereby complicating entanglement detection in practice. This follows from $|p_n| \leq (p_2)^{n/2}$ for $n \geq 2$, which shows that higher-order PT moments are exponentially suppressed, especially if the considered state is highly mixed.

On top of that, our proposed measurement scheme for p_n requires n replica states, which further increases the experimental burden for large n (see Section 5.3).

Hence, PT-moment-based entanglement criteria limited to third order, that is, conditions involving only the purity p_2 and moment p_3 , are of particular relevance for experimental applications. They are obtained from Eqs. (5.2) and (5.6) by setting $n = 3$, leading to what we refer to as the *quadratic* [80] and *linear* [169] (in p_2) p_3 -PPT criteria,

$$\hat{\rho} \text{ separable} \implies p_3 - p_2^2 \geq 0, \quad p_3 - \frac{3p_2 - 1}{2} \geq 0, \quad (5.7)$$

respectively. Here, we have used (for the quadratic criterion) that the positivity of a 2×2 matrix with positive elements is equivalent to the positivity of its determinant, and (for the linear criterion) the relations $3e_3 = e_2 - p_2 + p_3$ and $2e_2 = 1 - p_2$. Interestingly, the two conditions in Eq. (5.7) coincide for pure states ($p_2 = 1$), giving $p_3 \geq 1$, but also for mixed states with purity $p_2 = 1/2$, giving $p_3 \geq 1/4$. In between, the quadratic criterion is stronger than the linear criterion for $p_2 < 1/2$, and vice versa when $1/2 < p_2 < 1$.

This observation suggests the existence of an *optimal* criterion of third order, in the sense that given only the two lowest-order moments p_2 and p_3 , no stronger condition on the separability of a state can be found. This is achieved by minimizing p_3 with respect to p_2 over PPT states, which results in the optimal p_3 -PPT criterion [169, 252]

$$\hat{\rho} \text{ separable} \implies p_3 - \alpha u^3 - (1 - \alpha u)^3 \geq 0, \quad (5.8)$$

where $\alpha = \lfloor 1/p_2 \rfloor$ denotes the floor function of $1/p_2$ and $u = \frac{\alpha + \sqrt{\alpha[p_2(\alpha+1)-1]}}{\alpha(\alpha+1)}$ is a non-negative number. It is instructive to evaluate Eq. (5.8) when $1/2 < p_2 \leq 1$, in which case $\alpha = 1$, $u = (1 + \sqrt{2p_2 - 1})/2$, such that Eq. (5.8) reduces to the linear criterion. Thus, the linear criterion is optimal for states of purity $1/2 \leq p_2 \leq 1$. In contrast, both the quadratic and linear criteria in Eq. (5.7) are outperformed by Eq. (5.8) for mixed states with $p_2 < 1/2$ (see Figure 2 in Ref. [169]). As already noted in Ref. [252], the optimal p_3 -PPT criterion of Eq. (5.8) is inherently independent of the Hilbert space dimension, so it also constitutes a valid entanglement criterion in the CV regime.

Remarkably, the third-order criteria are both necessary and sufficient conditions for pure-state entanglement – also in the CV case. This follows directly from the Schmidt decomposition $|\psi\rangle = \sum_i c_i |i, i\rangle$, where $c_i \in \mathbb{C}$ are the Schmidt coefficients such that $\lambda_i = |c_i|^2$ defines a probability distribution over a countably infinite set. In this case, PT moments reduce to L^p -norms, i.e., $p_n = \|\lambda\|_n^n$ ($p_n = \|\lambda\|_{n/2}^n$) for odd (even) n , which implies $p_n \leq 1$. Equality for $n \geq 3$ holds if and only if all but one Schmidt coefficients vanish. Hence, pure product states are characterized by all PT moments being unity, while for pure entangled states, we have $p_n < 1$ for $n \geq 3$.

5.3 Implementation

We now devise measurement schemes to access PT moments in CV systems. In particular, we provide passive linear optical circuits to be applied on replicas of the state, from which the two lowest-order moments, p_2 and p_3 , can be extracted using particle-number measurements.

5.3.1 Multicopy method

On discrete-variable platforms like spin lattices, the PT moments are conveniently estimated from randomized measurements [80, 169]. However, the efficiency of this ansatz – in terms of the required measurements to extract a given PT moment with high accuracy – relies *heavily* on the Hilbert space dimension of a qubit being small. More precisely, the required number of measurements to ensure ϵ accuracy for an estimate of p_n is on the order of $d^{|AB|}/\epsilon^2$ [80], where d denotes the single-spin Hilbert space dimension and $|AB|$ the total number of spins of the bipartite system. Therefore, the randomized measurement toolbox breaks down in the CV regime.

To overcome this limitation, we propose a readout scheme using the multicopy method [29] already introduced in Chapter 4 (see also [8, 79]). This technique has been successfully applied to entanglement witnessing in various bosonic systems [53, 103, 105, 135, 146], as sketched at the end of Section 4.2.3. Let us briefly recall it hereafter.

The multicopy method relies on the central observation that any homogeneous polynomial functional of degree n in $\hat{\rho}$ can be written as

$$f_n(\hat{\rho}) = \text{Tr} [\hat{O}_n \hat{\rho}^{\otimes n}], \quad (5.9)$$

where $\hat{O}_n$ is a multicopy operator acting on n independently prepared replicas of the state. For the multicopy operators considered below, their expectation value can be accessed using passive linear-optical transformations followed by particle-number measurements, as was done in Refs. [9, 104, 105, 122]. Counting particles is well established experimentally and is a standard requirement for photonic quantum computing [147, 148, 191, 218], integrated circuits [166], quantum gas microscopes [14, 106, 134, 211], and when the system under scrutiny lacks a phase-stable local oscillator, e.g., in High Harmonic Generation [88].

Given a multicopy observable $\hat{O}_n$, the remaining task is to find a passive unitary transformation U appropriate for measuring $\hat{O}_n$ by counting particles at the output,

$$(\hat{a}_1, \dots, \hat{a}_n)^T \mapsto U(\hat{a}_1, \dots, \hat{a}_n)^T, \quad (5.10)$$

which can be efficiently decomposed into beam splitters (notice that we use a different convention than in Section 3.3.1.3)

$$B_{\mu\nu}(\tau) = \begin{pmatrix} \sqrt{\tau} & \sqrt{1-\tau} \\ \sqrt{1-\tau} & -\sqrt{\tau} \end{pmatrix}, \quad (5.11)$$

of transmissivity $\tau \in [0, 1]$, and phase shifts

$$R_\mu(\varphi) = e^{-i\varphi}, \quad (5.12)$$

with phase $\varphi \in [0, 2\pi)$, acting on modes $\mu, \nu = 1, \dots, n$.

5.3.2 Measuring PT moments

The required multicopy observable for accessing the n th PT moment p_n on the n -copy state $\hat{\rho}^{\otimes n}$ reads [80]

$$\begin{aligned} p_n &= \text{Tr} \left[\hat{\rho}^{\otimes n} \left(\hat{\Pi}_n^A \otimes \hat{\Pi}_n^B \right) \right] \\ &=: \left\langle \dots \left\langle \hat{\Pi}_n^A \otimes \hat{\Pi}_n^B \right\rangle \dots \right\rangle, \end{aligned} \quad (5.13)$$

where $\hat{\Pi}_n^A$ and $\hat{\Pi}_n^B$ are normal, unitary (but not Hermitian for $n > 2$) operators corresponding to cyclic shifts, i.e., permutations, of the Fock basis elements to the right on Alice's side and to the left on Bob's side, that is,

$$\begin{aligned} \hat{\Pi}_n^A |i_1, \dots, i_n\rangle_A &= |i_n, \dots, i_{n-1}\rangle_A, \\ \hat{\Pi}_n^B |i_1, \dots, i_n\rangle_B &= |i_2, \dots, i_n, i_1\rangle_B, \end{aligned} \quad (5.14)$$

respectively. Importantly, to evaluate the PT moment p_n , it suffices that Alice and Bob measure $\hat{\Pi}_n^A$ and $\hat{\Pi}_n^B$ *separately* on their respective subsystems of the n -copy state, then evaluate the product of their outcomes (by communicating classically), and finally average over sufficiently many experimental runs. Strictly speaking, the non-Hermitian shift operators are not measured directly. Rather, as shown below, they are diagonalized by a Fourier transform, after which their eigenvalues are inferred from measurements of the Hermitian particle-number operators appearing in their spectral decomposition.

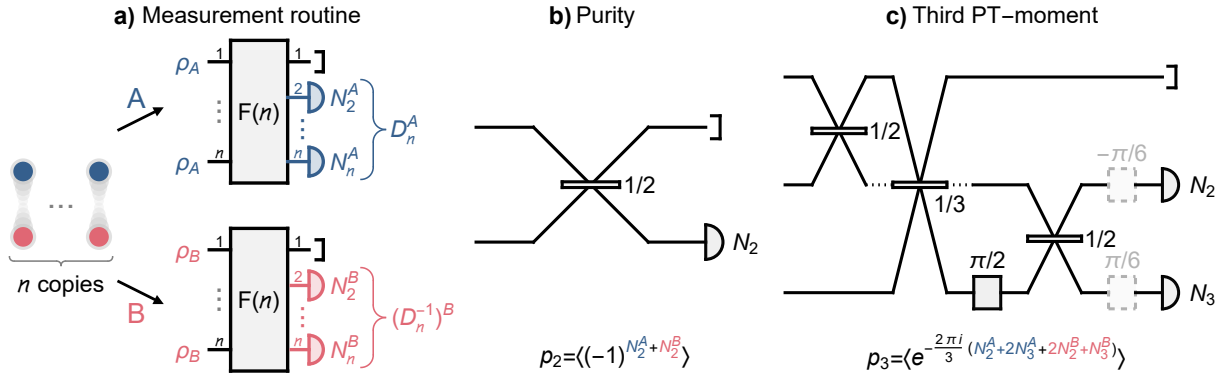


Figure 5.1: **a)** Measurement routine for detecting the n th PT moment p_n . After preparing n replicas, that is, independent and identical copies, of the bipartite state of interest, Alice and Bob separately assess their subsystems A (blue) and B (red), respectively. Both use a Fourier interferometer of order n , denoted by $F(n)$, see Eq. (5.15), and then measure the particle numbers on their output modes 2 to n . They obtain p_n from the multicopy observable $\hat{D}_n^A \otimes (\hat{D}_n^{-1})^B$, see Eq. (5.18). **b)** Readout of the purity p_2 . The second-order Fourier interferometer $F(2)$ is realized by a balanced beam splitter, upon which both Alice and Bob just measure particle numbers. **c)** Measurement of the third PT moment p_3 . The third-order Fourier interferometer $F(3)$ is implemented by a sequence of three beam splitters and three phase shifts (only the first is needed in practice), upon which incident particles are counted in the output modes 2 and 3. Here, Bob's multicopy observable corresponds to the inverse of Alice's, which is reflected in the particle numbers corresponding to the output modes 2 and 3 being interchanged for Bob.

To devise a suitable multicopy measurement scheme, let us first consider Alice's side. We introduce the discrete Fourier transform (DFT)

$$F(n) = \frac{1}{\sqrt{n}} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega_n & \cdots & \omega_n^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_n^{n-1} & \cdots & \omega_n^{(n-1)(n-1)} \end{pmatrix}, \quad (5.15)$$

with complex phases $\omega_n = e^{-2\pi i/n}$ and rank n . The matrix $F(n)$ describes the transformation of the mode operators. We denote by $\hat{U}_F^{(n)}$ the corresponding passive Fock-space unitary, defined through

$$\left(\hat{U}_F^{(n)}\right)^\dagger \hat{a}_k \hat{U}_F^{(n)} = \sum_{\ell=1}^n [F(n)]_{k\ell} \hat{a}_\ell. \quad (5.16)$$

The Fourier transformation diagonalizes the right cyclic shift,

$$\hat{\Pi}_n = \left(\hat{U}_F^{(n)}\right)^\dagger \hat{D}_n \hat{U}_F^{(n)}, \quad (5.17)$$

where

$$\hat{D}_n = \omega_n^{\sum_{j=1}^n (j-1)\hat{N}_j}. \quad (5.18)$$

The operator $\hat{D}_n$ is diagonal in the output particle-number basis. Since the coefficient associated with $\hat{N}_1$ vanishes, only the particle numbers in output modes 2 to n need to be recorded. The corresponding state transformation is

$$\hat{\rho}^{\otimes n} \mapsto \hat{U}_F^{(n)} \hat{\rho}^{\otimes n} \left(\hat{U}_F^{(n)} \right)^\dagger, \quad (5.19)$$

and therefore

$$\text{Tr} \left[\hat{\rho}^{\otimes n} \hat{\Pi}_n \right] = \text{Tr} \left[\hat{D}_n \hat{U}_F^{(n)} \hat{\rho}^{\otimes n} \left(\hat{U}_F^{(n)} \right)^\dagger \right]. \quad (5.20)$$

This procedure can easily be extended to the bipartite case, where Bob measures the left-shift operator $\hat{\Pi}_n$. Using the fact that the latter is the inverse of the right-shift operator together with the unitarity of the Fourier transform, we find

$$\hat{\Pi}_n = \left(\hat{U}_F^{(n)} \right)^\dagger \hat{D}_n^{-1} \hat{U}_F^{(n)}. \quad (5.21)$$

Therefore, for the readout of the n th PT moment, Alice and Bob need to perform the *same* n th-order Fourier interferometer on their respective sides. This results in the state transformation

$$\hat{\rho}^{\otimes n} \mapsto \left[\hat{U}_{F,A}^{(n)} \otimes \hat{U}_{F,B}^{(n)} \right] \hat{\rho}^{\otimes n} \left[\left(\hat{U}_{F,A}^{(n)} \right)^\dagger \otimes \left(\hat{U}_{F,B}^{(n)} \right)^\dagger \right] \quad (5.22)$$

and the subsequent measurements [see Figure 5.1a)]

$$p_n = \langle \cdots \langle \hat{D}_n^A \otimes (\hat{D}_n^{-1})^B \rangle \cdots \rangle. \quad (5.23)$$

We note here that this implementation is reminiscent of the method put forward in Ref. [53] to access Rényi entropies, i.e., moments of the state $\hat{\rho}$ itself. The key difference in our case – the partial transpose $\hat{\rho}^{T_B}$ – results in Bob having to measure the *inverse* of Alice's observable.

Alternatively, one may use the identity $\hat{\Pi}_n = \hat{U}_F^{(n)} \hat{D}_n \left(\hat{U}_F^{(n)} \right)^\dagger$, showing that the left shift operator can also be diagonalized by applying an *inverse* Fourier transform. In this case, Alice and Bob would have to measure precisely the same observables, but Bob would have to reverse the optical circuit Alice used.

5.3.3 Optical circuits for low-order moments

It is left to decompose the DFT of Eq. (5.15) into passive optical elements. In general, any n -mode passive linear-optical unitary can be decomposed into at most $n(n-1)/2$ beam splitters (each possibly followed by a phase shift), which stems from the decomposition of the n -dimensional unitary group $U(n)$ into $U(2)$ subgroups [37, 47, 187]. In the following, we explicitly devise optical circuits implementing F(2) and F(3), enabling the measurements of p_2 and p_3 , respectively.

5.3.3.1 Purity p_2

The protocol for obtaining the state's purity using two copies is arguably the most prominent application of the multicopy method [9, 29, 53, 104] (see Section 4.2.1) and has been successfully implemented in cold-atom systems [135, 146] and integrated photonic devices [101]. It is well known that the two-mode DFT is nothing but a 50:50 beam splitter operation, i.e.,

$$F(2) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (5.24)$$

Note that in this case, the DFT matrix is involutory; i.e., $F^{-1}(2) = F(2)$ since a shift in any direction corresponds to the swap operation, which is another way of seeing that p_2 equals the purity. Hence, a single beam splitter on each side suffices for Alice and Bob to access p_2 .

The multicopy observable after the beam splitter has been applied reads

$$p_2 = \left\langle \left\langle \omega_2^{\hat{N}_2^A} \otimes \omega_2^{\hat{N}_2^B} \right\rangle \right\rangle = \left\langle (-1)^{\hat{N}_2^A + \hat{N}_2^B} \right\rangle \quad (5.25)$$

and relates to a parity measurement of the total particle number across the second modes, specifically, the output modes displaying destructive interference. Thus, the experimental procedure consists of Alice and Bob counting the number of particles in these modes, sharing their information via classical communication, and assigning the values +1 or −1 for an even or odd *total* number of particles, respectively, and thereupon averaging over sufficiently many experimental runs [see Figure 5.1b)].

5.3.3.2 Third-order PT moment p_3

Following Eq. (5.15), the third-order DFT evaluates to

$$F(3) = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & e^{-i\frac{2\pi}{3}} & e^{i\frac{2\pi}{3}} \\ 1 & e^{i\frac{2\pi}{3}} & e^{-i\frac{2\pi}{3}} \end{pmatrix}, \quad (5.26)$$

whose implementation requires a combination of three beam splitters and three phase shifts, namely,

$$F(3) = R_3\left(\frac{\pi}{6}\right) R_2\left(-\frac{\pi}{6}\right) B_{23}\left(\frac{1}{2}\right) R_3\left(\frac{\pi}{2}\right) B_{13}\left(\frac{2}{3}\right) B_{12}\left(\frac{1}{2}\right) \quad (5.27)$$

[see Figure 5.1c)]. However, as we conduct particle-number measurements at the output, the two last phase shifts of $\pm\pi/6$ (light gray) can be left out.

Then, the third-order PT moment p_3 is encoded in the multicopy observable

$$\begin{aligned} p_3 &= \left\langle \left\langle \left\langle \omega_3^{\hat{N}_2^A + 2\hat{N}_3^A} \otimes \omega_3^{2\hat{N}_2^B + \hat{N}_3^B} \right\rangle \right\rangle \right\rangle \\ &= \left\langle \left(e^{-2\pi i/3} \right)^{\hat{N}_2^A + 2\hat{N}_3^A + 2\hat{N}_2^B + \hat{N}_3^B} \right\rangle. \end{aligned} \quad (5.28)$$

Importantly, the prefactors of Alice's number operators appearing in the exponent are now interchanged for Bob's. Although each experimental outcome is assigned a generally complex phase, their statistical average equals p_3 , which is real because $\hat{\rho}^{T_B}$ is Hermitian.

5.4 Benchmarks

Next, we benchmark the detection capabilities of the various third-order criteria discussed in Section 5.2.2. We discuss their performance in the form of a general comparison to other strong criteria, upon which we explore a variety of typical state families. This includes Gaussian states and strongly (also referred to as quantum [241]) non-Gaussian states – divided into superpositions of coherent or Fock states – including, e.g., NOON states, whose entanglement is known to be hard to certify.

5.4.1 Comparison

5.4.1.1 Shchukin–Vogel criteria

As already discussed in Section 4.2.3, the well-known Shchukin–Vogel hierarchy constitutes another complete set of criteria testing for PPT [163, 164, 209]. In contrast to PT-moment-based criteria, they involve moments of mode operators. The central idea is that the partially transposed state $\hat{\rho}^{T_B}$ is non-negative if and only if $\text{Tr}[(\hat{\rho}^{T_B}) \hat{f}^\dagger \hat{f}] \geq 0$ for all normally ordered $\hat{f}$. It is convenient to expand $\hat{f} = \sum_{i,j,k,l} c_{ijkl} \hat{a}^{\dagger i} \hat{a}^j \hat{b}^{\dagger k} \hat{b}^l$ for some complex-valued c , leading to a reformulation in terms of the matrix $D_{qrst,ijkl} = \text{Tr}[\hat{\rho} \hat{a}^{\dagger r} \hat{a}^q \hat{a}^{\dagger i} \hat{a}^j \hat{b}^{\dagger t} \hat{b}^s \hat{b}^{\dagger k} \hat{b}^l]$, namely [105],

$$\sum_{\substack{q,r,s,t \\ i,j,k,l}} c_{qrst}^* c_{ijkl} \tilde{D}_{qrst,ijkl} \geq 0, \quad (5.29)$$

for all complex c . This is, in fact, equivalent to all principal minors of $\tilde{D}$ being non-negative, as implied by Sylvester's criterion. A specific criterion is then obtained by choosing a certain $\hat{f}$ (or equivalently c), which corresponds to selecting some submatrix of $\tilde{D}$. For instance, setting $\hat{f} = c_1 + c_2 \hat{a} + c_3 \hat{b}$ leads to a second-order criterion that is necessary and sufficient for Gaussian entanglement.

It is natural to ask how Eq. (5.29) relates to PT-moment-based criteria. When considered over *all* c , the three *infinite* hierarchies from Eqs. (5.29), (5.2), and (5.6) are all equivalent to PPT. Still, their main conceptual difference is that while practically useful criteria of the Shchukin–Vogel hierarchy comprise only a few low-order mode-operator moments, PT moments can involve mode-operator moments of arbitrarily high order, depending on the state.

To make a more precise statement, we derive a relation between the linear p_3 -PPT criterion of Eq. (5.7) and Eq. (5.29). Starting from the state's normally ordered expansion $\hat{\rho} = \sum_{i,j,k,l} \rho_{ijkl} \hat{a}^{\dagger i} \hat{a}^j \hat{b}^{\dagger k} \hat{b}^l$ for complex-valued ρ , normally ordering its powers allows to write any PT moment as

$$p_n = \sum_{i,j,k,l} [\mathfrak{p}_{n-1}(\rho)]_{ijkl} \tilde{d}_{ijlk}, \quad (5.30)$$

where d denotes the matrix of mode-operator moments $d_{ijkl} = \text{Tr}[\hat{\rho} \hat{a}^{\dagger i} \hat{a}^j \hat{b}^{\dagger k} \hat{b}^l]$. Note here that $\tilde{d}_{ijlk} = d_{ijlk}$, and hence moments of the partially transposed state correspond to re-ordered moments of the state itself. Further, $\mathfrak{p}_n$'s, as defined in (5.30), are polynomials in ρ of order n that arise by repeatedly applying the bosonic commutation relations to ensure normal order in the trace. For the two lowest-order polynomials $\mathfrak{p}_1$ and $\mathfrak{p}_2$, which are required for p_2 and p_3 , respectively, we obtain

$$\begin{aligned} [\mathfrak{p}_1(x)]_{ijkl} &= x, \\ [\mathfrak{p}_2(x)]_{ijkl} &= \sum_{q,r,s,t} \sum_{u=0}^{\min(q,s)} \sum_{v=0}^{\min(r,t)} u! v! \binom{q}{u} \binom{s}{u} \binom{r}{v} \binom{t}{v} x_{i-s+u} x_{r-l-t+v} x_{s-j-q+u} x_{k-r+v} t. \end{aligned} \quad (5.31)$$

We find a similar decomposition for the Shchukin–Vogel criteria after normally ordering $\hat{f}^\dagger \hat{f}$, namely,

$$\text{Tr}[(\hat{\rho}^{T_B}) \hat{f}^\dagger \hat{f}] = \sum_{i,j,k,l} [\mathfrak{f}_2(c)]_{ijkl} \tilde{d}_{ijlk} \geq 0, \quad (5.32)$$

where $\mathfrak{f}_2$ is defined in analogy to $\mathfrak{p}_2$, with the main difference being that indices are slightly reordered since $\mathfrak{f}_2$ stems from $\hat{f}^\dagger \hat{f}$, whereas $\mathfrak{p}_2$ stems from $(\hat{\rho}^{T_B})^3$. In general, the map $c \mapsto \mathfrak{f}_2(c)$ is not surjective for complex-valued c . However, when identifying linear combinations of Eq. (5.30) with Eq. (5.32), $\mathfrak{f}_2(c)$ must be Hermitian, in which case it can always be expressed as a bilinear form of some complex-valued c .

Therefore, the linear p_3 -PPT criterion of Eq. (5.7) equals the Shchukin–Vogel criterion of Eq. (5.32), with the matrix c being defined implicitly via $[\mathfrak{f}_2(c)]_{ijkl} = [\mathfrak{p}_2(\rho)]_{ijkl} - (3/2)\rho_{ijkl} + (1/2)\delta_{i0}\delta_{j0}\delta_{k0}\delta_{l0}$. This shows that the set of mode-operator moments constrained by a PT-moment-based criterion depends solely on the state and can, in principle, range up to arbitrarily high orders – a crucial property for capturing higher-order correlations (we provide a worked-out example for NOON states in Appendix A.4.1). Given that the efficient readout of low-order Shchukin–Vogel criteria also requires multiple copies [105], the herein proposed scheme is expected to outrange the Shchukin–Vogel method for strongly non-Gaussian states in terms of both detection capability and experimental resources (see, e.g., Section 5.4.4.1).

5.4.1.2 Criteria based on non-local variables

In another common approach, PPT is assessed via the non-local quadrature operators $\hat{x}_\pm = \hat{x}_A \pm \hat{x}_B$ and $\hat{p}_\pm = \hat{p}_A \pm \hat{p}_B$ from the EPR argument [76]. By noting that taking the partial transpose in subsystem B translates into $p_B \rightarrow -p_B$ in phase space [216], physicality (meaning non-negativity) of $\hat{\rho}^{T_B}$ implies that uncertainty relations must be fulfilled by the mixed commuting variable pairs $(\hat{x}_\pm, \hat{p}_\mp)$ for all separable states. More generally, uncertainty relations provide another way to construct separability criteria and entanglement witnesses [57].

Many entanglement criteria have been formulated in terms of the outcome distributions of a homodyne measurement $f_\pm(x_\pm) = \langle x_\pm | \hat{\rho} | x_\pm \rangle$, $g_\mp(p_\mp) = \langle p_\mp | \hat{\rho} | p_\mp \rangle$. This includes constraints on their second moments, such as $\sigma_{x_\pm}^2 + \sigma_{p_\mp}^2 \geq 2$ [72] (see Theorem 2.7, and also [99, 151, 160, 216, 244]), and entropies, like $h(f_\pm) + h(g_\mp) \geq \log(2\pi e)$ [239] (see Theorem 2.10, and also [195, 200, 228, 238]). Recently, entanglement criteria were introduced for the Husimi-Q distribution $Q_\pm(x_\pm, p_\mp)$ associated with a heterodyne measurement, ranging from entropic variants $h(Q_\pm) \geq \log(2e)$ [84] (see also [85]) to majorization relations $Q_\pm < Q_{\pm, \text{vacuum}}$ [92, 93].

While all these methods fully characterize Gaussian entanglement, the entropic criteria also witness some forms of non-Gaussian entanglement, such as superpositions of Gaussian states, e.g., the entangled cat state (see Section 5.4.3.1) and close-to-Gaussian states, e.g., the Hermite-Gauss state [see Eq. (5.50)]. However, entanglement cannot be certified over the whole parameter range in both cases. Additionally, strongly non-Gaussian entanglement arising from finite superpositions in Fock space is entirely overlooked. Thus, such methods are efficient when dealing with Gaussian and Gaussian-like entanglement, but remain insensitive beyond.

Intuitively, this is rooted in how (quantum) correlations are analyzed: The non-local variables $(x_\pm, p_\mp)$ are linear and hence capture mostly linear correlations between the two subsystems, while lacking information about higher-order correlations. Therefore, we expect our method to outrange all criteria based on EPR-type variables in the strongly non-Gaussian regime (see, e.g., Section 5.4.4).

5.4.2 Gaussian entanglement

5.4.2.1 Symplectic formalism and Simon's criterion

It is convenient to group the two pairs of canonical operators into a single vector in phase space as

$$\hat{\chi} = (\hat{x}_1, \hat{p}_1, \hat{x}_2, \hat{p}_2)^T. \quad (5.33)$$

Then, the canonical commutation relations take the form $[\hat{\chi}_j, \hat{\chi}_{j'}] = i\Omega_{jj'}\hat{1}$, where Ω denotes the symplectic metric. We denote with the “ $\sim$ ” symbol partially transposed-state quantities.

A partially transposed Gaussian state is fully characterized by its mean vector and covariance matrix

$$\tilde{\chi} = \text{Tr}[(\hat{\rho}^{T_B}) \hat{\chi}], \quad \tilde{\gamma}_{jj'} = \text{Tr}[(\hat{\rho}^{T_B}) \{\hat{\chi}_j, \hat{\chi}_{j'}\}] - 2\tilde{\chi}_j \tilde{\chi}_{j'}, \quad (5.34)$$

respectively. While the former is irrelevant for separability, the latter is non-negative for all states but fulfills the stronger condition $\tilde{\gamma} + i\Omega \geq 0$ if and only if $\hat{\rho}$ is Gaussian and separable, which is known as Simon's criterion [216], see Theorem 2.9 (other well-known criteria have been introduced by Duan et al. [72], see Theorem 2.7, and Mancini et al. [99, 160], see Theorem 2.8). It can be recast into a simple form in terms of the so-called symplectic eigenvalues of $\tilde{\gamma}$. To this end, we employ Williamson's theorem [246] introduced in Section 3.3.1.1, which ensures the existence of a symplectic matrix $\mathcal{S}$ that diagonalizes the covariance matrix such that $\mathcal{S} \tilde{\gamma} \mathcal{S}^T = \text{diag}(\tilde{\nu}_1, \tilde{\nu}_1, \tilde{\nu}_2, \tilde{\nu}_2)$, where $\tilde{\nu}_1$ and $\tilde{\nu}_2$ denote the symplectic eigenvalues. Hence, a Gaussian state is separable if and only if

$$\tilde{\nu}_1, \tilde{\nu}_2 \geq 1 \quad (5.35)$$

[see the straight black curves in Figures 5.2a) and 5.2b)].

5.4.2.2 p_n -PPT criteria for Gaussian states

A closed formula for the n th moment of a Gaussian state in terms of its symplectic eigenvalues has been derived in [204, 206]. For the partially transposed state, it reads

$$p_n = \prod_{j=1}^2 \frac{2^n}{(\tilde{\nu}_j + 1)^n - (\tilde{\nu}_j - 1)^n}. \quad (5.36)$$

For $n = 2$ and $n = 3$, the formula evaluates to

$$p_2 = \frac{1}{\tilde{\nu}_1 \tilde{\nu}_2} = \frac{1}{\sqrt{\det \tilde{\gamma}}}, \quad p_3 = \frac{16}{(1 + 3\tilde{\nu}_1^2)(1 + 3\tilde{\nu}_2^2)}, \quad (5.37)$$

respectively. Hence, a Gaussian state is pure if and only if the product of symplectic eigenvalues equals unity, i.e., $\tilde{\nu}_1 \tilde{\nu}_2 = 1$. This implies that all states have to respect the physicality condition $\tilde{\nu}_1 \tilde{\nu}_2 \geq 1$ (see dashed black curves) Notice that $\tilde{\nu}_1 \tilde{\nu}_2 = \nu_1 \nu_2$ is ensured by symplectic matrices having unit determinant.

With Eq. (5.37) at hand, we can express the p_3 -PPT criteria in terms of the symplectic eigenvalues $(\tilde{\nu}_1, \tilde{\nu}_2)$ only. The quadratic p_3 -PPT criterion reads (blue curves)

$$7\tilde{\nu}_1^2 \tilde{\nu}_2^2 - 3(\tilde{\nu}_1^2 + \tilde{\nu}_2^2) - 1 \geq 0, \quad (5.38)$$

while the linear criterion evaluates to (green curves)

$$(\tilde{\nu}_1^2 + 3\tilde{\nu}_1^2 \tilde{\nu}_2^2 + \tilde{\nu}_2^2)(\tilde{\nu}_1 \tilde{\nu}_2 - 3) + 11\tilde{\nu}_1 \tilde{\nu}_2 - 1 \geq 0. \quad (5.39)$$

It is straightforward to check that both conditions and the optimal criterion (red curves) are implied by Simon's criterion of Eq. (5.35), showing that all three are necessary but not sufficient conditions for Gaussian separability. Sufficiency is retained in the limit $n \rightarrow \infty$, see Appendix A.1.2, in accordance with PPT being equivalent to separability for 1×1 Gaussian states [244].

5.4.2.3 Optimal PT-moment-based criterion for Gaussian entanglement

When expressing Simon's criterion of Eq. (5.35) through the two lowest-order PT moments p_2 and p_3 , we find that a Gaussian state is separable if and only if

$$p_3 \geq \frac{4p_2^2}{3 + p_2^2}, \quad (5.40)$$

which is the optimal p_3 -PPT criterion for Gaussian entanglement (see straight black curves). However, note that this criterion is biased in the sense that it only holds under the assumption that the underlying state is of Gaussian form. Translating also the physicality condition $\tilde{v}_1 \tilde{v}_2 \geq 1$ to PT moments by using $\tilde{v}_2 = 1/(\tilde{v}_1 p_2)$ for p_3 and maximizing over $\tilde{v}_1$ yields an upper bound constraining all Gaussian states (see dashed black curves)

$$p_3 \leq \left(\frac{4p_2}{3 + p_2} \right)^2. \quad (5.41)$$

Of course, one might argue that requiring copies of an entangled Gaussian state to witness its entanglement is unnecessarily demanding; however, the method becomes powerful in the non-Gaussian regime.

5.4.3 Strongly non-Gaussian entanglement from coherent states

5.4.3.1 Mixed entangled cat states

Non-Gaussianity naturally arises from mixtures of Gaussian states, while coherent superpositions are required to generate strongly non-Gaussian states, for which Schrödinger cat states display an important example [241, 253]. In the following, we consider the dephased [even (+) or odd (−)] entangled Schrödinger cat states

$$\hat{\rho} = \frac{1}{C} \left[|\alpha, \beta\rangle \langle \alpha, \beta| + |-\alpha, -\beta\rangle \langle -\alpha, -\beta| \pm (1 - z) (|\alpha, \beta\rangle \langle -\alpha, -\beta| + |-\alpha, -\beta\rangle \langle \alpha, \beta|) \right], \quad (5.42)$$

with normalization $C = 2[1 \pm (1 - z)e^{-2(|\alpha|^2 + |\beta|^2)}]$, displacements $\alpha, \beta \in \mathbb{C}$, and mixing (dephasing) parameter² $z \in [0, 1]$.

²Centering both cats ($\alpha = \beta = 0$), i.e., reducing the state to the vacuum, requires $z > 0$ for a well-defined state.

The state's purity decreases monotonically³ for increasing z , with $p_2 = 1$ if and only if $z = 0$, and $p_2 = 1/2$ for $z = 1$ and $|\alpha|, |\beta| \rightarrow \infty$. When $z = 1$, $\alpha = 0$, or $\beta = 0$, the state of Eq. (5.42) is separable, but entanglement is challenging to reveal for odd cat states: Second-moment-based criteria fail to witness entanglement, while entropic criteria flag entanglement only if the two cats are well separated, i.e., beyond $|\alpha| = |\beta| \gtrsim 3/2$ [84, 195, 239].

We provide expressions for the second and third PT moments in Appendix A.2. Since $p_2 \geq 1/2$ over the full parameter range, the linear criterion is optimal. Entanglement is always witnessed for even cat states (as in Ref. [95]). For odd cat states, we obtain the separability condition

$$|\alpha|^2 + |\beta|^2 \leq -\frac{1}{2} \ln(1 - z) \quad (5.43)$$

[see Figure 5.2c)]. For given z , entanglement is certified in the complement of a disk of radius $\sqrt{-(1/2) \ln(1 - z)}$ in the $(|\alpha|, |\beta|)$ -plane. For $|\alpha| = |\beta|$, this amounts to $|\alpha| \geq \sqrt{-\ln(1 - z)}/2$ (see inset), which is smaller than $3/2$ up to $z = 1 - e^{-9} \approx 0.99988$. Hence, all entropic criteria are outperformed by the p_3 -PPT criteria. Note also that entanglement is certified for all $|\alpha|, |\beta| > 0$ when the state is pure ($z = 0$).

5.4.3.2 High Harmonic Generation (HHG)

Assessing entanglement by counting photons is particularly relevant for applications where other standard readout methods are inapplicable. A prime example of recent interest is the process of High Harmonic Generation (HHG), in which intense laser fields are directed at atomic media to generate high-order harmonics of the driving field [86]. In the strongly interacting regime, i.e., when the generated light lies in the extreme ultraviolet (XUV) spectrum, a phase-stable reference beam is often lacking, thereby preventing quadrature measurements via both the homodyne and heterodyne detection protocols [88].

In a recently proposed multimode quantum description of HHG, the initial coherent-state driving field α (mode 1) is depleted by $\delta\alpha > 0$, which causes displacements χ_j in the $N - 1$ harmonic modes $j = 2, \dots, N$, leading to entanglement between the driving laser and the higher harmonics, as well as among the harmonics themselves [222]. The corresponding global state reads

$$|\psi\rangle = \frac{1}{\sqrt{C}} \left(|\alpha - \delta\alpha\rangle \otimes_{j=2}^N |\chi_j\rangle - \langle\alpha|\alpha - \delta\alpha\rangle |\alpha\rangle \otimes_{j=2}^N \langle 0_j|\chi_j\rangle |0_j\rangle \right), \quad (5.44)$$

with normalization $C = 1 - e^{-|\delta\alpha|^2 - \sum_{j=2}^N |\chi_j|^2}$.

In the following, we consider all harmonic modes to be displaced equally, i.e., $\chi = \chi_j$, in which case $|\chi| \propto |\delta\alpha|$ by energy conservation and only odd harmonics are populated.

³The only exception being when $|\alpha|, |\beta| \lesssim 1$ for the odd cat state.

We analyze entanglement across the bipartition of the driving field and one harmonic described by the mixed state

$$\hat{\rho} = \text{Tr}_{j>2} [|\psi\rangle\langle\psi|]; \quad (5.45)$$

see Appendix A.3 for an explicit expression and the corresponding PT moments.

Entanglement is detected by the linear p_3 -PPT criterion (optimal because $p_2 \geq 1/2$ in the considered regime), also for many harmonic modes [see Figure 5.2d)]. Note that the purity p_2 is monotonically increasing with $\delta\alpha$, and that $p_2 = 1$ for $N = 1$, while $p_2 \gtrsim 4/5$ for all $N > 1$. The amount of entanglement behaves very similarly when considered as a function of $\delta\alpha$ with a pure product state being attained in the limit $\delta\alpha \rightarrow \infty$, see also Figure 3 in Ref. [222], and hence entanglement is harder to certify for more pure states. Further, we observe a decrease in absolute violation for increasing N , as the entanglement is split among more modes.

It was proposed in Ref. [222] that entanglement can be enhanced by using conditioning measurements on complementary modes instead of the partial trace. In this case, the remaining bipartition is always pure, and thus entanglement is fully witnessed by any p_3 -PPT criterion.

5.4.4 Strongly non-Gaussian entanglement from Fock states

5.4.4.1 NOON states

Entanglement detection is hardest when considering finite superpositions in Fock space. Of particular importance is the class of bosonic NOON states [196]

$$|\psi\rangle = \alpha |N, 0\rangle + \beta |0, N\rangle, \quad (5.46)$$

with mode population $N \in \mathbb{N}^+$ and phases $\alpha, \beta \in \mathbb{C}$ constrained by $|\alpha|^2 + |\beta|^2 = 1$ due to state normalization. This class comprises the first two Bell states for $N = 1$ and $\alpha = 1/\sqrt{2}, \beta = \pm 1/\sqrt{2}$, and the Hong-Ou-Mandel state for $N = 2$ and $\alpha = -\beta = 1/\sqrt{2}$. States of the form of Eq. (5.46) have successfully been prepared deterministically for $N \leq 2$, e.g., using Hong-Ou-Mandel interferometry. In contrast, higher-order NOON states with $N > 2$ were generated probabilistically by additionally invoking post-selection (see Refs. [3, 136, 165, 168, 189, 219, 242]). Although NOON states are evidently entangled whenever neither of the complex phases is zero, detecting their entanglement is notoriously difficult – both theoretically and experimentally. Non-local EPR-type criteria – whether based on second moments or entropies – never flag entanglement [72, 84, 85, 92, 93, 99, 160, 195, 200, 228, 238, 239], while mode-operator-based criteria require moments up to order $2N$ [35], for which no practical readouts exist when N becomes large. For instance, a recent criterion formulated for a spin-like multicopy observable encoding fourth-order moments witnesses entanglement only up to $N = 2$ [105].

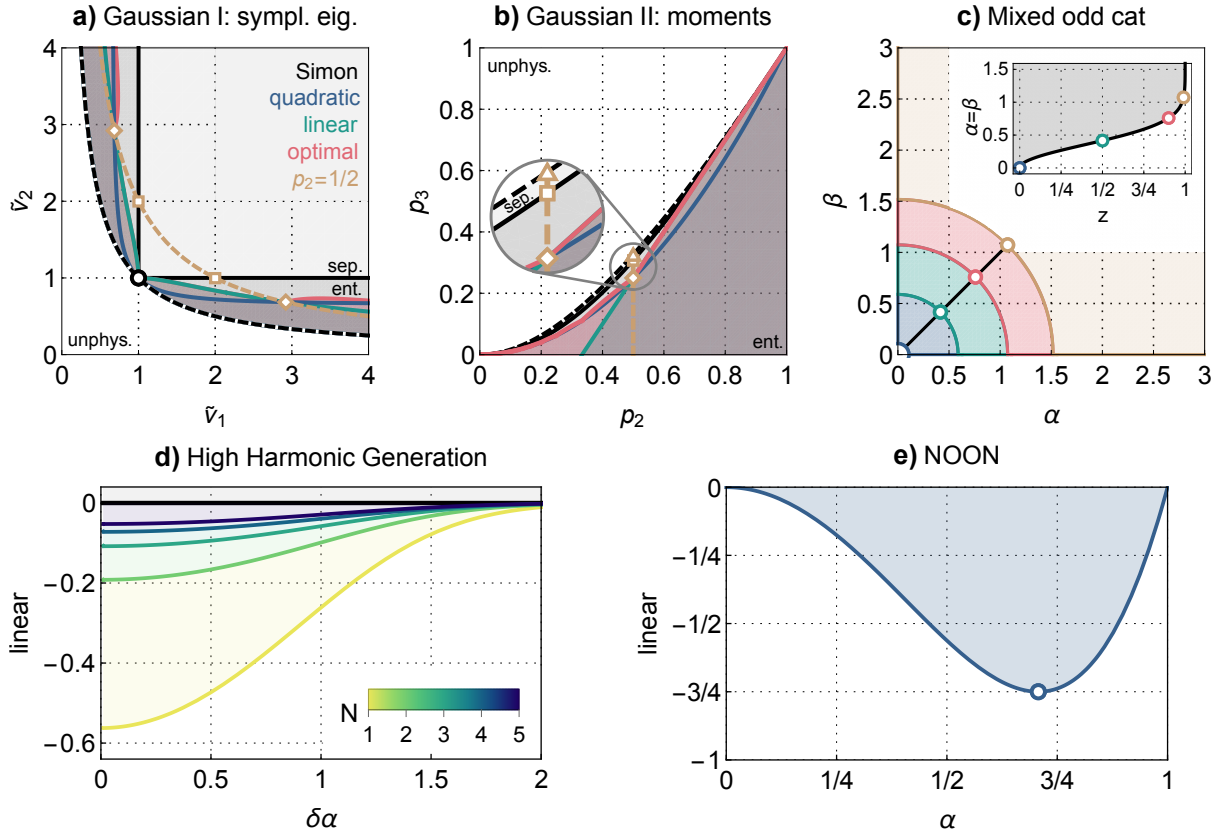


Figure 5.2: **a)** Detected Gaussian entanglement by PT-moment-based criteria of order 3 (shaded regions) in terms of the symplectic eigenvalues $\tilde{v}_1, \tilde{v}_2$. The set of physical Gaussian states constrained by $\tilde{v}_1 \tilde{v}_2 \geq 1$ (above and to the right of the black dashed line) is divided by Simon's criterion (5.35) (black straight line) into separable ($\tilde{v}_1 \geq 1$ and $\tilde{v}_2 \geq 1$) and entangled ($\tilde{v}_1 < 1$ or $\tilde{v}_2 < 1$) regions. The performance across the criteria is governed by the purity p_2 , with the tipping point at $p_2 = 1/2$ (yellow dashed line). This purity is realized, for instance, by a two-mode squeezed state at finite temperature, see Appendix A.1.1, with $\bar{n} = (\sqrt{2} - 1)/2$ particles per mode. No third-order criterion detects entanglement when $v_1 \lesssim 1, v_2 \gg 1$ (or vice versa). **b)** Complementary analysis in terms of the PT moments p_2 and p_3 . **c)** Certified entanglement for the odd Schrödinger cat state (5.42) by the optimal (= linear since $p_2 \geq 1/2$) p_3 -PPT criterion for various mixing parameters $z = 0, 0.5, 0.9, 0.99$ (blue, green, red, and yellow, respectively) and (without loss of generality real) phases α, β . Entanglement is flagged well below $\alpha = \beta \approx 3/2$ (see inset), thereby outperforming all entropic methods. **d)** Absolute violation of the linear p_3 -PPT criterion for the entanglement between the driving field and the first harmonic mode in High Harmonic Generation (HHG), see Eq. (5.45), as a function of the (without loss of generality real) depletion $\delta\alpha$ of the laser field for various total numbers of excited harmonics N . Entanglement is witnessed over the full parameter range. **e)** The same as in panel d) for the family of NOON states (5.46) as a function of the (without loss of generality real) phase α . Entanglement is detected for all mode numbers N and $\alpha \neq 0, 1$, with the strongest violation $-3/4$ occurring at $\alpha = 1/\sqrt{2} \approx 0.707$, independent of N . All other standard methods are significantly outperformed.

Since NOON states are pure ($p_2 = 1$), their entanglement is certified by all third-order criteria. More precisely, after a straightforward calculation, we find for p_3

$$\begin{aligned} p_3 &= \text{Tr} \left[|\alpha|^6 |N, 0\rangle \langle N, 0| + \alpha \beta^* |\alpha|^2 |\beta|^2 |N, N\rangle \langle 0, 0| \right. \\ &\quad \left. + \alpha^* \beta |\alpha|^2 |\beta|^2 |0, 0\rangle \langle N, N| + |\beta|^6 |0, N\rangle \langle 0, N| \right] \\ &= |\alpha|^6 + |\beta|^6. \end{aligned} \quad (5.47)$$

Remarkably, this result is independent of the mode population N (which generalizes to all PT moments) and p_3 is strictly less than unity if and only if α or $\beta \neq 0$ [see Figure 5.2e]. This demonstrates that PT-moment-based criteria significantly outperform all other approaches in the strongly non-Gaussian regime (see Appendix A.4.1 for an in-depth analysis of which mode-operator-moments appear in the PT moments). These findings also persist under realistic experimental constraints, which effectively render the underlying state mixed (see Section 5.5.2).

5.4.4.2 Other pure-state families

In the following, we list a few important pure entangled states that belong to the class of strongly non-Gaussian states, which are all automatically witnessed by any p_3 -PPT criterion. Some of them are typical benchmarks for entanglement criteria, while others are of particular experimental relevance.

- Entangled Fock-state superpositions

$$|\psi\rangle = \sum_{j,j'=0}^{N,N'} c_{jj'} |j, j'\rangle, \quad (5.48)$$

which includes, e.g., the photonic qutrit state [190, 239] $|\psi\rangle = (\sqrt{2}|0, 0\rangle + |2, 0\rangle + |0, 2\rangle)/2$, whose entanglement remains undetected by entropic methods [84, 85, 92, 93, 195, 200, 228, 238, 239].

- Photon-added and photon-subtracted states, which are obtained by adding or subtracting photons from a pure Gaussian state, i.e.,

$$|\psi_+\rangle \propto (\hat{a}^\dagger)^j |\psi_{\text{Gauss}}\rangle, \quad |\psi_-\rangle \propto (\hat{a})^j |\psi_{\text{Gauss}}\rangle, \quad (5.49)$$

respectively [241]. Both processes enhance the entanglement in the initial Gaussian state. Therefore, entanglement is typically witnessed also by second-moment criteria as long as j remains small [72, 99, 151, 160, 216, 244].

- The Hermite-Gauss state described by the wavefunction [190, 239]

$$\psi(x_1, x_2) = \frac{x_1 + x_2}{\sqrt{\pi\sigma - \sigma_+^3}} e^{-\left(\frac{(x_1+x_2)^2}{4\sigma_+^2} + \frac{(x_1-x_2)^2}{4\sigma_-^2}\right)}, \quad (5.50)$$

whose generation involves the non-Gaussian input $|0, 1\rangle$ fed into a Gaussian circuit [121]. No entropic criterion detects entanglement over the full parameter range, i.e., when $\sigma_+, \sigma_- > 0$ [84, 85, 92, 93, 195, 200, 228, 238, 239].

5.5 Application

Robustness to experimental imperfections is an essential requirement of any practically useful entanglement criterion. First, we discuss typical sources of errors encountered in our detection schemes and how to model them. Then, we explore their influence in detail for NOON states.

5.5.1 Imperfections

We are interested in the three main types of imperfections: the inherent losses that occur in state preparation, the optical circuit and detection; the noise-induced discrepancies between different state copies; and the limitations of finite measurement statistics.

5.5.1.1 Losses

Particle losses are inevitable in the preparation process of a given input state $\hat{\rho}$, the optical circuit, and the detection of single particles. With advances in superconducting photon-number-resolving detectors, efficiencies in photonic setups are now on the order of $\sim 95\%$ [162, 188] and single photons can be resolved up to 100 photons [40] (see also [218] and references therein). In ultracold atom experiments, quantum gas microscopes enable single-particle resolution up to ~ 1000 particles [14, 134, 211] (see [106] for a review). Therefore, losses are most detrimental during state preparation and optical processing.

The effect of losses is conveniently characterized by a lossy quantum channel $\mathcal{L}_\tau$, where τ is the loss parameter, being unity in the lossless case [243]. Since losses can primarily be attributed to the optical circuit, we assume equal losses for both subsystems A and B , as well as among the copies of $\hat{\rho}$ (on average, see below), which results in the input-state modification

$$\hat{\rho} \rightarrow (\mathcal{L}_\tau \otimes \mathcal{L}_\tau)(\hat{\rho}). \quad (5.51)$$

This quantum channel can be represented via its Stinespring dilation as the interference of the state of interest with a vacuum state $|0\rangle\langle 0|_E$ on a beam splitter with transmissivity $\tau \in [0, 1]$, after which the environmental mode E is traced out, that is,

$$\mathcal{L}_\tau(\hat{\rho}_A) = \text{Tr}_E \left[\hat{B}_{AE}(\tau) (\hat{\rho}_A \otimes |0\rangle\langle 0|_E) \hat{B}_{AE}^\dagger(\tau) \right], \quad (5.52)$$

for subsystem A , and analogously for B .

5.5.1.2 Noisy copies

In real-world experiments, input-state and circuit parameters can vary from one experimental realization to another. The resulting variations among the state copies and subsequent experimental runs are commonly referred to as noise. We account for both effects by drawing any continuous parameter x , such as the complex phase α of the input state or the loss parameter τ , from a narrow Gaussian distribution with standard deviation $\delta x \leq x/10$ centered around the corresponding expected value x .

We note that noise is sometimes accounted for by replacing the environmental vacuum state in Eq. (5.52) with a thermal state, where the mean particle number acts as the noise parameter, corresponding to a noisy and lossy channel [243]. Although we expect the resulting effects to be very similar, this modeling presents two practical drawbacks. First, mixing the input with a thermal state leads to (exponentially suppressed) excitations of arbitrarily high Fock states. Therefore, any reasonable numerical simulation of our method would require truncating the Fock spectrum above a specific threshold. Second, while this approach handles the noise during state preparation and within the optical circuit, it does not take into account fluctuations between different experimental realizations.

5.5.1.3 Finite statistics

In every experiment, an observable of interest is ultimately estimated from a finite number of repetitions of the experiment, which can greatly influence the statistical significance of the expected outcome. In the context of entanglement detection, where the standard smoking-gun signal is the negativity of some witness function, finite statistics can significantly deteriorate detection capabilities by resulting in small confidence intervals.

We analyze the statistical behavior of the PT moments and the third-order criteria in terms of the means and variances of their estimators when sampled from k independent and identically distributed (i.i.d.) experimental runs.

We define the estimate of p_n based on k samples as

$$p_n^{(k)} := \frac{1}{k} \sum_{i=1}^k p_{n,(i)}, \quad (5.53)$$

where $p_{n,(i)} = \omega_n^{\sum_{j=1}^n (j-1) [N_{j,(i)}^A - N_{j,(i)}^B]}$, with $N_{j,(i)}^A$ representing the number of particles in Alice's j th mode during experiment i (analogously for Bob) and $p_{n,(i)}$ being an i.i.d. complex random variable.

By linearity of the expectation value, we have

$$\mathbb{E} [p_n^{(k)}] = \mathbb{E} [p_{n,(i)}] = p_n. \quad (5.54)$$

Further, $\text{Var} [p_{n,(i)}] = \mathbb{E} [|p_{n,(i)}|^2] - |\mathbb{E} [p_{n,(i)}]|^2 = 1 - p_n^2$, where $|p_{n,(i)}| = 1$ since $p_{n,(i)}$ is one of the n roots of unity. Then, the variance of $p_n^{(k)}$ is simply

$$\text{Var} [p_n^{(k)}] = \frac{1 - p_n^2}{k}. \quad (5.55)$$

Upon defining the estimators for the linear and quadratic criteria given in Eq. (5.7) as

$$W_l^{(k)} := p_3^{(k)} - \frac{3p_2^{(k)} - 1}{2}, \quad W_q^{(k)} := p_3^{(k)} - \frac{k|p_2^{(k)}|^2 - 1}{k-1}, \quad (5.56)$$

which ensures unbiased estimators

$$\mathbb{E}[W_l^{(k)}] = p_3 - \frac{3p_2 - 1}{2}, \quad \mathbb{E}[W_q^{(k)}] = p_3 - p_2^2, \quad (5.57)$$

we obtain the variances

$$\text{Var} [W_l^{(k)}] = \frac{1 - p_3^2}{k} + \frac{9}{4} \frac{1 - p_2^2}{k}, \quad \text{Var} [W_q^{(k)}] = \frac{1 - p_3^2}{k} + \frac{2(1 - p_2^2)[1 + (2k-3)p_2^2]}{k(k-1)}. \quad (5.58)$$

5.5.2 NOON states – a case study

5.5.2.1 Analysis of individual errors

We analyze the influence of all considered experimental imperfections on the prospects of witnessing higher-order NOON-state entanglement with our proposed method. To this end, we first incorporate losses following Eq. (5.52), which renders the underlying state mixed, thereby impeding entanglement detection. Expressions for the resulting state and its PT moments can be found in Appendix A.4.2.

We show the absolute violation of the optimal p_3 -PPT criterion as a function of the loss parameter τ for balanced ($\alpha = 1/\sqrt{2}$) NOON states up to $N = 10$ in Figure 5.3a). As N increases, fewer losses are tolerable for the state's entanglement to remain certified (see colored rings for the zero crossings). However, even for high occupations ($N = 10$), our method detects entanglement up to $\sim 20\%$ loss (note that entanglement is never certified when losses exceed 50% – regardless of N – although the states remain entangled).

To investigate noise effects (and to be able to sample the PT moments), we have to calculate the photon-number distributions at the outputs of the two circuits required for measuring p_2 and p_3 , see Figures 5.1b) and 5.1c), respectively – with generally different state and loss parameters for both subsystems and all copies. Note that this task resembles boson sampling [1], which becomes inefficient on a classical computer rapidly with increasing N . Therefore, we restrict to $N = 1$ for the analysis of noise and also for a full simulation of the experiment. We provide explicit expressions for the case where all parameters are equal and show the resulting output distributions for various loss parameters τ in Appendix A.4.3.

We depict the impact of imperfect state copies on entanglement detection for $N = 1$ in Figure 5.3b). In both panels, the first copy remains unaltered ($\alpha_1 = 1/\sqrt{2}$ and $\tau_1 = 0.9, 0.75, 0.6$ corresponding to the blue, red, and green regions, respectively), while the second and third copies are described by arbitrary (α_2, τ_2) and (α_3, τ_3) , respectively. We examine noisy phases in the left panel and noisy losses in the right panel. In all cases, entanglement is certified substantially beyond the perfect-copy scenario (see points). We observe a maximum of $\approx 30\%$ allowed variations for $\tau = 0.6$ (green). This marks the minimum tolerance among the considered cases, thereby demonstrating strong robustness against noisy state copies.

At last, we study statistical effects. Using Eq. (5.58), we compute the least amount of samples needed to witness entanglement with the quadratic criterion within one standard deviation [see Figure 5.3c)]. For the Bell state ($N = 1$), only ≈ 10 samples are required at 10% losses (blue point), while $\approx 10^3$ samples are still sufficient in the very lossy regime $\tau = 0.6$ (green point). Remarkably, our method certifies higher-order NOON entanglement up to $N = 10$ at $\approx 15\%$ losses for a reasonable measurement budget of 10^5 experimental runs, showcasing its efficiency also in the strongly non-Gaussian regime.

5.5.2.2 Full simulation

To complete the picture, we combine all sources of imperfections by sampling from the photon-number-output distribution of the lossy and noisy Bell state ($N = 1, \alpha = 1/\sqrt{2}$). For each individual copy and each experimental run, we draw the phases α and the losses τ from Gaussian distributions centered around $\alpha = 1/\sqrt{2}$ and $\tau = 0.9, 0.75, 0.6$ (blue, red, green), respectively, and with standard deviations 5% and 0.05, respectively, to represent the overall noisiness of a real-world experiment. Then, we compute the optimal p_3 -PPT criterion for a given number of samples, which we repeat 500 times to gather statistics on the mean and its error.

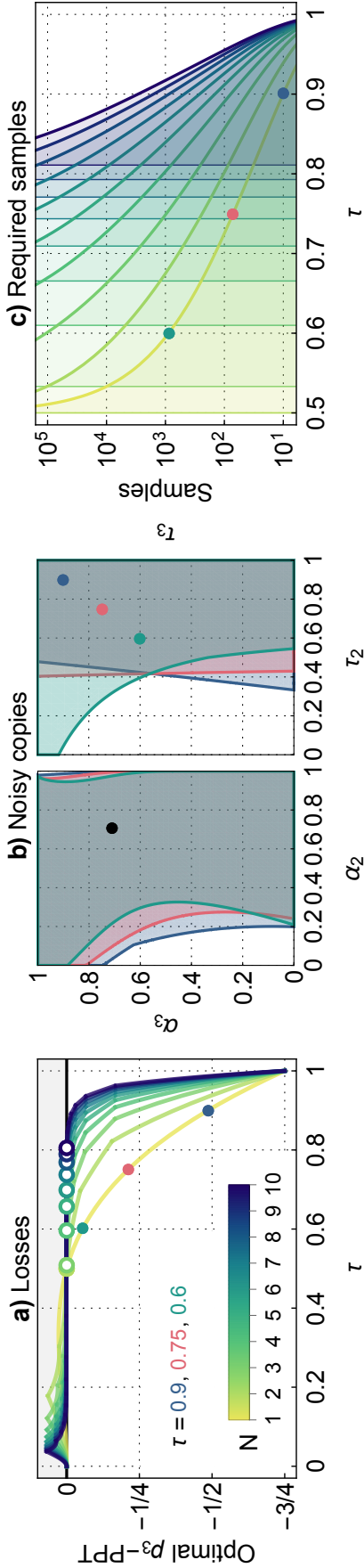


Figure 5.3: The three types of imperfections discussed in Section 5.5.1 applied to balanced ($\alpha = 1/\sqrt{2}$) NOON states. **a)** Optimal p_3 -PPT criterion as a function of the loss parameter τ of the lossy channel $\mathcal{L}_\tau$. From yellow to blue, the NOON state's mode population number N ranges from 1 to 10. Colored rings are placed at points where the criterion first flags entanglement, with a tolerance of up to 20% losses until $N = 10$ (note that $\tau > 1/2$ is required for all N). The colored dots at $N = 1$ and $\tau = 0.9$ (blue), $\tau = 0.75$ (red), and $\tau = 0.6$ (green) allow for an intercomparison of typical loss values between the subfigures (other imperfections) and Figure 5.4 (full simulation). **b)** Regions of detected entanglement by the optimal p_3 -PPT criterion for noisy copies of the first NOON state and the three typical loss parameters $\tau = 0.9, 0.75, 0.6$. The first copy serves as a reference with the standard values for $\alpha_1 := \alpha$ and $\tau_1 := \tau$; the second copy is described by α_2 and τ_2 (for both p_2 and p_3), while α_3 and τ_3 parametrize the third copy used solely in the circuit for p_3 . We illustrate the dependence on α in the left panel (with the black dot marking all α 's being $1/\sqrt{2}$). Entanglement is also detected for large discrepancies, except when α_2 approaches zero, i.e., when the second copy becomes separable. The right panel depicts the dependence on τ , where entanglement is witnessed whenever $\tau_2 \gtrsim 1/2$. **c)** Number of samples required to detect entanglement via the quadratic criterion within one standard deviation for given loss τ and various mode occupations N (vertical lines indicate minimal τ for infinite statistics). Even for 15% losses, entanglement is still witnessed up to $N = 10$ with 10^5 experimental runs only.

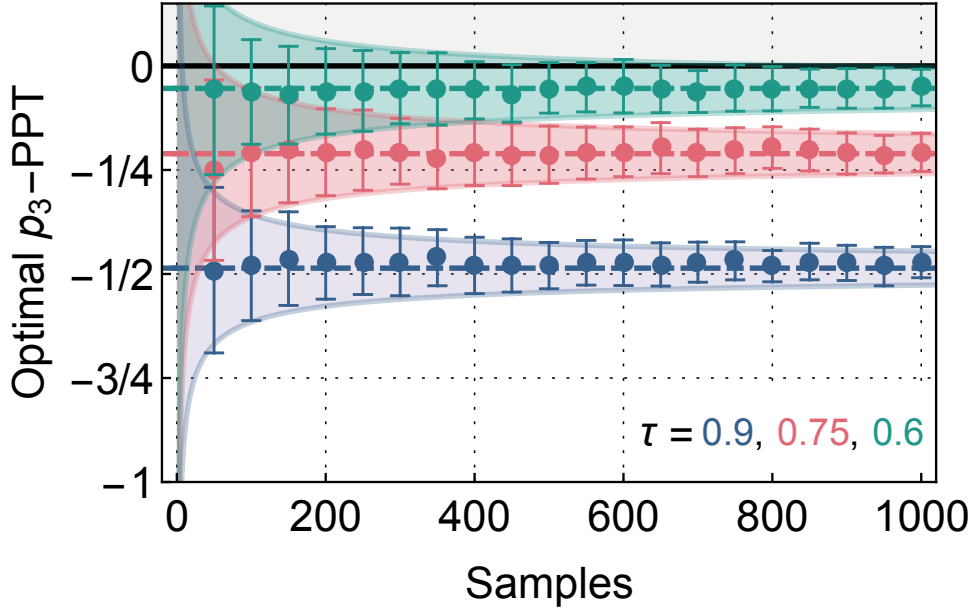


Figure 5.4: Simulated experiment for detecting the entanglement of the first NOON state ($N = 1$), including the effects of losses, noisy copies, and finite statistics. We plot the absolute violation of the optimal p_3 -PPT criterion over the number of samples. The data points and their error bars representing one standard deviation are obtained after averaging over 500 repetitions. The dashed horizontal lines depict the perfect-copy (but still lossy) case. At the same time, the surrounding shaded areas indicate the one standard deviation intervals from the analytical error model for the linear criterion [see Eq. (5.58)]. Given that all curves fall below zero for less than 10^3 samples, our method is capable of detecting strongly non-Gaussian entanglement in a close-to-experimental setting.

We showcase the performance of the optimal criterion as a function of the sample size in Figure 5.4. The simulated data points fluctuate around the analytical predictions (dashed lines, with losses taken into account), and their one-standard-deviation intervals fall within the analytical error bounds of the linear witness [cf. Eq. (5.58)]. For a detailed comparison of the three third-order criteria and the error models without additional noise effects, we refer to Appendix A.4.4. We observe that entanglement is witnessed within one standard deviation up to 40% loss using 10^3 samples only (green points and curves), which is well within the capabilities of current experimental technologies. To conclude, our method remains robust against the typical errors occurring in a realistic experimental setting for the studied cases.

5.6 Discussion

In summary, we generalized entanglement criteria based on moments of the density operator's partial transpose to continuous variables and exploited the multicopy approach in order to access low-order PT moments.

Our measurement schemes use linear interferometers and a few replicas of the state, together with particle-number-resolving detectors. Compared to standard approaches, our method exhibits strong advantages regarding both detection capability – especially in the strongly non-Gaussian regime – and robustness against experimental imperfections, including losses, noise, and finite statistics. In particular, we demonstrated that our criteria can witness higher-order NOON-type entanglement under realistic experimental conditions, highlighting their usefulness in practical applications such as photonic quantum processors.

In comparison, in Ref. [95], a metrological witness is proposed. It relies on local variances and provides a necessary and sufficient criterion for pure-state entanglement, but does not address the role of imperfections or the regime of strongly non-Gaussian entanglement. More recently, in Ref. [90], neural networks are employed to certify non-Gaussian-state entanglement using homodyning. They also account for losses, using the same model as in this work, in the context of single-photon-subtracted and entangled coherent states, whereas our focus is on NOON-state entanglement. A more detailed comparative study of these different approaches would be an interesting direction for future work.

The need for interferometric stability over several state copies – though not necessarily perfect – is arguably the most significant experimental burden of our method. Hence, we believe that it is of particular interest to extend our criteria, and more generally the multicopy method, to arbitrary auxiliary states, which would further reduce this experimental hurdle. We refer to Ref. [35] for a recently introduced entanglement detection scheme where higher-order correlations of a given bipartite state are accessed via its interference with an arbitrary two-mode auxiliary state.

A remaining challenge lies in generating non-Gaussian states. While probabilistic preparation makes producing several copies more demanding, not all schemes are probabilistic [176], and the success rate depends on the specific state [226]. Moreover, probabilistic preparation is effectively equivalent to reduced statistics, and our results show that only about 10^3 samples are sufficient to witness NOON-state entanglement within realistic experimental conditions.

We consider several other directions to be interesting for future work. First, since we have exclusively discussed a two-mode bipartition, one may generalize our setup to both multimode and multipartite cases to analyze many-body and/or multipartite entanglement. The extension to a multimode bipartition is straightforward: The optical circuits for measuring PT moments have to be applied to each mode individually, upon which the particle numbers within one subsystem are summed up. As the number of copies remains constant, the total system size still scales linearly with the number of target state modes. Further, a possible generalization that would allow one to certify multipartite entanglement could be based on the corresponding variant of the PPT criterion [145].

Second, in a somewhat different direction, one may ask whether the preferred readout scheme in the discrete-variable case, that is, randomized measurements in spin bases, can be transferred directly to the CV regime via the Jordan-Schwinger map [73] (see also Refs. [17, 89] for CV shadow tomography). This would not only constitute another way of accessing PT moments, but could be a general alternative to the multicopy method. Since the required number of randomized measurements diverges for an infinite-dimensional Hilbert space, one would have to impose a finite energy (or particle-number) cutoff, which is fairly reasonable from an experimental perspective and a standard procedure in CV quantum information [243].

6 Measurable realignment criteria for entanglement in continuous-variable quantum systems

The content of this chapter is based on the following work: Serge Deside, Vojtěch Kala, and Nicolas J. Cerf. “Measurable realignment criteria for entanglement in continuous-variable quantum systems” [65].

As discussed in previous chapters, certifying entanglement is a central problem in quantum information theory, especially in continuous-variable systems where full state tomography is experimentally demanding. Here, we introduce separability criteria based on the realignment map and show how the corresponding quantities can be accessed within a quantum-optical measurement scheme. We illustrate the method on Gaussian and non-Gaussian states, including two-mode entangled cats and lossy NOON states.

6.1 Preliminaries

Similar to the Schmidt decomposition for pure states, bipartite mixed states can be expanded using the operator Schmidt decomposition [120, 254], which is essentially a singular value decomposition of the bipartite density operator $\hat{\rho}$, namely,

$$\hat{\rho} = \sum_i \lambda_i \hat{A}_i \otimes \hat{B}_i, \quad (6.1)$$

where $\lambda_i \geq 0$ for all i , and the operators $\hat{A}_i$ and $\hat{B}_i$ form orthonormal sets with respect to the Hilbert–Schmidt inner product, i.e.,

$$\text{Tr} [\hat{A}_i^\dagger \hat{A}_j] = \text{Tr} [\hat{B}_i^\dagger \hat{B}_j] = \delta_{ij}. \quad (6.2)$$

We refer to the subsystem associated with the operators $\hat{A}_i$ as Alice’s side and to the one associated with the operators $\hat{B}_i$ as Bob’s side. The singular values λ_i are unique, whereas the associated operators are not. Since $\hat{\rho}$ is Hermitian, the operators can be chosen Hermitian too, although this choice is not necessary.

Using this notation, the realignment criterion (i.e., Theorem 2.6), also known as the computable cross-norm criterion [194], states that every separable state satisfies

$$\sum_i \lambda_i \leq 1. \quad (6.3)$$

Therefore, if $\sum_i \lambda_i > 1$, the state is necessarily entangled.

In the realignment picture, this translates as follows. We introduce the realignment map R , defined by its action on product operators as

$$R(|i\rangle\langle k| \otimes |j\rangle\langle l|) := |i\rangle\langle j| \otimes |k\rangle\langle l|. \quad (6.4)$$

Then, using the following vectorization [44, 140]

$$\hat{X} = \sum_{ij} X_{ij} |i\rangle\langle j| \mapsto |X\rangle = \sum_{ij} X_{ij} |i\rangle \otimes |j\rangle, \quad (6.5)$$

which associates vectors $|A\rangle$ and $|B\rangle$ with the corresponding operators $\hat{A}$ and $\hat{B}$, we find [120]

$$R(\hat{\rho}) = \sum_i \lambda_i |A_i\rangle\langle B_i^*|. \quad (6.6)$$

Here, $|B_i^*\rangle$ denotes the vectorization of the complex conjugate of $\hat{B}_i$. Since the vectorized operators satisfy

$$\langle A_i | A_j \rangle = \langle B_i | B_j \rangle = \delta_{ij}, \quad (6.7)$$

we see that Eq. (6.6) is a singular-value decomposition of $R(\hat{\rho})$. Hence, the singular values of $R(\hat{\rho})$ are precisely the coefficients λ_i . Defining the trace norm of an operator $\hat{X}$ as

$$\|\hat{X}\|_1 := \text{Tr} \sqrt{\hat{X}^\dagger \hat{X}}, \quad (6.8)$$

we obtain

$$\|R(\hat{\rho})\|_1 = \sum_i \lambda_i. \quad (6.9)$$

The realignment criterion can therefore be written in the equivalent form

$$\hat{\rho} \text{ separable} \implies \|R(\hat{\rho})\|_1 \leq 1. \quad (6.10)$$

Although the trace norm of the realigned operator gives a powerful separability criterion, it is generally not directly accessible experimentally in continuous-variable systems. This motivates the weak realignment criterion, which is based instead on the trace of the realigned operator and can be measured [120].

In the following, whenever $\text{Tr}[R(\hat{\rho})]$ is used, we assume that the Hilbert spaces of Alice and Bob satisfy

$$\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathcal{H}. \quad (6.11)$$

Consequently,

$$\mathcal{H}_A \otimes \mathcal{H}_A \simeq \mathcal{H}_B \otimes \mathcal{H}_B, \quad (6.12)$$

so that the vectorized operators $|A_i\rangle$ and $|B_i^*\rangle$ belong to the same Hilbert–Schmidt space and $R(\hat{\rho})$ can be regarded as a square operator.

For every trace-class operator $\hat{X}$, Hölder’s inequality gives [215]

$$|\text{Tr}[\hat{X}]| = |\text{Tr}[\hat{X}\hat{1}]| \leq \|\hat{X}\|_1 \|\hat{1}\|_\infty = \|\hat{X}\|_1. \quad (6.13)$$

Applying this inequality to $\hat{X} = R(\hat{\rho})$, together with Eq. (6.10), gives the weak realignment criterion in the following form [120]

$$\hat{\rho} \text{ separable} \implies |\text{Tr}[R(\hat{\rho})]| \leq 1. \quad (6.14)$$

Thus, a violation of Eq. (6.14) certifies entanglement.

Using Eq. (6.6), the weak quantity can be written as

$$\text{Tr}[R(\hat{\rho})] = \sum_i \lambda_i \langle B_i^* | A_i \rangle. \quad (6.15)$$

The weak and full realignment criteria coincide for Schmidt-symmetric states [120], for which

$$\langle B_i^* | A_j \rangle = \delta_{ij}, \quad \forall i, j. \quad (6.16)$$

In that case,

$$\text{Tr}[R(\hat{\rho})] = \sum_i \lambda_i = \|R(\hat{\rho})\|_1. \quad (6.17)$$

Outside this Schmidt-symmetric case, however, the weak criterion may be strictly weaker than the full realignment criterion. The aim of the criteria introduced below is to partially bridge this gap while keeping the required quantities experimentally measurable.

6.2 Beyond the weak realignment criterion

Here, we present a criterion that improves upon the weak realignment criterion and is based on easily measurable quantum-optical quantities. The measurement schemes are provided in Section 6.3. Under the same identification of the local spaces as in the previous section, we introduce the two families of moments

$$r_n := \text{Tr} \left[\left(R(\hat{\rho})^\dagger R(\hat{\rho}) \right)^{n/2} \right], \quad w_n := \text{Tr} [R(\hat{\rho})^n], \quad (6.18)$$

which satisfy $w_n \geq 0$, $r_n \geq 0$, and $w_n \leq r_n$, as well as $r_n \geq r_{n+1}$, for all $n \in \mathbb{N}_{>0}$ (see Appendix B.1 for the different proofs). With these new notations, the cross-norm and weak realignment criteria are simply rephrased, for all separable states, as $r_1 \leq 1$ and $w_1 \leq 1$, respectively.

Using the vectorized expression of Eq. (6.6), we can rewrite w_1 as

$$w_1 = \sum_i \lambda_i \langle B_i^* | A_i \rangle. \quad (6.19)$$

To simplify the discussion, we define the matrix M as

$$M_{ij} := \langle B_i^* | A_j \rangle = \text{Tr} [\hat{B}_i^T \hat{A}_j]. \quad (6.20)$$

Therefore,

$$w_1 = \sum_i \lambda_i M_{ii}. \quad (6.21)$$

Similarly, the trace norm of the realigned state reads

$$r_1 = \|R(\hat{\rho})\|_1 = \sum_i \lambda_i. \quad (6.22)$$

Generally, the coefficients of the operator Schmidt decomposition and their sum are not directly accessible from w_1 , since w_1 also depends on the overlaps

$$M_{ii} = \text{Tr} [\hat{B}_i^T \hat{A}_i]. \quad (6.23)$$

This can be illustrated by the example of a balanced NOON state

$$|\psi_N\rangle = \frac{1}{\sqrt{2}} (|N, 0\rangle + |0, N\rangle), \quad (6.24)$$

with $N \in \mathbb{N}_{>0}$ the photon number.

This state has the following vector of operator Schmidt coefficients,

$$\lambda = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right), \quad (6.25)$$

together with the operator sets

$$\{\hat{A}_i\}_{i=1}^4 = \{|0\rangle\langle 0|, |0\rangle\langle N|, |N\rangle\langle 0|, |N\rangle\langle N|\}, \quad (6.26)$$

and

$$\{\hat{B}_i\}_{i=1}^4 = \{|N\rangle\langle N|, |N\rangle\langle 0|, |0\rangle\langle N|, |0\rangle\langle 0|\}. \quad (6.27)$$

While $r_1 = 2$, thereby certifying inseparability, $w_1 = 0$ since

$$M_{ii} = \text{Tr}[\hat{B}_i^T \hat{A}_i] = 0 \quad (6.28)$$

for $i = 1, \dots, 4$. The weak realignment criterion is therefore inconclusive. Unlike Schmidt-symmetric states, for which $M = I$ (see Eq. (6.16)), the NOON state in Eq. (6.24) has a corresponding matrix M whose diagonal terms all vanish.

This motivates us to call an operator Schmidt decomposition Schmidt-asymmetric when $M_{ii} = 0$ for all i . For such a decomposition, the first moment of the realigned density matrix satisfies $w_1 = 0$. Thus, the weak realignment criterion cannot certify entanglement for these states. This Schmidt asymmetry can be circumvented by considering the trace of the second power of the realigned density matrix,

$$w_2 = \text{Tr} [R(\hat{\rho})^2]. \quad (6.29)$$

Using the vectorized expression of Eq. (6.6), it can be expressed as

$$w_2 = \sum_{ij} \lambda_i \lambda_j M_{ij} M_{ji}. \quad (6.30)$$

Information on the Schmidt coefficients is then accessible through the off-diagonal elements of M , providing information not contained in w_1 . To make use of the moment w_2 , we need a criterion involving products of Schmidt coefficients, which can be obtained from r_1^2 . Note that the realignment criterion can equivalently be written as $r_1^2 \leq 1$ for all separable states.

Let us now prove the following two inequalities, valid for all states,

$$r_2 \pm w_2 \leq r_1^2 \pm w_1^2. \quad (6.31)$$

Using Eqs. (6.21) and (6.30), together with the fact that $r_n = \sum_i \lambda_i^n$, we obtain

$$r_1^2 \pm w_1^2 - (r_2 \pm w_2) = \sum_{i \neq j} \lambda_i \lambda_j [1 \pm (M_{ii} M_{jj} - M_{ij} M_{ji})]. \quad (6.32)$$

In Appendix B.3, we prove that

$$|M_{ii} M_{jj} - M_{ij} M_{ji}| \leq 1. \quad (6.33)$$

The quantities $d_{ij} := M_{ii} M_{jj} - M_{ij} M_{ji}$ need not be real. Since the left-hand side of Eq. (6.32) is real (because $w_n \geq 0$ and $r_n \geq 0$, see Appendix B.1), we take the real part of d_{ij} and obtain

$$r_1^2 \pm w_1^2 - (r_2 \pm w_2) = \sum_{i \neq j} \lambda_i \lambda_j [1 \pm \text{Re}(d_{ij})]. \quad (6.34)$$

The bound $|d_{ij}| \leq 1$ implies $|\text{Re}(d_{ij})| \leq 1$, so every term in the last sum is non-negative. This proves Eq. (6.31).

From the realignment criterion, separable states satisfy $r_1 \leq 1$, and thus $r_1^2 \leq 1$. Therefore, the inequalities in Eq. (6.31) imply the following separability criteria

$$\hat{\rho} \text{ separable} \implies r_2 \pm (w_2 - w_1^2) \leq 1. \quad (6.35)$$

Equivalently, entanglement is certified whenever

$$r_2 + |w_2 - w_1^2| > 1. \quad (6.36)$$

Moreover, since $w_2 \leq r_2$ (see Appendix B.1), any violation of the weak criterion, $w_1 > 1$, also implies

$$r_2 - w_2 + w_1^2 > 1. \quad (6.37)$$

In this sense, the order-2 criterion contains the weak realignment criterion. Also, r_2 is simply the purity of $\hat{\rho}$, as can be checked straightforwardly. The development for the order-2 criteria can be extended to higher orders (see Appendix B.4). However, these criteria have not been proven to lead to better entanglement certification, since only even moments r_{2n} can be measured using the multicopy method (see Appendix B.6).

Let us now reexamine the NOON state case. Since the state is pure, $r_2 = 1$, and since it is Schmidt-asymmetric, $w_1 = 0$. Further, $r_2 = w_2$ since $R(\hat{\rho})$ is Hermitian in this case. Hence,

$$r_2 + |w_2 - w_1^2| = 2, \quad (6.38)$$

and entanglement is certified. In Section 6.4, we pursue this analysis for lossy NOON states.

Let us now examine the behavior of the criteria for states interpolating between a Schmidt-asymmetric case and a Schmidt-symmetric one. In addition to the asymmetric state from Eq. (6.24) with $N = 1$, which we denote by $|\psi_a\rangle$, we consider the Schmidt-symmetric state

$$|\psi_s\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + |1,1\rangle). \quad (6.39)$$

The density operator $|\psi_s\rangle\langle\psi_s|$ has the following operator Schmidt decomposition:

$$\lambda = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \quad (6.40)$$

with

$$\{\hat{A}_i\}_{i=1}^4 = \{|0\rangle\langle 0|, |0\rangle\langle 1|, |1\rangle\langle 0|, |1\rangle\langle 1|\}, \quad (6.41)$$

and

$$\{\hat{B}_i\}_{i=1}^4 = \{|0\rangle\langle 0|, |0\rangle\langle 1|, |1\rangle\langle 0|, |1\rangle\langle 1|\}. \quad (6.42)$$

Thus, it is Schmidt-symmetric, with $M = I$. We then consider the superposition

$$|\psi_\alpha\rangle = \sqrt{\frac{1}{2} - \alpha} |\psi_a\rangle + \sqrt{\frac{1}{2} + \alpha} |\psi_s\rangle, \quad (6.43)$$

where α ranges from $-\frac{1}{2}$, corresponding to the asymmetric case, to $\frac{1}{2}$, corresponding to the symmetric case. We have

$$r_1 = 1 + 2|\alpha|, \quad w_1 = 1 + 2\alpha, \quad w_2 = r_2 = 1. \quad (6.44)$$

and the dependence of the criteria on α is displayed in Figure 6.1.

Similarly, we can evaluate the criteria for the mixture of the symmetric and asymmetric states

$$\hat{\rho}_\alpha = \left(\frac{1}{2} - \alpha\right) |\psi_a\rangle\langle\psi_a| + \left(\frac{1}{2} + \alpha\right) |\psi_s\rangle\langle\psi_s|. \quad (6.45)$$

Here again, $\alpha \in [-1/2, 1/2]$ so that the coefficients define a valid convex mixture. We have

$$r_1 = 1 + 2|\alpha|, \quad w_1 = 1 + 2\alpha, \quad w_2 = r_2 = \frac{1}{2} + 2\alpha^2, \quad (6.46)$$

and the results are shown in Figure 6.2.

The Werner state of two qubits, a typical discrete-variable entangled state, is considered in Appendix B.5.

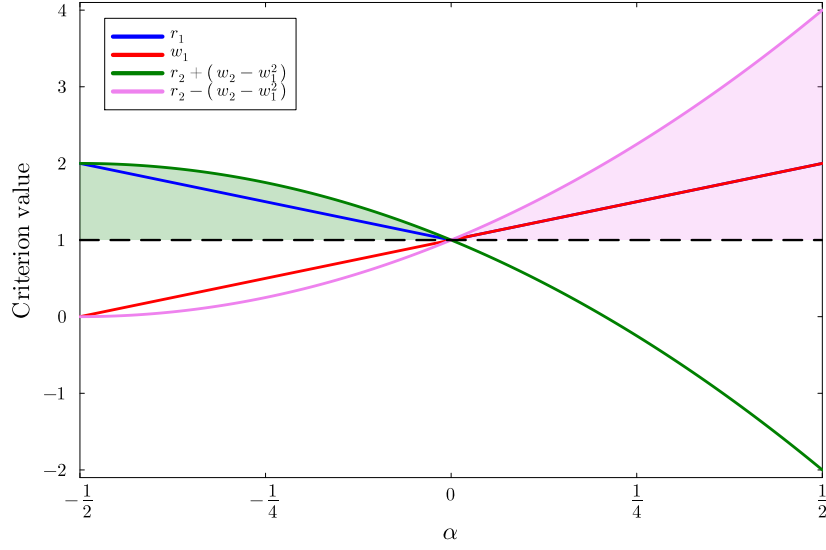


Figure 6.1: Different criteria for the state $|\psi_\alpha\rangle$ from Eq. (6.43). As soon as a criterion attains a value greater than 1, entanglement is certified. We observe that the original realignment criterion, based solely on r_1 , flags entanglement for all $\alpha \neq 0$, whereas the weak realignment criterion, based solely on w_1 , flags entanglement only in the region $\alpha > 0$ (notice that in the range $\alpha \in [0, 1/2]$, the blue and red curves coincide). The order-2 criteria, based on r_2 , w_2 , and w_1 , certify entanglement for the whole entangled parameter range.

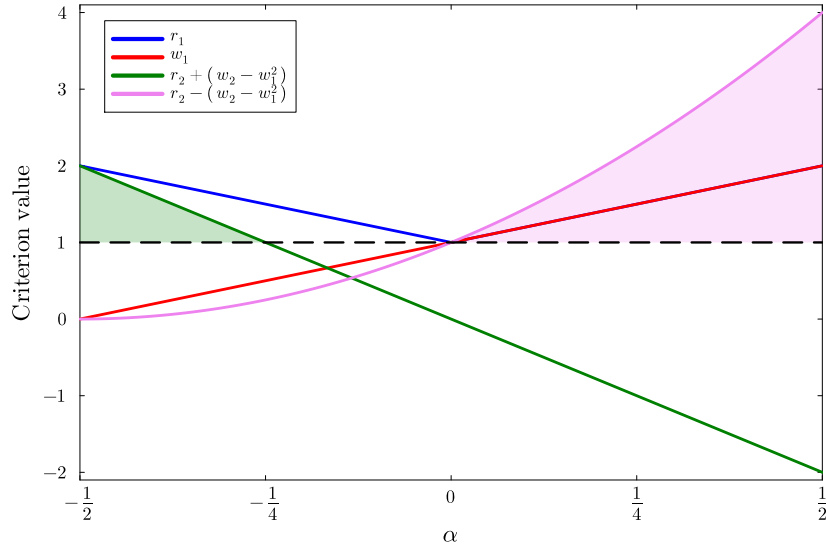


Figure 6.2: Different criteria for the state $\hat{\rho}_\alpha$ from Eq. (6.45). As soon as a criterion attains a value greater than 1, entanglement is certified. We observe that the original realignment criterion, based solely on r_1 , flags entanglement for all $\alpha \neq 0$, whereas the weak realignment criterion, based solely on w_1 , flags entanglement only in the region $\alpha > 0$ (notice that in the range $\alpha \in [0, 1/2]$, the blue and red curves coincide). The order-2 criteria, based on r_2 , w_2 , and w_1 , certify entanglement for $\alpha \in [-1/2, -1/4] \cup (0, 1/2]$.

6.3 Implementation

In this section, we describe measurement schemes for the quantities r_2 , w_1 , and w_2 using multicopy [29] and interferometric techniques (see Appendix B.6 for the multicopy operators associated with w_n and r_{2n} , with $n \in \mathbb{N}_{>0}$, as well as the corresponding quantum-optical measurement schemes). For the optical implementation, we assume that subsystems A and B are encoded in bosonic modes with annihilation operators $\hat{a}_A$ and $\hat{a}_B$.

We first consider the purity of the bipartite state $\hat{\rho}$, namely,

$$r_2 = \text{Tr}[\hat{\rho}^2]. \quad (6.47)$$

Let $\hat{\rho}_{A_1 B_1} \otimes \hat{\rho}_{A_2 B_2}$ denote two identical copies of the state. The swap operator between two copies of a system α is denoted by $\hat{S}_{\alpha_1 \alpha_2}$ and is defined as

$$\hat{S}_{\alpha_1 \alpha_2}(|i\rangle_{\alpha_1} \otimes |j\rangle_{\alpha_2}) = |j\rangle_{\alpha_1} \otimes |i\rangle_{\alpha_2}. \quad (6.48)$$

Using the following identity, valid for all operators $\hat{X}$, $\hat{Y}$ acting on system α ,

$$\text{Tr}[\hat{S}_{\alpha_1 \alpha_2}(\hat{X} \otimes \hat{Y})] = \text{Tr}[\hat{X} \hat{Y}], \quad (6.49)$$

one obtains

$$r_2 = \text{Tr}[(\hat{S}_{A_1 A_2} \otimes \hat{S}_{B_1 B_2})(\hat{\rho}_{A_1 B_1} \otimes \hat{\rho}_{A_2 B_2})]. \quad (6.50)$$

Thus, r_2 can be measured by applying local swap measurements between the two copies of Alice's and Bob's modes.

For bosonic modes, the swap measurement can be implemented by interfering the two copies on balanced beam splitters followed by parity detection [104, 135]. We define the parity operator on mode X as

$$\hat{\Pi}_X = (-1)^{\hat{n}_X}, \quad (6.51)$$

where $\hat{n}_X = \hat{a}_X^\dagger \hat{a}_X$. The balanced beam splitter between modes X and Y (see Section 3.3.1.3) is defined as

$$\hat{B}_{XY} = e^{\frac{\pi}{4}(\hat{a}_X^\dagger \hat{a}_Y - \hat{a}_X \hat{a}_Y^\dagger)}, \quad (6.52)$$

such that

$$\hat{B}_{XY}^\dagger \hat{a}_X \hat{B}_{XY} = \frac{\hat{a}_X + \hat{a}_Y}{\sqrt{2}}, \quad \hat{B}_{XY}^\dagger \hat{a}_Y \hat{B}_{XY} = \frac{-\hat{a}_X + \hat{a}_Y}{\sqrt{2}}. \quad (6.53)$$

With this convention, the swap operator is mapped onto the parity operator of the second output mode, namely,

$$\hat{S}_{XY} = \hat{B}_{XY}^\dagger (\hat{\Pi}_X \otimes \hat{\Pi}_Y) \hat{B}_{XY}. \quad (6.54)$$

Therefore, the purity can be written as

$$r_2 = \text{Tr} \left[\left(\hat{\Pi}_{A_2} \otimes \hat{\Pi}_{B_2} \right) \left(\hat{B}_{A_1 A_2} \otimes \hat{B}_{B_1 B_2} \right) \left(\hat{\rho}_{A_1 B_1} \otimes \hat{\rho}_{A_2 B_2} \right) \left(\hat{B}_{A_1 A_2}^\dagger \otimes \hat{B}_{B_1 B_2}^\dagger \right) \right]. \quad (6.55)$$

Equivalently, r_2 is obtained from the expectation value of the product of the parities measured in the output modes A_2 and B_2 after locally interfering the two copies on balanced beam splitters. For the measurement, only the photon-number parity of each antisymmetric output mode is required. Thus, the detectors need only discriminate between even and odd photon numbers (resolving the exact photon number is unnecessary).

We now introduce the maximally correlated vector that will be used in the measurement of both w_1 and w_2 . It is given by

$$|\Omega\rangle_{AB} = \sum_i |i\rangle_A \otimes |i\rangle_B. \quad (6.56)$$

In the continuous-variable optical setting, $|\Omega\rangle_{AB}$ is understood as the corresponding unnormalizable EPR-type vector. With the quadrature convention

$$\hat{x} = \frac{\hat{a} + \hat{a}^\dagger}{\sqrt{2}}, \quad \hat{p} = \frac{\hat{a} - \hat{a}^\dagger}{i\sqrt{2}}, \quad [\hat{x}, \hat{p}] = i, \quad (6.57)$$

and with generalized eigenstates normalized as

$$\langle x|x'\rangle = \delta(x - x'), \quad \langle p|p'\rangle = \delta(p - p'), \quad (6.58)$$

one has [120]

$$|\Omega\rangle_{AB} = \sqrt{\pi} \hat{B}_{AB}^\dagger (|p=0\rangle_A \otimes |x=0\rangle_B). \quad (6.59)$$

Equivalently,

$$|\Omega\rangle\langle\Omega|_{AB} = \pi \hat{B}_{AB}^\dagger (|p=0\rangle\langle p=0|_A \otimes |x=0\rangle\langle x=0|_B) \hat{B}_{AB}. \quad (6.60)$$

The projectors onto $|x=0\rangle$ and $|p=0\rangle$ are understood in the usual distributional sense. Experimentally, the corresponding probability density is estimated from finite-resolution homodyne data.

The measurement of w_1 follows directly from this representation [120]. By definition,

$$w_1 = \text{Tr} [\hat{\rho} |\Omega\rangle\langle\Omega|_{AB}]. \quad (6.61)$$

Using the optical expression for $|\Omega\rangle\langle\Omega|_{AB}$, this becomes

$$w_1 = \pi \text{Tr} \left[(|p=0\rangle\langle p=0| \otimes |x=0\rangle\langle x=0|) \hat{B}_{AB} \hat{\rho} \hat{B}_{AB}^\dagger \right]. \quad (6.62)$$

Thus, after applying a balanced beam splitter $\hat{B}_{AB}$ to the two modes A and B , one performs homodyne detection of the p quadrature in the symmetric output mode and of the x quadrature in the antisymmetric output mode. If $P_{p_A, x_B}(0, 0)$ denotes the joint probability density at the origin,

$$P_{p_A, x_B}(0, 0) = \text{Tr} \left[(|p=0\rangle\langle p=0| \otimes |x=0\rangle\langle x=0|) \hat{B}_{AB} \hat{\rho} \hat{B}_{AB}^\dagger \right], \quad (6.63)$$

then

$$w_1 = \pi P_{p_A, x_B}(0, 0). \quad (6.64)$$

The measurement of w_2 can also be implemented using two copies of the bipartite state. We write

$$w_2 = \text{Tr} \left[(\hat{\rho}_{A_1 B_1} \otimes \hat{\rho}_{A_2 B_2}) \hat{O}_2 \right], \quad (6.65)$$

where $\hat{O}_2$ is a two-copy observable. This observable can be expressed using two copies of $|\Omega\rangle\langle\Omega|$ and a swap between Bob's two modes:

$$\hat{O}_2 = \hat{S}_{B_1 B_2} (|\Omega\rangle\langle\Omega|_{A_1 B_1} \otimes |\Omega\rangle\langle\Omega|_{A_2 B_2}) \hat{S}_{B_1 B_2}. \quad (6.66)$$

Hence, using the cyclicity of the trace,

$$w_2 = \text{Tr} \left[\hat{S}_{B_1 B_2} (\hat{\rho}_{A_1 B_1} \otimes \hat{\rho}_{A_2 B_2}) \hat{S}_{B_1 B_2} (|\Omega\rangle\langle\Omega|_{A_1 B_1} \otimes |\Omega\rangle\langle\Omega|_{A_2 B_2}) \right]. \quad (6.67)$$

For compactness, define

$$\hat{Q}_{A_j B_j} = |p=0\rangle\langle p=0|_{A_j} \otimes |x=0\rangle\langle x=0|_{B_j}, \quad j = 1, 2. \quad (6.68)$$

Substituting the optical representation of $|\Omega\rangle\langle\Omega|$ into the above expression gives

$$w_2 = \pi^2 \text{Tr} \left[(\hat{Q}_{A_1 B_1} \otimes \hat{Q}_{A_2 B_2}) (\hat{B}_{A_1 B_1} \otimes \hat{B}_{A_2 B_2}) \hat{S}_{B_1 B_2} (\hat{\rho}_{A_1 B_1} \otimes \hat{\rho}_{A_2 B_2}) \hat{S}_{B_1 B_2} (\hat{B}_{A_1 B_1}^\dagger \otimes \hat{B}_{A_2 B_2}^\dagger) \right]. \quad (6.69)$$

Thus, w_2 can be obtained by swapping Bob's two copies, applying balanced beam splitters to the pairs $A_1 B_1$ and $A_2 B_2$, and measuring the joint homodyne probability density at

$$p_{A_1} = 0, \quad x_{B_1} = 0, \quad p_{A_2} = 0, \quad x_{B_2} = 0. \quad (6.70)$$

Equivalently, one may avoid a physical swap by routing the modes so that the two beam splitters mix the pairs $A_1 B_2$ and $A_2 B_1$ instead of $A_1 B_1$ and $A_2 B_2$. The quantum-optical devices required for measuring the purity r_2 , w_1 and w_2 are illustrated in Figure 6.3.

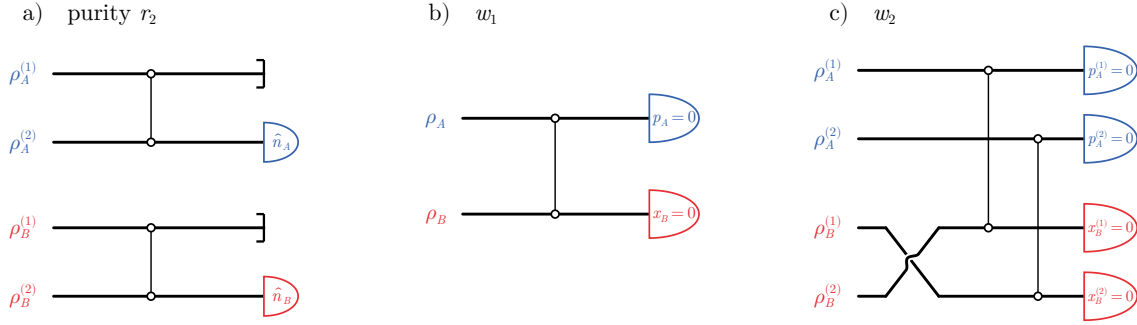


Figure 6.3: Quantum-optical measurement schemes for the three quantities entering the criterion in Eq. (6.36). Vertical bars represent balanced beam splitters. Half ellipses denote photon-number-parity detectors in panel a) and homodyne detectors in panels b) and c). **a)** The purity r_2 is measured by locally interfering two copies on balanced beam splitters and measuring the photon-number parity of each antisymmetric output mode. Only even–odd discrimination is required. The product of the two parity outcomes is averaged over the experimental runs. **b)** The quantity w_1 is accessed using a balanced beam splitter followed by homodyne detection of two conjugate quadratures. **c)** The quantity w_2 is accessed by extending the scheme of panel b) to two copies, up to an exchange of Bob’s two modes.

A useful consequence of the homodyne protocol is that, for a certain class of states, it also provides an indirect determination of the purity. Since, for every bipartite state,

$$r_2 = \text{Tr} \left[R(\hat{\rho})^\dagger R(\hat{\rho}) \right] = \text{Tr} \left[\hat{\rho}^2 \right], \quad (6.71)$$

if, in the fixed local bases used to define the realignment map, the realigned operator is Hermitian, i.e.,

$$R(\hat{\rho}) = R(\hat{\rho})^\dagger, \quad (6.72)$$

then

$$r_2 = \text{Tr} \left[R(\hat{\rho})^2 \right] = w_2. \quad (6.73)$$

In this case, one may estimate w_2 with the two-copy homodyne protocol described above and then infer the purity $r_2 = \text{Tr}(\hat{\rho}^2)$, without performing the photon-number-parity measurement that would directly access r_2 . This is an indirect determination of the purity, that is, the homodyne experiment estimates w_2 , while the identification with r_2 relies on the independently established Hermiticity of $R(\hat{\rho})$.

6.4 Benchmarks

In this section, we investigate whether the order-2 criteria we proposed can be used to certify entanglement for several typical quantum-optical states. We begin by looking at Gaussian states (that were introduced in Section 3.2.2). Then, we look at two important families of non-Gaussian states, namely the two-mode entangled cats (introduced in Section 3.2.3.2) and the lossy NOON states.

6.4.1 Gaussian states

We consider an $n \times n$ continuous-variable system, where Alice and Bob each hold n bosonic modes. We collect the quadrature operators in the vector

$$\hat{\mathbf{r}} := (\hat{x}_{A_1}, \hat{p}_{A_1}, \dots, \hat{x}_{A_n}, \hat{p}_{A_n}, \hat{x}_{B_1}, \hat{p}_{B_1}, \dots, \hat{x}_{B_n}, \hat{p}_{B_n})^T. \quad (6.74)$$

The covariance matrix γ of a state $\hat{\rho}$ is defined by

$$\gamma_{ij} := \frac{1}{2} \text{Tr} [\hat{\rho} \{\Delta \hat{r}_i, \Delta \hat{r}_j\}], \quad \Delta \hat{r}_i := \hat{r}_i - \text{Tr}[\hat{\rho} \hat{r}_i], \quad (6.75)$$

where $\{\cdot, \cdot\}$ denotes the anti-commutator. The displacement vector is

$$\mathbf{d} := \text{Tr}[\hat{\rho} \hat{\mathbf{r}}]. \quad (6.76)$$

A Gaussian state is fully characterized by its first and second moments, $\mathbf{d}$ and γ . Its Wigner function is

$$W_{\hat{\rho}}(\mathbf{r}) = \frac{1}{(2\pi)^{2n} \sqrt{\det \gamma}} e^{-\frac{1}{2}(\mathbf{r}-\mathbf{d})^T \gamma^{-1}(\mathbf{r}-\mathbf{d})}. \quad (6.77)$$

Since first moments can be removed by local displacement operations, which do not affect separability, we focus below on centered Gaussian states, namely states with $\mathbf{d} = \mathbf{0}$.

The purity of a Gaussian state is directly obtained from its covariance matrix. For a total of $2n$ modes, one has

$$r_2 = \text{Tr}[\hat{\rho}^2] = \frac{1}{2^{2n} \sqrt{\det \gamma}}. \quad (6.78)$$

We now turn to w_1 , whose expression for Gaussian states was already given in Ref. [120]. Let

$$S_{\text{BS}} = \frac{1}{\sqrt{2}} \begin{pmatrix} I_{2n} & I_{2n} \\ -I_{2n} & I_{2n} \end{pmatrix} \quad (6.79)$$

be the symplectic matrix describing n balanced beam splitters, each mixing Alice's j th mode with Bob's j th mode.

After this beam-splitter transformation, the covariance matrix becomes

$$\gamma^{(1)} = S_{\text{BS}} \gamma S_{\text{BS}}^T. \quad (6.80)$$

The measurement of w_1 consists, for each balanced beam splitter, in measuring the p quadrature of the first, symmetric output mode and the x quadrature of the second, antisymmetric output mode. We therefore introduce the restricted covariance matrix

$$\gamma_{w_1} = \Pi_{w_1} \gamma^{(1)} \Pi_{w_1}^T, \quad (6.81)$$

where Π_{w_1} is the projection matrix that keeps only the measured quadratures

$$(p_{A_1}, \dots, p_{A_n}, x_{B_1}, \dots, x_{B_n})^T. \quad (6.82)$$

at the output of the beam splitters. The joint probability density of obtaining zero outcomes for all these quadratures is

$$P_{w_1}(0) = \frac{1}{(2\pi)^n \sqrt{\det \gamma_{w_1}}}. \quad (6.83)$$

Using the relation between the maximally correlated vector $|\Omega\rangle$ and the homodyne projection discussed above, we obtain

$$w_1 = \pi^n P_{w_1}(0) = \frac{1}{2^n \sqrt{\det \gamma_{w_1}}}. \quad (6.84)$$

The same reasoning applies to w_2 , but now two copies of the Gaussian state are used. The covariance matrix of the two-copy state is

$$\gamma \oplus \gamma. \quad (6.85)$$

Let $P_B^{(12)}$ denote the permutation matrix that exchanges Bob's modes between the two copies. This is the covariance-matrix representation of the swap $\hat{S}_{B_1 B_2}$ appearing in the operator expression of w_2 . After this swap, we apply balanced beam splitters to the pairs $A_1 B_1$ and $A_2 B_2$. At the covariance-matrix level, this gives

$$\gamma^{(2)} = S_{\text{BS}}^{(2)} P_B^{(12)} (\gamma \oplus \gamma) \left(P_B^{(12)}\right)^T \left(S_{\text{BS}}^{(2)}\right)^T, \quad (6.86)$$

where

$$S_{\text{BS}}^{(2)} := S_{\text{BS}} \oplus S_{\text{BS}}. \quad (6.87)$$

For both copies, the measured quadratures are the p quadratures of the symmetric output modes and the x quadratures of the antisymmetric output modes. We define the corresponding restricted covariance matrix as

$$\gamma_{w_2} = \Pi_{w_2} \gamma^{(2)} \Pi_{w_2}^T, \quad (6.88)$$

where Π_{w_2} keeps the quadratures

$$\left(p_{A_1}^{(1)}, \dots, p_{A_n}^{(1)}, x_{B_1}^{(1)}, \dots, x_{B_n}^{(1)}, p_{A_1}^{(2)}, \dots, p_{A_n}^{(2)}, x_{B_1}^{(2)}, \dots, x_{B_n}^{(2)} \right)^T. \quad (6.89)$$

The corresponding zero-outcome probability density is

$$P_{w_2}(\mathbf{0}) = \frac{1}{(2\pi)^{2n} \sqrt{\det \gamma_{w_2}}}, \quad (6.90)$$

and hence

$$w_2 = \pi^{2n} P_{w_2}(\mathbf{0}) = \frac{1}{2^{2n} \sqrt{\det \gamma_{w_2}}}. \quad (6.91)$$

Let us restrict the following discussion to a 1×1 Gaussian state. The covariance matrix can be brought by local Gaussian unitaries to the standard form [243]

$$\gamma = \begin{pmatrix} a & 0 & c & 0 \\ 0 & a & 0 & d \\ c & 0 & b & 0 \\ 0 & d & 0 & b \end{pmatrix}. \quad (6.92)$$

In this case, the above expressions reduce to simple closed formulas. The purity is

$$r_2 = \frac{1}{4 \sqrt{(ab - c^2)(ab - d^2)}}. \quad (6.93)$$

For w_1 , the restricted covariance matrix is

$$\gamma_{w_1} = \frac{1}{2} \begin{pmatrix} a + b + 2d & 0 \\ 0 & a + b - 2c \end{pmatrix}. \quad (6.94)$$

so that

$$w_1 = \frac{1}{\sqrt{(a + b + 2d)(a + b - 2c)}}. \quad (6.95)$$

For w_2 , after swapping Bob's two copies and applying the two balanced beam splitters, the restricted covariance matrix is

$$\gamma_{w_2} = \frac{1}{2} \begin{pmatrix} a+b & 0 & 2d & 0 \\ 0 & a+b & 0 & -2c \\ 2d & 0 & a+b & 0 \\ 0 & -2c & 0 & a+b \end{pmatrix}. \quad (6.96)$$

Therefore,

$$w_2 = \frac{1}{\sqrt{[(a+b)^2 - 4c^2][(a+b)^2 - 4d^2]}}. \quad (6.97)$$

Let us now consider a locally phase-flipped noisy two-mode squeezed vacuum (TMSV) state (see Eq. (3.51) for the covariance matrix of a noiseless TMSV state with no phase-flip). It is related to the noisy TMSV by a local phase shift of π on Bob's mode and therefore has the same entanglement (in the entropy of entanglement sense, see Eq. (2.19)). Its covariance matrix is of the form of Eq. (6.92), with

$$a = b = \frac{\nu \cosh(2r)}{2}, \quad c = -\frac{\nu \sinh(2r)}{2}, \quad d = \frac{\nu \sinh(2r)}{2}, \quad (6.98)$$

where $r > 0$ and $\nu \geq 1$. It is pure when $\nu = 1$, and mixed otherwise. Its entanglement is not detected by the weak realignment criterion of Eq. (6.14), but it can be detected by the improved criterion of Eq. (6.35). Indeed, using Eq. (6.95), we obtain

$$w_1 = \frac{e^{-2r}}{\nu}. \quad (6.99)$$

Since $r > 0$ and $\nu \geq 1$, one has $w_1 \leq 1$. Hence, the weak realignment criterion does not certify entanglement for this family of states. Let us now evaluate the improved criterion. From Eq. (6.93), the purity is $r_2 = 1/\nu^2$. Similarly, using Eq. (6.97), we obtain $w_2 = 1/\nu^2$. Therefore,

$$r_2 + (w_2 - w_1^2) = \frac{2 - e^{-4r}}{\nu^2}. \quad (6.100)$$

The improved criterion detects entanglement whenever

$$\nu < \sqrt{2 - e^{-4r}}. \quad (6.101)$$

Thus, for all parameters satisfying

$$1 \leq \nu < \sqrt{2 - e^{-4r}}, \quad (6.102)$$

the state is detected as entangled by the improved criterion, while remaining undetected by the weak realignment criterion.

In particular, for pure TMSV states, corresponding to $v = 1$, the improved criterion gives

$$r_2 + (w_2 - w_1^2) = 2 - e^{-4r}. \quad (6.103)$$

Since $r > 0$, one has

$$2 - e^{-4r} > 1, \quad (6.104)$$

so every pure TMSV state with nonzero squeezing is certified as entangled.

6.4.2 Non-Gaussian states

6.4.2.1 Schrödinger cat states

We next consider a partially dephased two-mode Schrödinger cat state,

$$\hat{\rho}_{\pm} = \frac{1}{\mathcal{N}_{\pm}} [|\alpha, \beta\rangle\langle\alpha, \beta| + |-\alpha, -\beta\rangle\langle-\alpha, -\beta| \pm (1-z) (|\alpha, \beta\rangle\langle-\alpha, -\beta| + |-\alpha, -\beta\rangle\langle\alpha, \beta|)], \quad (6.105)$$

where

$$\mathcal{N}_{\pm} = 2 \left[1 \pm (1-z) \exp^{-2(|\alpha|^2 + |\beta|^2)} \right] \quad (6.106)$$

ensures normalization. Here, $\alpha, \beta \in \mathbb{C}$, and the parameter $z \in [0, 1]$ controls the coherence between the two product states $|\alpha, \beta\rangle$ and $|-\alpha, -\beta\rangle$. For $z = 0$, the coherence is maximal. By contrast, for $z = 1$, the coherence is minimal (equal to zero). The $+$ and $-$ signs correspond to the even and odd cat states, respectively.

The first realignment moment of the state in Eq. (6.105) evaluates to

$$w_1 = \frac{2[1 \pm (1-z)]}{\mathcal{N}_{\pm}} e^{-|\alpha|^2 - |\beta|^2 + 2\text{Re}(\alpha\beta)}. \quad (6.107)$$

The second realignment moment is expressed as

$$w_2 = \frac{4e^{-2(|\alpha|^2 + |\beta|^2)}}{\mathcal{N}_{\pm}^2} \{ [1 + (1-z)^2] \cosh[4\text{Re}(\alpha\beta)] \pm 2(1-z) \cos[4\text{Im}(\alpha\beta)] \}. \quad (6.108)$$

Finally, the purity, $r_2 = \text{Tr}[\hat{\rho}_{\pm}^2]$, is given by

$$r_2 = \frac{1}{\mathcal{N}_{\pm}^2} \left\{ 2 \left[1 + e^{-4(|\alpha|^2 + |\beta|^2)} \right] [1 + (1-z)^2] \pm 8(1-z) e^{-2(|\alpha|^2 + |\beta|^2)} \right\}. \quad (6.109)$$

These moments allow us to evaluate the measurable order-two realignment criterion of Eq. (6.35) as displayed in Figure 6.4.

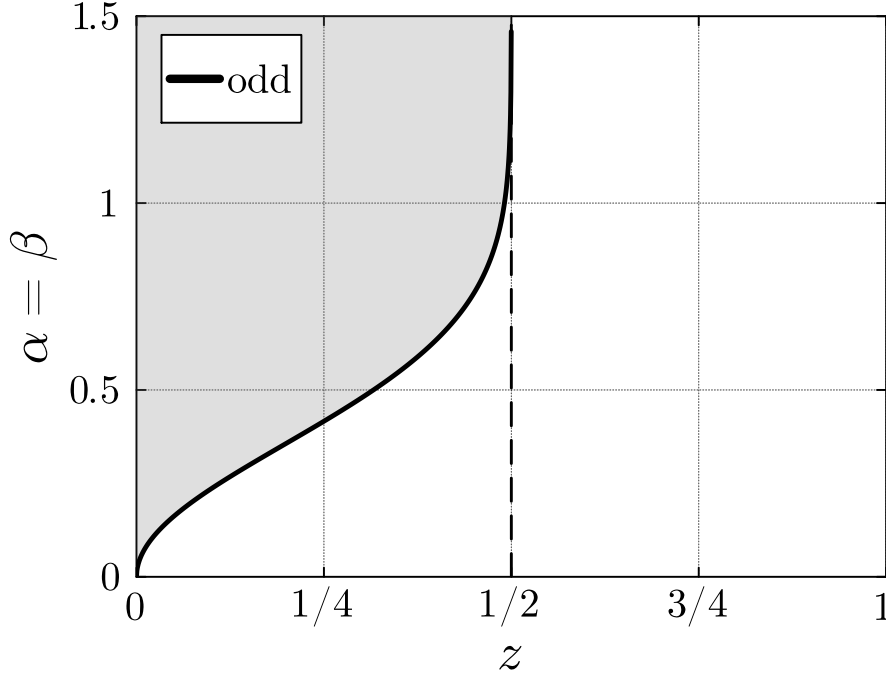


Figure 6.4: Detection boundary for odd cat states with positive and real parameters $\alpha = \beta$. The order-two criterion detects entanglement above the curve $a(z) = \frac{1}{2}\sqrt{-\ln(1-2z)}$, with $0 \leq z < \frac{1}{2}$. No violation occurs for $z \geq 1/2$. For the even cats, the weak criterion detects all states with $\alpha = \beta > 0$ and $z < 1$; the cases $\alpha = \beta = 0$ or $z = 1$ are separable.

6.4.2.2 Lossy NOON states

We now consider the experimentally motivated case of lossy NOON states. We start from the pure two-mode state (often simply called generalized NOON state)

$$|\psi_N\rangle = \alpha|N, 0\rangle + \beta|0, N\rangle, \quad |\alpha|^2 + |\beta|^2 = 1, \quad (6.110)$$

and assume that the two modes undergo independent pure-loss channels $\mathcal{L}$ with transmissivities τ_A and τ_B . The resulting state is

$$\hat{\rho}_N^{\text{loss}} = (\mathcal{L}_{\tau_A} \otimes \mathcal{L}_{\tau_B})(|\psi_N\rangle\langle\psi_N|). \quad (6.111)$$

Explicitly, this gives

$$\hat{\rho}_N^{\text{loss}} = (p_0 + q_0)|0, 0\rangle\langle 0, 0| + \sum_{k=1}^N p_k|k, 0\rangle\langle k, 0| + \sum_{k=1}^N q_k|0, k\rangle\langle 0, k| + \kappa|N, 0\rangle\langle 0, N| + \kappa^*|0, N\rangle\langle N, 0|, \quad (6.112)$$

where the different terms introduced are defined as

$$p_k := |\alpha|^2 \binom{N}{k} \tau_A^k (1 - \tau_A)^{N-k}, \quad q_k := |\beta|^2 \binom{N}{k} \tau_B^k (1 - \tau_B)^{N-k}, \quad \kappa := \alpha\beta^* (\tau_A \tau_B)^{N/2}. \quad (6.113)$$

Applying the realignment map gives

$$R(\hat{\rho}_N^{\text{loss}}) = (p_0 + q_0)|0, 0\rangle\langle 0, 0| + \sum_{k=1}^N p_k |k, k\rangle\langle 0, 0| + \sum_{k=1}^N q_k |0, 0\rangle\langle k, k| + \kappa |N, 0\rangle\langle 0, N| + \kappa^* |0, N\rangle\langle N, 0|. \quad (6.114)$$

Therefore,

$$w_1 = \text{Tr} \left[R(\hat{\rho}_N^{\text{loss}}) \right] = p_0 + q_0. \quad (6.115)$$

Equivalently,

$$w_1 = |\alpha|^2 (1 - \tau_A)^N + |\beta|^2 (1 - \tau_B)^N. \quad (6.116)$$

Since $0 \leq \tau_A, \tau_B \leq 1$ and $|\alpha|^2 + |\beta|^2 = 1$, one has $w_1 \leq 1$. Hence, the weak realignment criterion never certifies entanglement for this family.

Let us now compute the quantities entering the improved criterion. The purity is

$$r_2 = (p_0 + q_0)^2 + \sum_{k=1}^N (p_k^2 + q_k^2) + 2|\kappa|^2. \quad (6.117)$$

Similarly, the w_2 term is

$$w_2 = (p_0 + q_0)^2 + 2 \sum_{k=1}^N p_k q_k + 2|\kappa|^2. \quad (6.118)$$

For this family,

$$w_2 - w_1^2 = 2 \sum_{k=1}^N p_k q_k + 2|\kappa|^2 \geq 0. \quad (6.119)$$

Hence, the absolute-value criterion of Eq. (6.36) reduces here to the plus branch,

$$r_2 + |w_2 - w_1^2| = r_2 + (w_2 - w_1^2). \quad (6.120)$$

Combining the above expressions for w_1 , w_2 , and r_2 , we obtain

$$r_2 + (w_2 - w_1^2) = \sum_{k=0}^N (p_k + q_k)^2 + 4|\kappa|^2. \quad (6.121)$$

Therefore, the improved criterion detects entanglement whenever

$$\sum_{k=0}^N (p_k + q_k)^2 + 4|\kappa|^2 > 1. \quad (6.122)$$

In Figure 6.5, we plot the value of the improved criterion as a function of the transmissivity $\tau = \tau_A = \tau_B$ for balanced NOON states, i.e., $\alpha = \beta = 1/\sqrt{2}$, and for photon numbers $N = 1, \dots, 5$. For these states, the realigned operator is Hermitian, and hence $r_2 = w_2$. The purity can therefore be inferred from the homodyne data used to estimate w_2 .

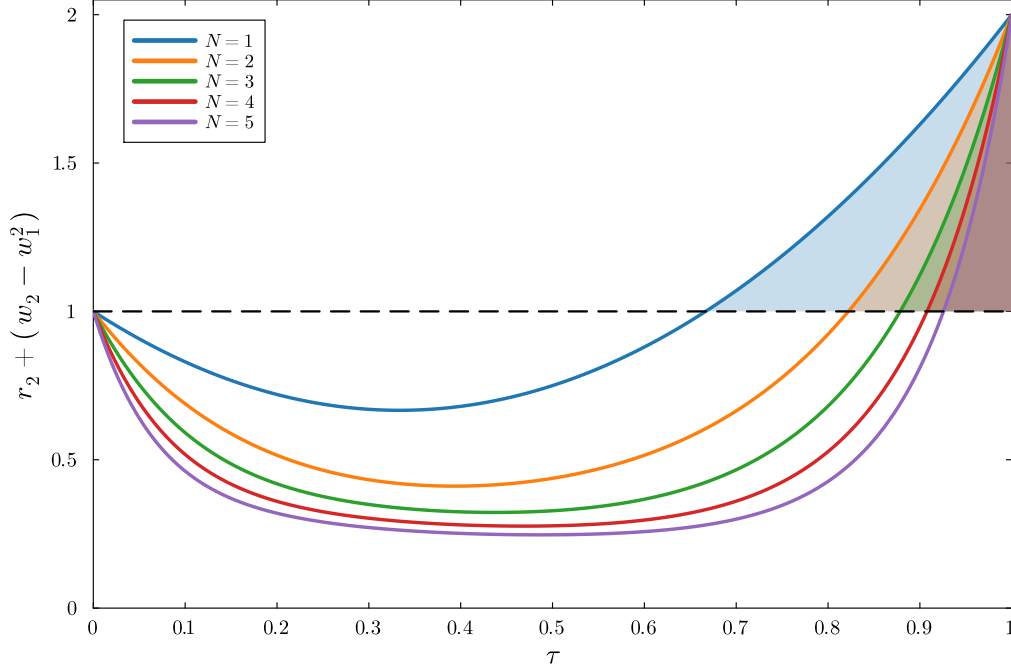


Figure 6.5: Value of the realignment criterion $r_2 + (w_2 - w_1^2)$ for balanced NOON states subjected to symmetric losses, $\tau_A = \tau_B = \tau$, for photon numbers $N = 1, \dots, 5$. The horizontal dashed line indicates the separability bound equal to 1. Entanglement is certified whenever a curve lies above this line. For each N , the criterion is therefore violated when the transmissivity exceeds a critical value, increasing with N .

In the lossless limit, when $\tau_A = \tau_B = 1$, one has

$$w_1 = 0, \quad r_2 = 1, \quad w_2 = 4|\alpha|^2|\beta|^2. \quad (6.123)$$

Thus,

$$r_2 + (w_2 - w_1^2) = 1 + 4|\alpha|^2|\beta|^2. \quad (6.124)$$

Hence, every entangled NOON state (every state with $\alpha\beta \neq 0$) is detected by the improved criterion in the lossless case, whereas it remains undetected by the weak realignment criterion.

6.5 Discussion

We have introduced experimentally accessible continuous-variable separability criteria based on the realignment criterion. In particular, our approach improves upon the weak realignment criterion proposed in Ref. [120], which can be interpreted as the first-order member of a hierarchy (see Appendix B.4). Although higher-order criteria may also be useful in certain situations, we have shown that the second-order criteria already certify entanglement for several classes of states that are not detected by the weak realignment criterion, including Gaussian states, two-mode entangled cat states, and NOON states.

So far, however, we have not demonstrated that these detection capabilities outperform the separability criteria based on partial transposition presented in Chapter 5. A potential advantage of the present approach is instead experimental. The moments of the realigned operator can be accessed using homodyne detection together with photon-number-parity measurements when the realigned operator is not known to be Hermitian¹. If $R(\hat{\rho}) = R(\hat{\rho})^\dagger$, the purity satisfies $r_2 = w_2$ and can therefore be inferred from the homodyne measurement of w_2 , without requiring parity detection.

The total number of copies required by the realignment-based second-order criteria and the partial-transpose-based third-order criteria schemes is the same, namely five. Their optical implementations are nevertheless different. In the realignment-based protocol, the measurements involving one and two copies can be performed separately. The corresponding interferometric networks therefore require coherence between fewer copies at a time. By contrast, the third-order criteria based on partial transposition require all three copies to be processed within a single interferometric network, hence it requires a higher interferometric stability.

An important direction for future work is to determine whether there exist physically relevant states that are not detected by the partial-transpose-based criteria considered here but are detected by one or more realignment-based criteria. It would also be worthwhile to identify specific regimes in which higher-order realignment moments provide a significant advantage over the second-order criteria, and to investigate the robustness of the resulting protocols against realistic experimental imperfections.

¹In specific settings, photon-number parity can be measured more easily than the exact photon number [184].

Part III

Majorization lattice

7 Lattice structure of majorization

Introduced and developed independently in the early twentieth century by mathematicians and economists¹ [161], the theory of majorization gives a means to compare vectors in terms of their intrinsic disorder. It has seen applications, e.g., in econometrics to measure the distribution of income within a given population [167], in information theory to bound entropic quantities to estimate the capacity of communication channels [109], or in quantum information theory to characterize the allowed transformations between bipartite pure states under physically meaningful constraints [171].

In lattice theory, non-increasingly arranged probability vectors can be ordered by majorization, yielding the so-called *majorization lattice* introduced in the early twenty-first century in the context of (classical) information theory [46]. This nontrivial structure unveils two essential operations: the greatest lower bound, called the *meet*, and the least upper bound, called the *join*. The majorization lattice has recently been used in quantum information theory [26, 149, 199, 250]. In particular, in Chapter 8, based on Ref. [62], we use this structure to propose new conversion protocols between bipartite pure states.

In this chapter, we first review the basic notions of majorization theory in Section 7.1. We then recall the relevant lattice-theoretic terminology in Section 7.2, before presenting the majorization lattice in Section 7.3.

7.1 Theory of majorization

When two vectors are related by majorization, this relation can be understood from several equivalent viewpoints. Informally, if $\mathbf{u}$ is majorized by $\mathbf{v}$, denoted $\mathbf{u} \prec \mathbf{v}$, then $\mathbf{u}$ is spread out more evenly than $\mathbf{v}$, conversely $\mathbf{v}$ is more concentrated than $\mathbf{u}$. This can be expressed in terms of inequalities between partial sums of ordered entries of $\mathbf{u}$ and $\mathbf{v}$ or, equivalently, by the existence of a random mixing of the entries of $\mathbf{v}$ to obtain $\mathbf{u}$, or also, by an inequality relation satisfied for every convex function. These formulations are made precise in this section.

7.1.1 Definition with inequalities

We work with finite-dimensional vectors. Although majorization can be defined for arbitrary real vectors with the same total sum, this manuscript considers only non-negative vectors, and in particular, probability vectors.

¹Notably, Lorenz (1905), Dalton (1920), Schur (1923), and Hardy, Littlewood and Pólya (1929).

We therefore restrict ourselves to the finite-dimensional, discrete setting; see, e.g., Refs. [161, 231] for extensions to infinite-dimensional vectors, vectors with negative entries, and densities. Throughout this section, we primarily rely on the broad overview of inequalities and majorization that is Ref. [161].

Given a vector $\mathbf{u} \in \mathbb{R}^d$, let $\mathbf{u}^\downarrow$ denote the vector obtained by arranging the entries of $\mathbf{u}$ in non-increasing order, and let $\mathbf{u}^\uparrow$ denote the vector obtained by arranging them in non-decreasing order.

Definition 7.1 (Cumulative sums). For $\mathbf{u} \in \mathbb{R}^d$, we define the decreasing and increasing cumulative sums by

$$s_k^\downarrow(\mathbf{u}) := \sum_{i=1}^k u_i^\downarrow, \quad s_k^\uparrow(\mathbf{u}) := \sum_{i=1}^k u_i^\uparrow, \quad k = 1, \dots, d. \quad (7.1)$$

We also fix $s_0^\downarrow(\mathbf{u}) = s_0^\uparrow(\mathbf{u}) = 0$.

Definition 7.2 (Majorization [116]). For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, we say that $\mathbf{u}$ is majorized by $\mathbf{v}$, and write $\mathbf{u} < \mathbf{v}$, if

$$s_k^\downarrow(\mathbf{u}) \leq s_k^\downarrow(\mathbf{v}), \quad \forall k = 1, \dots, d-1, \quad \text{and} \quad s_d^\downarrow(\mathbf{u}) = s_d^\downarrow(\mathbf{v}). \quad (7.2)$$

Equivalently, the same relation can be expressed in terms of increasing cumulative sums as

$$s_k^\uparrow(\mathbf{u}) \geq s_k^\uparrow(\mathbf{v}), \quad \forall k = 1, \dots, d-1, \quad \text{and} \quad s_d^\uparrow(\mathbf{u}) = s_d^\uparrow(\mathbf{v}). \quad (7.3)$$

The equivalence between Eqs. (7.2) and (7.3) follows, for $k = 1, \dots, d$, from the identity

$$s_k^\downarrow(\mathbf{u}) = s_d(\mathbf{u}) - s_{d-k}^\uparrow(\mathbf{u}), \quad (7.4)$$

together with the equality $s_d(\mathbf{u}) = s_d(\mathbf{v})$. Since $s_d^\downarrow(\mathbf{u}) = s_d^\uparrow(\mathbf{u})$, we simply write $s_d(\mathbf{u})$ for the total sum of the entries of $\mathbf{u}$. In the remainder of this thesis, we usually make use of Eq. (7.2) rather than Eq. (7.3).

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be comparable under majorization if either $\mathbf{u} < \mathbf{v}$ or $\mathbf{v} < \mathbf{u}$. If both relations hold, then $\mathbf{u}$ and $\mathbf{v}$ have the same decreasing rearrangement, i.e., they differ at most by a permutation of their entries. In this case, we write $\mathbf{u} \equiv \mathbf{v}$ and say that $\mathbf{u}$ and $\mathbf{v}$ are equivalent under majorization. If neither $\mathbf{u} < \mathbf{v}$ nor $\mathbf{v} < \mathbf{u}$ holds, then $\mathbf{u}$ and $\mathbf{v}$ are incomparable².

²For $d = 2$, any two vectors with the same total sum are comparable under majorization. Incomparability can occur only from dimension $d = 3$ onward.

The vectors considered in this manuscript are often probability vectors. We denote the probability simplex in dimension d by

$$\mathcal{P}_d := \left\{ \mathbf{u} \in \mathbb{R}^d \mid u_i \geq 0 \text{ for all } i = 1, \dots, d, \text{ and } \sum_{i=1}^d u_i = 1 \right\}. \quad (7.5)$$

If $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d$, the equality condition in the definition of majorization is automatically satisfied, since $s_d(\mathbf{u}) = s_d(\mathbf{v}) = 1$.

Importantly, for all $\mathbf{u} \in \mathcal{P}_d$, we have

$$\left(\frac{1}{d}, \dots, \frac{1}{d} \right)^T < \mathbf{u} < (1, 0, \dots, 0)^T. \quad (7.6)$$

That is, within $\mathcal{P}_d$, the uniform vector is the least concentrated element, whereas the peaked vector is the most concentrated one. For instance, in dimension 3, let

$$\mathbf{u} = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)^T, \quad \mathbf{v} = \left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6} \right)^T. \quad (7.7)$$

Then, $\mathbf{u} < \mathbf{v}$, since

$$s_1^\downarrow(\mathbf{u}) = \frac{1}{3} \leq \frac{1}{2} = s_1^\downarrow(\mathbf{v}), \quad s_2^\downarrow(\mathbf{u}) = \frac{2}{3} \leq \frac{5}{6} = s_2^\downarrow(\mathbf{v}), \quad s_3(\mathbf{u}) = 1 = s_3(\mathbf{v}). \quad (7.8)$$

Thus, $\mathbf{v}$ is more concentrated than $\mathbf{u}$. By contrast, the vectors

$$\mathbf{u} = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4} \right)^T, \quad \mathbf{v} = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5} \right)^T \quad (7.9)$$

are incomparable. Indeed, $s_1^\downarrow(\mathbf{u}) = \frac{1}{2} > \frac{2}{5} = s_1^\downarrow(\mathbf{v})$, while $s_2^\downarrow(\mathbf{u}) = \frac{3}{4} < \frac{4}{5} = s_2^\downarrow(\mathbf{v})$.

When the equality condition in majorization is removed, two variants of majorization can prove useful.

Definition 7.3 (Submajorization [161]). For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, we say that $\mathbf{u}$ is submajorized by $\mathbf{v}$, and write $\mathbf{u} <_w \mathbf{v}$, if

$$s_k^\downarrow(\mathbf{u}) \leq s_k^\downarrow(\mathbf{v}), \quad \forall k = 1, \dots, d. \quad (7.10)$$

Definition 7.4 (Supermajorization [161]). For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, we say that $\mathbf{u}$ is supermajorized by $\mathbf{v}$, and write $\mathbf{u} <^w \mathbf{v}$, if

$$s_k^\uparrow(\mathbf{u}) \geq s_k^\uparrow(\mathbf{v}), \quad \forall k = 1, \dots, d. \quad (7.11)$$

If both $\mathbf{u} <_w \mathbf{v}$ and $\mathbf{u} <^w \mathbf{v}$ hold, then the inequalities at $k = d$ imply $s_d(\mathbf{u}) = s_d(\mathbf{v})$. Hence, the usual majorization relation $\mathbf{u} < \mathbf{v}$ is recovered.

7.1.2 Equivalent formulations

Depending on the context, several equivalent formulations of majorization are useful. We now review some of them. We first introduce doubly stochastic matrices, namely square matrices with non-negative entries whose rows and columns each sum to 1.

Definition 7.5 (Doubly stochastic matrix). We say that $D \in \mathbb{R}^{d \times d}$ is doubly stochastic if

$$D_{ij} \geq 0, \forall i, j = 1, \dots, d, \quad \sum_{i=1}^d D_{ij} = 1, \forall j = 1, \dots, d, \quad \sum_{j=1}^d D_{ij} = 1, \forall i = 1, \dots, d. \quad (7.12)$$

Majorization has the following characterization in terms of doubly stochastic matrices [117].

Theorem 7.1. For $u, v \in \mathbb{R}^d$, we have $u < v$ if and only if there exists a doubly stochastic matrix D such that

$$u = Dv. \quad (7.13)$$

Notice that such a doubly stochastic matrix is not unique in general.

Birkhoff's theorem [161] states that the extreme points of the set of doubly stochastic matrices are the permutation matrices. Hence, any doubly stochastic matrix can be written as a convex combination of permutation matrices, i.e.,

$$D = \sum_i p_i P_i, \quad (7.14)$$

where p is a probability vector and the P_i are permutation matrices. Combining this decomposition with Eq. (7.13), the relation $u < v$ can be interpreted as obtaining u by randomly permuting the entries of v and averaging the result.

We can adopt a similar view by introducing T -transforms, which are elementary averaging operations that act on two coordinates. When applied to a larger and a smaller coordinate, such an operation transfers weight from the larger to the smaller, making the vector more evenly distributed. For this reason, T -transforms are also called Robin-Hood transformations.

Definition 7.6 (T -transform). A T -transform is a matrix of the form

$$T = \lambda I + (1 - \lambda)P_{ij}, \quad (7.15)$$

where $0 \leq \lambda \leq 1$, $I \in \mathbb{R}^{d \times d}$ is the identity matrix, and $P_{ij} \in \mathbb{R}^{d \times d}$ is the permutation matrix that swaps coordinates i and j , with $i \neq j$.

Given $\mathbf{u} \in \mathbb{R}^d$, we have

$$T\mathbf{u} = (u_1, \dots, \lambda u_i + (1 - \lambda)u_j, \dots, \lambda u_j + (1 - \lambda)u_i, \dots, u_d), \quad (7.16)$$

with $1 \leq i < j \leq d$. Thus, applying T to $\mathbf{u}$ leaves all coordinates unchanged except the i th and j th ones, which are replaced by convex combinations of each other. The following theorem makes the link with majorization [116].

Theorem 7.2. *For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, we have $\mathbf{u} < \mathbf{v}$ if and only if $\mathbf{u}$ can be obtained from $\mathbf{v}$ by a finite sequence of T -transforms; that is, there exist T -transforms $T_1, \dots, T_m$ such that*

$$\mathbf{u} = T_m \cdots T_1 \mathbf{v}. \quad (7.17)$$

When working with decreasing rearrangements, one may choose $m \leq d - 1$.

Finally, majorization can also be characterized by inequalities involving convex (or concave) functions [116].

Theorem 7.3. *For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, we have $\mathbf{u} < \mathbf{v}$ if and only if, for every convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$,*

$$\sum_{i=1}^d \phi(u_i) \leq \sum_{i=1}^d \phi(v_i). \quad (7.18)$$

Equivalently, $\mathbf{u} < \mathbf{v}$ if and only if, for every concave function $\psi : \mathbb{R} \rightarrow \mathbb{R}$,

$$\sum_{i=1}^d \psi(u_i) \geq \sum_{i=1}^d \psi(v_i). \quad (7.19)$$

There exist order-preserving functions under majorization, which we introduce now.

Definition 7.7 (Schur-convexity and Schur-concavity [203]). Let $\mathcal{A} \subseteq \mathbb{R}^d$ be a set on which majorization is defined. A function $f : \mathcal{A} \rightarrow \mathbb{R}$ is said to be Schur-convex if, for all $\mathbf{u}, \mathbf{v} \in \mathcal{A}$,

$$\mathbf{u} < \mathbf{v} \implies f(\mathbf{u}) \leq f(\mathbf{v}). \quad (7.20)$$

Similarly, a function $g : \mathcal{A} \rightarrow \mathbb{R}$ is said to be Schur-concave if, for all $\mathbf{u}, \mathbf{v} \in \mathcal{A}$,

$$\mathbf{u} < \mathbf{v} \implies g(\mathbf{u}) \geq g(\mathbf{v}). \quad (7.21)$$

Obviously, if f is Schur-convex, then $-f$ is Schur-concave. If $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function, the function

$$f_\phi : \mathbb{R}^d \rightarrow \mathbb{R} : \mathbf{u} \mapsto f_\phi(\mathbf{u}) = \sum_{i=1}^d \phi(u_i), \quad (7.22)$$

is Schur-convex by Theorem 7.3. For instance, choosing $\phi(x) = x^2$ shows that $\mathbf{u} \mapsto \sum_i u_i^2$ is Schur-convex.

7.1.3 Lorenz curves

In order to get a visual representation of majorization, we introduce the notion of a (concave) Lorenz curve. For vectors with the same total sum, $\mathbf{u} < \mathbf{v}$ if and only if the Lorenz curve of $\mathbf{u}$ lies below that of $\mathbf{v}$.

Definition 7.8 (Lorenz curve [157]). Given $\mathbf{u} \in \mathbb{R}^d$, its (concave) Lorenz curve is the concave polygonal curve passing through points $(k, s_k^\downarrow(\mathbf{u}))$, for $k = 0, \dots, d$. This curve is concave because its successive slopes are $u_1^\downarrow, \dots, u_d^\downarrow$, which form a non-increasing sequence.

A similar definition could be given using the increasing cumulative sums $s_k^\uparrow(\mathbf{u})$, which would yield a convex polygonal curve. Since majorization is formulated throughout this manuscript in terms of Eq. (7.2), we work with concave Lorenz curves in what follows. Figure 7.1 illustrates Lorenz curves for both comparable and incomparable vectors.

7.1.4 The Shannon entropy: a weaker notion of disorder

A standard real-valued measure of disorder on probability vectors is the Shannon entropy, often simply called entropy. As such, it quantifies the uncertainty associated with a probability distribution and can be interpreted as a measure of how spread out the distribution is.

Definition 7.9 (Shannon entropy [208]). Given $\mathbf{u} \in \mathcal{P}_d$, we define its Shannon entropy $H(\mathbf{u})$ as

$$H(\mathbf{u}) := - \sum_{i=1}^d u_i \log u_i, \quad (7.23)$$

with the convention $0 \log 0 = 0$.

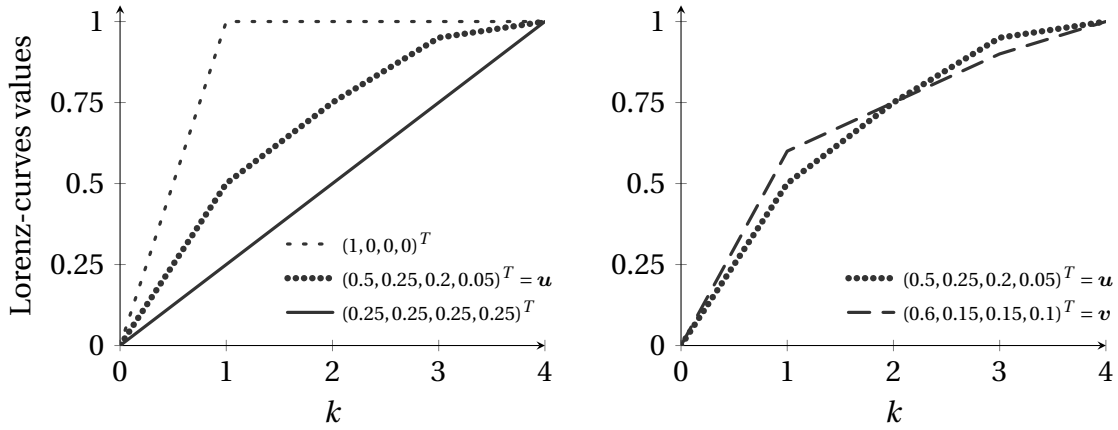


Figure 7.1: Lorenz curves for different choices of probability vectors in $\mathcal{P}_4$. At each integer $k = 0, \dots, d = 4$, the curve takes the value $s_k^\downarrow(u)$, and consecutive points are joined by straight line segments. For vectors with the same total sum, if one Lorenz curve lies everywhere above another, then the associated vector majorizes the other (as can be deduced from Eq. (7.2)). On the left, we exemplify the general relation of Eq. (7.6), i.e., the comparable case, with the choice $u = (0.5, 0.25, 0.2, 0.05)^T$, the peaked vector $(1, 0, 0, 0)^T$ in light dashed line (majorizing every other vector in $\mathcal{P}_4$), and the uniform vector $(0.25, 0.25, 0.25, 0.25)^T$ in solid line (majorized by every other vector in $\mathcal{P}_4$). On the right, we depict the case of two incomparable vectors in $\mathcal{P}_4$, as illustrated by the crossing of the curves. Indeed, if $u = (0.5, 0.25, 0.2, 0.05)^T$ and $v = (0.6, 0.15, 0.15, 0.1)^T$, we have $s_1^\downarrow(u) < s_1^\downarrow(v)$ together with $s_3^\downarrow(u) > s_3^\downarrow(v)$, which shows that neither vector majorizes the other.

It is straightforward to check that H is minimal on the peaked vectors, namely on the permutations of $(1, 0, \dots, 0)^T$, for which H vanishes. Its maximum is attained at the uniform vector, for which

$$H\left(\left(\frac{1}{d}, \dots, \frac{1}{d}\right)^T\right) = \log d. \quad (7.24)$$

Since the function $\eta : [0, \infty) \rightarrow \mathbb{R}$ defined by $\eta(x) = -x \log x$, with $\eta(0) = 0$, is concave, the Shannon entropy is Schur-concave (see the development surrounding Eq. (7.22)), i.e., if $u, v \in \mathcal{P}_d$, then

$$u < v \implies H(u) \geq H(v). \quad (7.25)$$

Thus, majorization implies an ordering by entropy. The converse implication is false in general: knowing only that $H(u) \geq H(v)$ does not imply $u < v$. In this sense, entropy provides a weaker, scalar notion of disorder, whereas majorization gives a preorder on $\mathcal{P}_d$. For example, consider

$$u = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)^T, \quad v = \left(\frac{2}{5}, \frac{2}{5}, \frac{1}{5}\right)^T. \quad (7.26)$$

These vectors are incomparable under majorization, since

$$s_1^\downarrow(u) > s_1^\downarrow(v), \quad s_2^\downarrow(u) < s_2^\downarrow(v). \quad (7.27)$$

Nevertheless, one has $H(\mathbf{u}) < H(\mathbf{v})$. Thus, entropy ranks these two distributions, although majorization does not.

Interestingly, an improvement over Schur-concavity of entropy [see Eq. (7.25)] has been proven, and reads as follows.

Theorem 7.4 (Ho and Verdú [124]). *Let $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d$, and assume that, $\forall i = 1, \dots, d$, $u_i^\downarrow = 0$ implies $v_i^\downarrow = 0$. If $\mathbf{u} < \mathbf{v}$, then*

$$H(\mathbf{u}) - H(\mathbf{v}) \geq D(\mathbf{v}^\downarrow \parallel \mathbf{u}^\downarrow), \quad (7.28)$$

where $D(\mathbf{v}^\downarrow \parallel \mathbf{u}^\downarrow)$ denotes the Kullback–Leibler divergence

$$D(\mathbf{v}^\downarrow \parallel \mathbf{u}^\downarrow) := \sum_{i=1}^d v_i^\downarrow \log \frac{v_i^\downarrow}{u_i^\downarrow}. \quad (7.29)$$

The Kullback–Leibler divergence [49] $D(\mathbf{v} \parallel \mathbf{u})$ (also called relative entropy), defined in Eq. (7.29), can be understood as a measure of “how far” the probability vector $\mathbf{u}$ is from the probability vector $\mathbf{v}$. As can be immediately deduced from its expression, this measure is not a true distance since it does not satisfy the symmetry property, nor the triangle inequality. However, it can be made symmetric with the notion of Jensen–Shannon entropy [154].

7.2 A brief look at the theory of lattices

Before studying the majorization lattice, we recall a few basic notions from order and lattice theory. The goal is not to give a complete account of lattice theory, but only to introduce the terminology needed to discuss partial order, meet, join, down-sets, up-sets, and a lattice in the context of majorization. This section is based on the following introduction to lattices and order [56].

Definition 7.10 (Partially ordered set). Let S be a set. A partial order on S is a binary relation $\leq$ on S that is reflexive, antisymmetric, and transitive. That is, for all $x, y, z \in S$,

$$x \leq x, \quad (7.30)$$

$$x \leq y \text{ and } y \leq x \implies x = y, \quad (7.31)$$

and

$$x \leq y \text{ and } y \leq z \implies x \leq z. \quad (7.32)$$

The pair $(S, \leq)$ is called a partially ordered set, or poset.

If the binary relation is reflexive and transitive but not necessarily antisymmetric, it is called a preorder. This distinction is important for majorization: On arbitrary probability vectors, majorization is only a preorder, because two vectors that differ by a permutation majorize each other. After restricting to non-increasingly arranged probability vectors, this ambiguity disappears, and majorization becomes a partial order.

Definition 7.11 (Down-set and up-set). Let $(S, \leq)$ be a partially ordered set. A subset $D \subseteq S$ is called a down-set if, whenever $x \in D$ and $y \leq x$, one also has $y \in D$. Similarly, a subset $U \subseteq S$ is called an up-set if, whenever $x \in U$ and $x \leq y$, one also has $y \in U$.

For any $x \in S$, the set of all elements below x is a down-set, while the set of all elements above x is an up-set.

Before defining the so-called meet and join operations, it is useful to recall the notions of lower and upper bounds. Let $(S, \leq)$ be a partially ordered set and let $x, y \in S$. An element $z \in S$ is a lower bound of x and y if

$$z \leq x, \quad \text{and} \quad z \leq y. \quad (7.33)$$

Similarly, z is an upper bound of x and y if

$$x \leq z, \quad \text{and} \quad y \leq z. \quad (7.34)$$

There may be many lower or upper bounds, or none at all. When a greatest lower bound or a least upper bound exists, it is unique.

Definition 7.12 (Meet). Let $(S, \leq)$ be a partially ordered set and let $x, y \in S$. The meet of x and y , denoted $x \wedge y$, is the greatest lower bound of x and y , if it exists. Thus,

$$x \wedge y \leq x, \quad x \wedge y \leq y, \quad (7.35)$$

and for every lower bound z of x and y , one has

$$z \leq x \wedge y. \quad (7.36)$$

Definition 7.13 (Join). Let $(S, \leq)$ be a partially ordered set and let $x, y \in S$. The join of x and y , denoted $x \vee y$, is the least upper bound of x and y , if it exists. Thus,

$$x \leq x \vee y, \quad y \leq x \vee y, \quad (7.37)$$

and for every upper bound z of x and y , one has

$$x \vee y \leq z. \quad (7.38)$$

The definition of a lattice then follows naturally.

Definition 7.14 (Lattice). Let $(S, \leq)$ be a non-empty partially ordered set. If, for all $x, y \in S$, a greatest lower bound $x \wedge y$ and a least upper bound $x \vee y$ exist, then $(S, \leq, \wedge, \vee)$ is a lattice.

Thus, a lattice is a partially ordered set in which any two elements can be combined in two ways: their meet is the best approximation from below, and their join is the best approximation from above.

The greatest lower bound and the least upper bound can be straightforwardly generalized to sets of more than two elements. Given $P \subseteq S$, we write $\bigwedge P$ the greatest lower bound (that is, the meet) of all elements of P , and $\bigvee P$ the least upper bound (that is, the join) of all elements of P .

Definition 7.15 (Complete lattice). Let $(S, \leq)$ be a non-empty partially ordered set. If, for every subset $P \subseteq S$, both $\bigwedge P$ and $\bigvee P$ exist, then $(S, \leq, \wedge, \vee)$ is a complete lattice.

In particular, taking $P = S$ shows that a complete lattice has both a least element, namely $\bigwedge S$, and a greatest element, namely $\bigvee S$.

The meet and join operations are not arbitrary binary operations; they satisfy several algebraic identities.

Theorem 7.5. Let $(S, \leq, \wedge, \vee)$ be a lattice. The operations $\wedge$ and $\vee$ satisfy, for all $a, b, c \in S$, the associative, commutative, idempotency, and absorption laws, i.e.,

- $(a \vee b) \vee c = a \vee (b \vee c); (a \wedge b) \wedge c = a \wedge (b \wedge c).$
- $a \vee b = b \vee a; a \wedge b = b \wedge a.$
- $a \vee a = a; a \wedge a = a.$
- $a \vee (a \wedge b) = a; a \wedge (a \vee b) = a.$

Conversely, these identities can be used to define lattices algebraically.

Some lattices satisfy additional compatibility properties between meet and join. Two important such properties are distributivity and modularity. Distributivity is stronger than modularity, and both properties can fail in general lattices.

Definition 7.16 (Distributive lattice). A lattice $(S, \leq, \wedge, \vee)$ is distributive if, for all $a, b, c \in S$,

$$a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c). \quad (7.39)$$

Definition 7.17 (Modular lattice). A lattice $(S, \leq, \wedge, \vee)$ is modular if, for all $a, b, c \in S$ such that $c \leq a$,

$$a \wedge (b \vee c) = (a \wedge b) \vee c. \quad (7.40)$$

Since every distributive lattice is modular (which is easily deduced from Eqs. (7.39) and (7.40)), showing that a lattice is not modular immediately implies that it is not distributive.

7.3 Theory of the majorization lattice

The preceding section explains why we restrict the majorization order to non-increasingly arranged probability vectors. Indeed, on this set, majorization is antisymmetric and hence defines a genuine partial order. Studying the lattice structure of non-increasingly arranged probability vectors, ordered by majorization, actually reveals some of the complexity hidden by this order and helps clarify the structure of majorization.

7.3.1 General definitions

First, we denote by $\mathcal{P}_d^\downarrow$ the set of non-increasingly ordered probability vectors, i.e.,

$$\mathcal{P}_d^\downarrow := \{u \in \mathcal{P}_d \mid u_i \geq u_{i+1}, \quad i = 1, \dots, d-1\}. \quad (7.41)$$

Thus, according to Definition 7.10, $(\mathcal{P}_d^\downarrow, <)$ is a partially ordered set. Second, we define two sets of interest in the majorization lattice:

- The *majorized set* of u in $\mathcal{P}_d^\downarrow$ is

$$\mathcal{S}_u^< := \{w \in \mathcal{P}_d^\downarrow \mid w < u\}; \quad (7.42)$$

- The *majorizing set* of u in $\mathcal{P}_d^\downarrow$ is

$$\mathcal{S}_u^> := \{w \in \mathcal{P}_d^\downarrow \mid u < w\}. \quad (7.43)$$

In the terminology of Definition 7.11, $\mathcal{S}_u^<$ is a down-set of $\mathcal{P}_d^\downarrow$, while $\mathcal{S}_u^>$ is an up-set of $\mathcal{P}_d^\downarrow$.

Definition 7.18 (Majorization lattice [46]). The lattice $(\mathcal{P}_d^\downarrow, <, \wedge, \vee)$ is called the *majorization lattice*, and the operations $\wedge$ and $\vee$ are defined as follows. For $u, v \in \mathcal{P}_d^\downarrow$:

- The *meet* $u \wedge v$ is the greatest lower bound of u and v with respect to majorization. That is,

$$u \wedge v < u, \quad u \wedge v < v, \quad (7.44)$$

and for every $w \in \mathcal{P}_d^\downarrow$ such that $w < u$ and $w < v$, one has

$$w < u \wedge v; \quad (7.45)$$

- The *join* $u \vee v$ is the least upper bound of u and v with respect to majorization. That is,

$$u < u \vee v, \quad v < u \vee v, \quad (7.46)$$

and for every $w \in \mathcal{P}_d^\downarrow$ such that $u < w$ and $v < w$, one has

$$u \vee v < w. \quad (7.47)$$

It follows from these definitions that we can write the majorized set of the meet $u \wedge v$ as the intersection between the majorized sets of u and v , i.e., $\mathcal{S}_{u \wedge v}^\prec = \mathcal{S}_u^\prec \cap \mathcal{S}_v^\prec$. Similarly, we can write the majorizing set of the join $u \vee v$ as the intersection between the majorizing sets of u and v , i.e., $\mathcal{S}_{u \vee v}^\succ = \mathcal{S}_u^\succ \cap \mathcal{S}_v^\succ$. Notice that the majorization lattice is complete in the sense of Definition 7.15.

If u and v are comparable, e.g., if $u < v$, we find $u \wedge v = u$ and $u \vee v = v$. When u and v are incomparable, the lattice becomes nontrivial; see Figure 7.2 for a graphical depiction of the majorization lattice for two incomparable vectors $u, v \in \mathcal{P}_d^\downarrow$ to grasp some intuition.

7.3.2 Constructing the meet and the join

Let us now present how knowing the components of u and v enables finding the components of the meet $u \wedge v$ and the join $u \vee v$ [46]. The idea is simple: the meet is obtained directly by taking the lower envelope of the Lorenz curves of u and v . The join is subtler: taking the upper envelope of the Lorenz curves of u and v yields the correct cumulative sums only if the corresponding components are in non-increasing order. Otherwise, the resulting pseudo-Lorenz curve must be “concavified”.

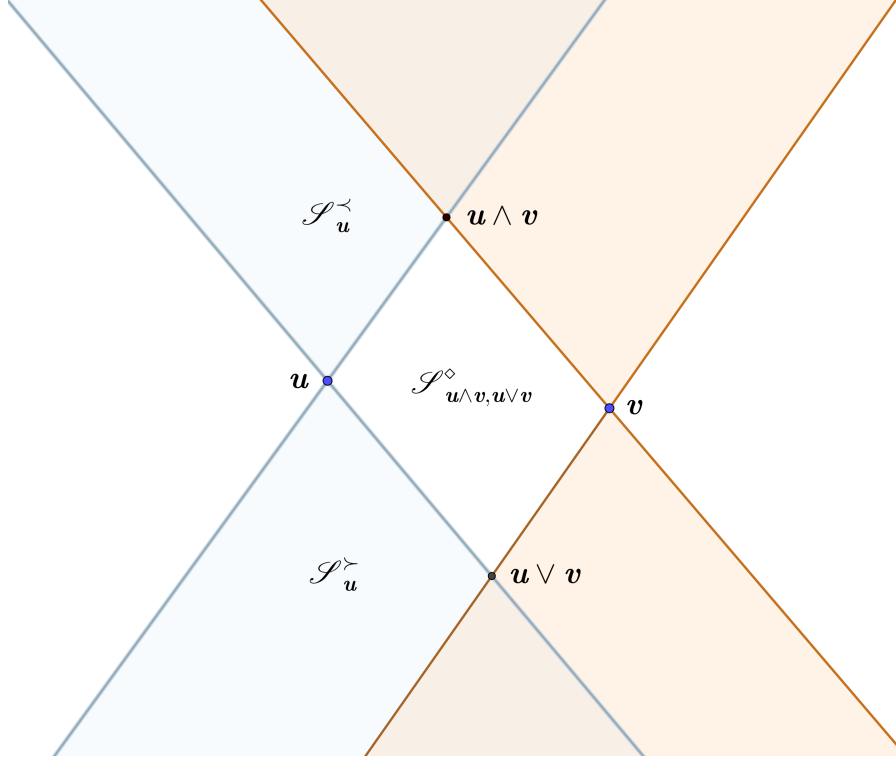


Figure 7.2: Schematic representation of the majorization lattice. Let u and v in $\mathcal{P}_d^\downarrow$. The blue cone up from u contains all vectors in $\mathcal{P}_d^\downarrow$ that are majorized by u (i.e., the majorized set of u , denoted by $\mathcal{S}_u^<$), while the blue cone down from u contains all vectors in $\mathcal{P}_d^\downarrow$ that majorize u (i.e., the majorizing set of u , denoted by $\mathcal{S}_u^>$). Similarly, the orange cone up from v contains all vectors in $\mathcal{P}_d^\downarrow$ that are majorized by v , while the orange cone down from v contains all vectors in $\mathcal{P}_d^\downarrow$ that majorize v . Note as an important caveat that, in order to match the usual schematic representation of the majorization lattice having the thermodynamic arrow of time pointing upwards [149], we call an “up cone” the actual down-set, whereas we call a “down-cone” the actual up-set. The meet $u \wedge v$ is located at the intersection of the two up cones, whereas the join $u \vee v$ is located at the intersection of the two down cones. Since the up cones contain more disordered (in the sense of majorization) elements, the meet $u \wedge v$ stands for the most ordered element among those that are more disordered than both u and v . Conversely, since the down cones contain more ordered elements, the join $u \vee v$ stands for the most disordered element among those that are more ordered than both u and v . The majorization diamond $\mathcal{S}_{u \wedge v, u \vee v}^\diamond$, introduced in Section 9.3, corresponds to the intersection of the cone down from $u \wedge v$ and the cone up from $u \vee v$ (see Eq. (9.8)).

More precisely, for the meet, since $\mathbf{u} \wedge \mathbf{v} < \mathbf{u}$ and $\mathbf{u} \wedge \mathbf{v} < \mathbf{v}$, we write, using Eq. (7.2),

$$s_k^\downarrow(\mathbf{u} \wedge \mathbf{v}) \leq \min(s_k^\downarrow(\mathbf{u}), s_k^\downarrow(\mathbf{v})), \quad \forall k = 1, \dots, d-1, \quad (7.48)$$

and saturate it for all $k = 1, \dots, d-1$, giving its components when subtracting the $(k-1)$ th cumulative sum from the k th, that is,

$$(\mathbf{u} \wedge \mathbf{v})_k = \min(s_k^\downarrow(\mathbf{u}), s_k^\downarrow(\mathbf{v})) - \min(s_{k-1}^\downarrow(\mathbf{u}), s_{k-1}^\downarrow(\mathbf{v})), \quad k = 1, \dots, d, \quad (7.49)$$

which is a vector belonging to $\mathcal{P}_d^\downarrow$ by construction. Remember that $s_d(\mathbf{u}) = s_d(\mathbf{v}) = 1$, and $s_0(\mathbf{u}) = s_0(\mathbf{v}) = 0$. The left panel of Figure 7.3 illustrates the Lorenz curve of the meet.

For the join, the reasoning follows the same path but with an important caveat; since $\mathbf{u} < \mathbf{u} \vee \mathbf{v}$ and $\mathbf{v} < \mathbf{u} \vee \mathbf{v}$, we write similarly, using Eq. (7.2),

$$\max(s_k^\downarrow(\mathbf{u}), s_k^\downarrow(\mathbf{v})) \leq s_k^\downarrow(\mathbf{u} \vee \mathbf{v}), \quad \forall k = 1, \dots, d-1. \quad (7.50)$$

If we formally saturate these inequalities for all $k = 1, \dots, d-1$, we obtain the intermediate vector $\beta(\mathbf{u}, \mathbf{v})$. However, doing so leads to a vector that does not always belong to $\mathcal{P}_d^\downarrow$ (since it is not always in non-increasing order, see the left panel of Figure 7.3 for an illustration), which implies that it cannot always be the join $\mathbf{u} \vee \mathbf{v}$. Nonetheless, because this intermediate vector is useful in the development for finding the join, we write it explicitly, subtracting its $(k-1)$ th cumulative sum from its k th, that is

$$\beta(\mathbf{u}, \mathbf{v})_k = \max(s_k^\downarrow(\mathbf{u}), s_k^\downarrow(\mathbf{v})) - \max(s_{k-1}^\downarrow(\mathbf{u}), s_{k-1}^\downarrow(\mathbf{v})), \quad k = 1, \dots, d. \quad (7.51)$$

As mentioned, two situations occur: either $\beta(\mathbf{u}, \mathbf{v})$ is a non-increasingly ordered vector³ and $\mathbf{u} \vee \mathbf{v} = \beta(\mathbf{u}, \mathbf{v})$, or $\beta(\mathbf{u}, \mathbf{v})$ is not an ordered vector and an algorithm is used to obtain $\mathbf{u} \vee \mathbf{v}$ from $\beta(\mathbf{u}, \mathbf{v})$. At first, one might try to reorder $\beta(\mathbf{u}, \mathbf{v})$ non-increasingly and take $\beta(\mathbf{u}, \mathbf{v})^\downarrow$ as the join. However, this generally gives an upper bound that is too large in majorization order. More precisely, the true join satisfies

$$\mathbf{u} \vee \mathbf{v} < \beta(\mathbf{u}, \mathbf{v})^\downarrow, \quad (7.52)$$

so $\beta(\mathbf{u}, \mathbf{v})^\downarrow$ is more concentrated than necessary. To find the actual least upper bound, we use the following algorithm to correct “just enough” where concavity breaks in Lorenz coordinates, that is, we use a linear correction where needed. This gives a non-increasingly ordered vector, which is the least upper bound. The right panel of Figure 7.3 illustrates the discrepancies between the different curves in Lorenz coordinates. We now detail the complete construction of the join.

³This is always the case for $d = 2$, which is a trivial lattice since all vectors are comparable in this situation, and also for $d = 3$.

Let $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$. We start by introducing singleton blocks $I_i = \{i\}$ with $i = 1, \dots, d$, and attach to each block a value $h_i = \beta(\mathbf{u}, \mathbf{v})_i$. Hence, we begin with $\mathbf{h} = \beta(\mathbf{u}, \mathbf{v})$. After the first step, we use Greek indices α to label the current blocks, since these blocks may contain multiple coordinates. At every step, h_α denotes the constant slope assigned to the current block I_α . Now, we scan the list of blocks from left to right and, as soon as $h_\alpha < h_{\alpha+1}$, we merge the two blocks I_α and $I_{\alpha+1}$. The value of the new block is the weighted average of the two block values, i.e.,

$$h_{\text{new}} = \frac{|I_\alpha| h_\alpha + |I_{\alpha+1}| h_{\alpha+1}}{|I_\alpha| + |I_{\alpha+1}|}. \quad (7.53)$$

We then relabel the current blocks consecutively. After merging, we still need to check with the previous block because the new value may have created a violation with the left block. We continue until

$$h_1 \geq \dots \geq h_r, \quad (7.54)$$

for all blocks (we suppose there are $r \leq d$ at the end of the process). Then, the join is obtained by setting

$$(\mathbf{u} \vee \mathbf{v})_i = h_\alpha, \quad \forall i \in I_\alpha. \quad (7.55)$$

This is the pool-adjacent-violators algorithm [58] applied to the slopes of the pseudo-Lorenz curve. Equivalently, it constructs the least concave majorant of the pseudo-Lorenz curve of $\beta(\mathbf{u}, \mathbf{v})$. Notice that Ref. [46] proposes an alternative way of constructing $\mathbf{u} \vee \mathbf{v}$ from $\beta(\mathbf{u}, \mathbf{v})$.

7.3.3 Properties

We now present some structural properties of the majorization lattice. In particular, although it has well-defined meet and join operations, it does not satisfy some of the stronger identities that hold in special classes of lattices.

There exist different pairs $(\mathbf{u}, \mathbf{v})$ and $(\mathbf{u}', \mathbf{v}')$ such that

$$\mathbf{u} \wedge \mathbf{v} = \mathbf{u}' \wedge \mathbf{v}' \quad \text{and} \quad \mathbf{u} \vee \mathbf{v} = \mathbf{u}' \vee \mathbf{v}'. \quad (7.56)$$

This occurs, for example, by choosing $\mathbf{u} = \mathbf{u}' = (0.6, 0.15, 0.15, 0.1)^T$, $\mathbf{v} = (0.5, 0.25, 0.2, 0.05)^T$, and $\mathbf{v}' = (0.5, 0.275, 0.175, 0.05)^T$. These vectors also provide a counterexample to modularity, and therefore to distributivity, of the majorization lattice [118].

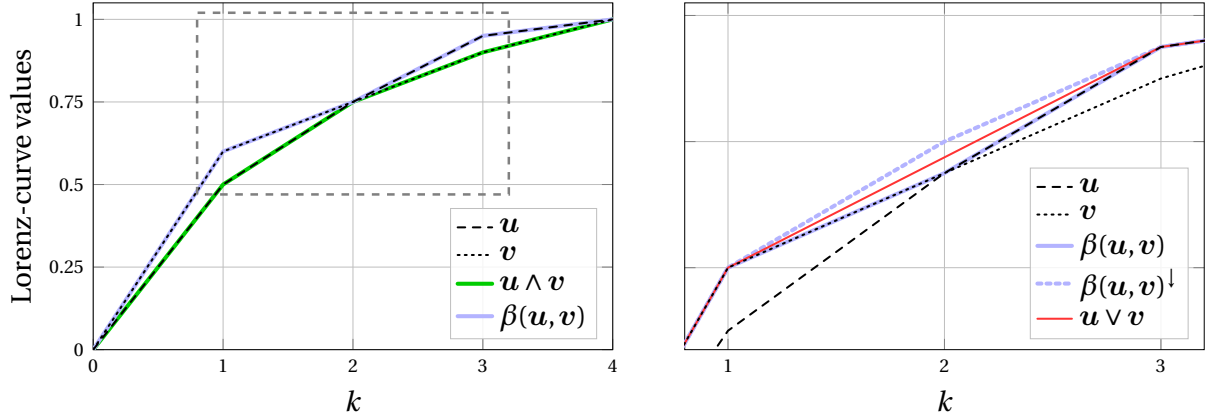


Figure 7.3: Lorenz curves associated with the meet and join constructions for the incomparable vectors $u = (0.5, 0.25, 0.2, 0.05)^T$ and $v = (0.6, 0.15, 0.15, 0.1)^T$. In the left panel, we depict the Lorenz curves of the original vectors (in dotted and dashed black lines) together with that of their meet $u \wedge v = (0.5, 0.25, 0.15, 0.1)^T$ in green, and the pseudo-Lorenz curve of the intermediate vector $\beta(u, v) = (0.6, 0.15, 0.2, 0.05)^T$ in blue. We observe that, while the Lorenz curve of the meet is concave, the pseudo-Lorenz curve of the intermediate vector is not concave, and hence it is not its actual Lorenz curve. To resolve this issue, the intermediate vector is reordered (its actual Lorenz curve is in a dotted blue line in the right panel). In the right panel, we zoom on the region where concavity fails for $\beta(u, v)$ (see the dotted rectangle in the left panel) and compare the pseudo-Lorenz curve of $\beta(u, v)$ in solid blue line, the Lorenz curve of its reordered version $\beta(u, v)^\downarrow = (0.6, 0.2, 0.15, 0.05)^T$ in dotted blue line, and that of the join $u \vee v = (0.6, 0.175, 0.175, 0.05)^T$ in solid red line, obtained by “concavifying” the pseudo-Lorenz curve of $\beta(u, v)$. Since the Lorenz curve of $\beta(u, v)^\downarrow$ lies above that of the join, we have $u \vee v < \beta(u, v)^\downarrow$. Thus, $\beta(u, v)^\downarrow$ is an upper bound, but it is not the least one.

Proposition 7.1. *The majorization lattice is non-modular and, thus, non-distributive.*

Proof. Choose

$$u = u' = (0.6, 0.15, 0.15, 0.1)^T, \quad v = (0.5, 0.25, 0.2, 0.05)^T, \quad v' = (0.5, 0.275, 0.175, 0.05)^T. \quad (7.57)$$

Since $v < v'$, we require that (see Definition 7.17)

$$v \wedge (u \vee v') = (v \wedge u) \vee v'. \quad (7.58)$$

However, in this example,

$$v \wedge (u \vee v') = v \neq v' = (v \wedge u) \vee v'. \quad (7.59)$$

Thus, the lattice is not modular, and therefore it is not distributive. $\square$

When the majorization lattice was introduced in information theory, two entropic properties were proposed [46].

Proposition 7.2 (Subadditivity of entropy on the majorization lattice). *For $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$, we have*

$$H(\mathbf{u} \wedge \mathbf{v}) \leq H(\mathbf{u}) + H(\mathbf{v}). \quad (7.60)$$

Proof. We present an alternative, shorter proof than the original from Ref. [46]. When comparing probability vectors of different lengths, we pad the shorter ones with zeros. This does not affect either majorization relations or Shannon entropy. Let (U, V) be any coupling of the distributions $\mathbf{u}$ and $\mathbf{v}$, and let p_{UV} be its joint distribution. Its marginals are $p_U = \mathbf{u}$ and $p_V = \mathbf{v}$. Since p_{UV} is a refinement of each marginal distribution, we have

$$p_{UV}^\downarrow \prec p_U = \mathbf{u}, \quad p_{UV}^\downarrow \prec p_V = \mathbf{v}. \quad (7.61)$$

Hence, p_{UV} is a lower bound of both $\mathbf{u}$ and $\mathbf{v}$ in majorization order. By the definition of the meet, we can thus write

$$p_{UV}^\downarrow \prec \mathbf{u} \wedge \mathbf{v}. \quad (7.62)$$

Using Schur-concavity of entropy (see Eq. (7.25)), we get

$$H(p_{UV}) \geq H(\mathbf{u} \wedge \mathbf{v}). \quad (7.63)$$

On the other hand, by the subadditivity of Shannon entropy [49],

$$H(p_{UV}) = H(U, V) \leq H(U) + H(V) = H(\mathbf{u}) + H(\mathbf{v}). \quad (7.64)$$

Combining the last two inequalities gives

$$H(\mathbf{u} \wedge \mathbf{v}) \leq H(\mathbf{u}) + H(\mathbf{v}), \quad (7.65)$$

as desired. □

Proposition 7.3 (Supermodularity of entropy on the majorization lattice). *For $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$, we have*

$$H(\mathbf{u}) + H(\mathbf{v}) \leq H(\mathbf{u} \wedge \mathbf{v}) + H(\mathbf{u} \vee \mathbf{v}). \quad (7.66)$$

In Ref. [223], Proposition 7.3 is obtained from the stronger majorization relation

$$(\mathbf{u} \wedge \mathbf{v}) \oplus (\mathbf{u} \vee \mathbf{v}) \prec \mathbf{u} \oplus \mathbf{v}, \quad (7.67)$$

where $\oplus$ denotes the direct sum. By Schur-concavity of entropy,

$$H\left(\frac{(\mathbf{u} \wedge \mathbf{v}) \oplus (\mathbf{u} \vee \mathbf{v})}{2}\right) \geq H\left(\frac{\mathbf{u} \oplus \mathbf{v}}{2}\right). \quad (7.68)$$

Using the following entropy identity, valid for every $\mathbf{a}, \mathbf{b} \in \mathcal{P}_d^\downarrow$,

$$H\left(\frac{\mathbf{a} \oplus \mathbf{b}}{2}\right) = \frac{1}{2}H(\mathbf{a}) + \frac{1}{2}H(\mathbf{b}) + \log 2, \quad (7.69)$$

we recover Eq. (7.66). The majorization lattice also admits a distance defined in terms of the Shannon entropy H and the join operation.

Definition 7.19 (Entropic distance in the majorization lattice [45]). For $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$, the entropic distance between $\mathbf{u}$ and $\mathbf{v}$ in the majorization lattice is defined as

$$d_H(\mathbf{u}, \mathbf{v}) := H(\mathbf{u}) + H(\mathbf{v}) - 2H(\mathbf{u} \vee \mathbf{v}). \quad (7.70)$$

If $\mathbf{u}$ and $\mathbf{v}$ are comparable, e.g., if $\mathbf{u} \prec \mathbf{v}$, we get $d_H(\mathbf{u}, \mathbf{v}) = H(\mathbf{u}) - H(\mathbf{v})$ since $\mathbf{u} \vee \mathbf{v} = \mathbf{v}$. Notice that we could think of defining an analogous distance using the meet instead of the join, such as

$$\tilde{d}_H(\mathbf{u}, \mathbf{v}) := 2H(\mathbf{u} \wedge \mathbf{v}) - H(\mathbf{u}) - H(\mathbf{v}). \quad (7.71)$$

Similarly, if $\mathbf{u}$ and $\mathbf{v}$ are comparable, e.g., if $\mathbf{u} \prec \mathbf{v}$, then $\tilde{d}_H(\mathbf{u}, \mathbf{v}) = H(\mathbf{u}) - H(\mathbf{v})$, since $\mathbf{u} \wedge \mathbf{v} = \mathbf{u}$. However, $\tilde{d}_H(\mathbf{u}, \mathbf{v})$ does not satisfy all the requirements of a distance measure; in particular, it can fail the triangle inequality. Indeed,

$$\tilde{d}_H(\mathbf{u}, \mathbf{v}) \not\leq \tilde{d}_H(\mathbf{u}, \mathbf{u} \vee \mathbf{v}) + \tilde{d}_H(\mathbf{u} \vee \mathbf{v}, \mathbf{v}), \quad (7.72)$$

since it can be rewritten as

$$H(\mathbf{u}) + H(\mathbf{v}) \geq H(\mathbf{u} \wedge \mathbf{v}) + H(\mathbf{u} \vee \mathbf{v}), \quad (7.73)$$

which is forbidden for incomparable $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$ in general, due to Proposition 7.3. In contrast, d_H defines a proper distance on the majorization lattice, as shown in Ref. [45]. The supermodularity of entropy can also be rephrased using the entropic distance d_H as

$$d_H(\mathbf{u} \wedge \mathbf{v}, \mathbf{u}) \geq d_H(\mathbf{v}, \mathbf{u} \vee \mathbf{v}). \quad (7.74)$$

In Chapter 8, we make use of the majorization lattice notions of meet and join to propose two new optimal protocols for converting bipartite pure states, whereas in Chapter 9, we investigate new geometric aspects of the majorization lattice in order to strengthen entropic inequalities.

8 Probabilistic pure-state conversion on the majorization lattice

The content of this chapter is based on the following published work: Serge Deside, Matthieu Arnhem, Célia Griffet, and Nicolas J. Cerf. “Probabilistic pure state conversion on the majorization lattice”. Physical Review Research 6, 023156, May 2024 [62].

In this chapter, we show that the majorization lattice provides an efficient framework for characterizing the allowed transformations of pure entangled states under local operations and classical communication (LOCC). The underlying notions of meet $\wedge$ and join $\vee$ in the majorization lattice lead us to define, respectively, the optimal common resource and optimal common product states. Based on these two states, we introduce two optimal probabilistic protocols for the (single-copy) conversion of incomparable bipartite pure states, which we name greedy and thrifty. Both protocols reduce to Vidal’s protocol [233] when the initial and final states are comparable. When they are incomparable, however, the thrifty protocol is superior to the greedy protocol in the sense that it yields a more entangled residual state upon failure, while both protocols yield the target state with the same optimal success probability upon success. Finally, we consider the generalization of these protocols to entanglement transformations involving multiple initial or final states.

8.1 Introduction

As discussed in Chapter 2, quantum entanglement has long been recognized as a necessary resource in many quantum information protocols [21, 22, 133] and is often regarded as the paramount quantum resource [42]. In this context, separable states are defined as resource-free states and LOCC are viewed as free operations, that is, operations that map separable states onto separable states (hence, do not create any resource). An important aspect of such a resource theory of entanglement concerns the study of allowed entanglement transformations, i.e., the conversions between entangled states of a system shared between several parties that can be achieved using LOCC only. Note that the conversions of bipartite pure states are asymptotically (in the limit of many copies) always allowed with some yield [19]. Instead, here, we focus on the case of single-copy transformations between bipartite pure states. The most notable result on this subject, due to Nielsen [171], consists in a simple characterization of deterministic bipartite pure state conversions using LOCC by means of majorization theory. A related result, due to Vidal [233], further extends this characterization to probabilistic transformations.

As presented in Chapter 7, the theory of majorization allows one to compare probability distributions in terms of intrinsic disorder. These majorization relations play a central role in entanglement theory, as implied by Nielsen's theorem (see Theorem 2.1), i.e., if Alice and Bob share a bipartite pure state $|\psi\rangle_{AB}$,

$$|\psi\rangle_{AB} = \sum_{i=1}^d \sqrt{\lambda_{\psi}^{(i)}} |i\rangle_A |i\rangle_B. \quad (8.1)$$

It is convertible into another pure state

$$|\phi\rangle_{AB} = \sum_{i=1}^d \sqrt{\lambda_{\phi}^{(i)}} |i\rangle_A |i\rangle_B \quad (8.2)$$

if and only if the two Schmidt vectors λ_{ψ} , λ_{ϕ} obey the following majorization relation,

$$\lambda_{\psi} < \lambda_{\phi}. \quad (8.3)$$

Notice that, by an abuse of language, we write $|\psi\rangle_{AB} < |\phi\rangle_{AB}$, and say that $|\psi\rangle_{AB}$ is majorized by $|\phi\rangle_{AB}$, whenever $\lambda_{\psi} < \lambda_{\phi}$. Protocols achieving such a transformation have been discussed, for example, in Refs. [141, 230]. The characterization of entanglement transformations via majorization relations has been generalized to probabilistic LOCC transformations by Vidal (see Theorem 2.2), and an optimal protocol achieving the desired probabilistic transformation was also provided (it will be detailed in Section 8.3). Optimality refers here to a protocol that reaches the exact state $|\phi\rangle_{AB}$ with the highest possible success probability.

In this chapter, we exploit the *majorization lattice* (presented in Section 7.3) in order to revisit the probabilistic entanglement transformations between bipartite pure states. Our analysis builds upon the two central elements of a lattice theory, the so-called *meet* and *join* (see Figure 7.2), and yields two corresponding optimal protocols (see Figure 8.1). Considering an entanglement lattice, where each node is a bipartite pure state and the lattice structure emerges from majorization relations, we associate the meet and the join with two specific states that we call, respectively, the optimal common resource (OCR)¹ and optimal common product (OCP) states. We first build a protocol using the OCP state and show that it is akin to Vidal's optimal protocol. More interestingly, we then build a second protocol making use of the OCR state and prove that, while being again optimal, it better preserves average entanglement. Both protocols trivially reduce to Vidal's in the special case where $|\psi\rangle$ and $|\phi\rangle$ are comparable states in the sense of majorization theory, so we focus on the interesting case of incomparable states.

¹This term was first coined in Ref. [110] but with no reference to the majorization lattice.

The protocol based on the OCP state can be viewed as a *greedy* protocol. In the context of optimization, a greedy algorithm evolves towards the solution by choosing the local optimal move at each stage. In analogy, our greedy protocol favors the immediate gain of a deterministic move and postpones the probabilistic move (see red arrows in Figure 8.1). Our second protocol based on the OCR state can be understood, in contrast, as a *thrifty* protocol. A thrifty optimization algorithm prefers not to choose immediate gain at each stage. Here, our thrifty protocol indeed tolerates starting with a probabilistic move (see green arrows in Figure 8.1). The main results of this chapter consist in Theorems 8.1 and 8.2, which prove, respectively, the optimality and better average entanglement preservation of the thrifty protocol.

This chapter is organized as follows. In Section 8.2, we first introduce the majorization lattice and, specifically for bipartite entanglement, the OCR and OCP states. Then, in Section 8.3, we present Vidal's theorem for the probabilistic conversion of bipartite pure states as well as the corresponding optimal protocol. This provides us with the tools for introducing the greedy and thrifty protocols in Section 8.4. We further establish in Section 8.5 the generalization of the greedy and thrifty protocols to an arbitrary number of initial or final states. Finally, we give our conclusions in Section 8.6.

8.2 From the majorization lattice to the entanglement lattice

Let us briefly recall the majorization lattice introduced in Definition 7.18, that is, $(\mathcal{P}_d^\downarrow, <, \wedge, \vee)$. The meet of two vectors $u, v \in \mathcal{P}_d^\downarrow$, denoted $u \wedge v$, is the sole element of $\mathcal{P}_d^\downarrow$ such that, for every $w \in \mathcal{P}_d^\downarrow$ with $w < u$ and $w < v$, we have $w < u \wedge v$. Analogously, the join of two vectors $u, v \in \mathcal{P}_d^\downarrow$, denoted $u \vee v$, is the sole element of $\mathcal{P}_d^\downarrow$ such that, for every $w \in \mathcal{P}_d^\downarrow$ with $u < w$ and $v < w$, we have $u \vee v < w$. Thus, it implies $u \wedge v < u < u \vee v$ and $u \wedge v < v < u \vee v$.

In the following, we will consider an entanglement lattice whose elements are bipartite pure states (see Figure 8.1). Since the entanglement of a bipartite pure state is uniquely characterized by its Schmidt vector, we may equivalently view the elements of the lattice as probability vectors² in $\mathcal{P}_d^\downarrow$. Hence, according to Nielsen's theorem, a state can be obtained from any state belonging to its majorized set by using only deterministic LOCC. Conversely, a state can be converted into any state of its majorizing set by using only deterministic LOCC. In other words, the lattice structure translates the allowed entanglement transformations.

²In what follows, when we say that one bipartite pure state majorizes another, we mean that the corresponding relation holds between their Schmidt vectors.

Accordingly, the meet of two (or more) states denotes the least entangled state that can still be deterministically converted into any one of them. We refer to it as the optimal common resource (OCR)¹ state. Reciprocally, the join of two (or more) states denotes the most entangled state that can still be deterministically produced from any one of them. We refer to it as the optimal common product (OCP) state, where “product” refers to an output of the conversion process. The OCP and OCR states will be used in Section 8.4 in order to construct the greedy and thrifty protocols, respectively, as pictured in Figure 8.1.

Instructively, the optimality of the OCP and OCR states can be rephrased as follows. For ease of notation, from now on, we drop the subscript AB for bipartite states (all pure states are bipartite in the following). Considering two bipartite pure states $|\psi\rangle$ and $|\phi\rangle$, any state $|\tau\rangle$ that is majorized by both $|\psi\rangle$ and $|\phi\rangle$ (i.e., any state $|\tau\rangle$ lying inside the intersection of the two up cones emerging from $|\psi\rangle$ and $|\phi\rangle$ in Figure 8.1) must also be majorized by the OCR state $|\psi \wedge \phi\rangle$. Thus, according to Nielsen’s theorem, all “resource states” (i.e., states $|\tau\rangle$ that are convertible into either $|\psi\rangle$ or $|\phi\rangle$ via a deterministic LOCC) can also produce the OCR state $|\psi \wedge \phi\rangle$. Since the latter is itself convertible into either $|\psi\rangle$ or $|\phi\rangle$, it is the resource state that requires the lowest amount of resource. Conversely, any state $|\sigma\rangle$ that majorizes both $|\psi\rangle$ and $|\phi\rangle$ (i.e., any state $|\sigma\rangle$ lying inside the intersection of the two down cones emerging from $|\psi\rangle$ and $|\phi\rangle$ in Figure 8.1) must necessarily also majorize the OCP state $|\psi \vee \phi\rangle$. This means that all “producible states” (i.e., states $|\sigma\rangle$ that can be deterministically produced with LOCC either from $|\psi\rangle$ or from $|\phi\rangle$) can also be produced from the OCP state $|\psi \vee \phi\rangle$. Since the latter is itself obtainable either from $|\psi\rangle$ or from $|\phi\rangle$, it is the producible state that contains the highest amount of resource.

8.3 Probabilistic state-conversion protocol

When a deterministic state conversion is impossible according to Nielsen’s theorem, the same transformation may sometimes be achieved probabilistically. This is the content of Vidal’s theorem (see Theorem 2.2), which we now recall, before giving its associated protocol.

Let $|\psi\rangle$ be the initial state of a bipartite system shared between Alice and Bob, and let $|\phi\rangle$ be the target state they wish to obtain using only LOCC. We suppose that it is possible to perform the desired transformation from $|\psi\rangle$ to $|\phi\rangle$ with non-vanishing probability, i.e., the number of non-zero coefficients of λ_ψ must be at least as large as that of λ_ϕ (in the following, we consider, without loss of generality, that they are equal)³. Vidal’s theorem states that there exists a probabilistic protocol converting $|\psi\rangle$ into $|\phi\rangle$ with probability p if and only if

$$E_l(\psi) \geq p E_l(\phi), \quad \forall l = 1, \dots, d, \quad (8.4)$$

where $E_l(\psi) := \sum_{i=l}^d \lambda_\psi^{(i)\downarrow}$, as defined in Section 2.2.2.

³If the number of Schmidt coefficients of the initial state is strictly smaller than the number of Schmidt coefficients of the target state, the optimal conversion probability is equal to zero by Theorem 2.2.

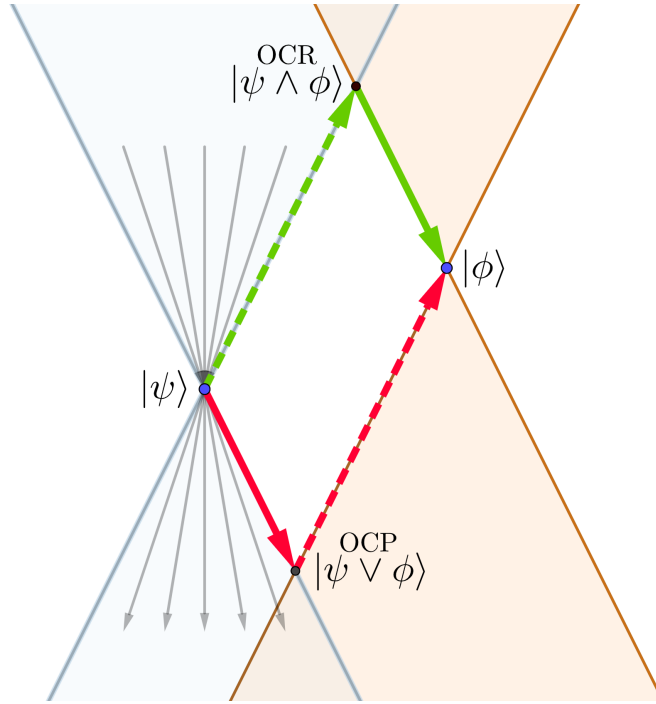


Figure 8.1: Schematic representation of the majorization lattice applied to bipartite entanglement. Let $|\psi\rangle$ and $|\phi\rangle$ be two bipartite pure states that are assumed to be incomparable. Any state lying in the up cone of $|\psi\rangle$ is convertible into $|\psi\rangle$ via deterministic LOCC, while any state in the down cone of $|\psi\rangle$ is reachable from $|\psi\rangle$ via deterministic LOCC (this is pictured with a collection of grey arrows pointing downwards). The same holds true, of course, for the cones associated with $|\phi\rangle$. The OCR state $|\psi \wedge \phi\rangle$ is located at the intersection of the two up cones and can be understood as the least entangled state capable of deterministically producing either $|\psi\rangle$ or $|\phi\rangle$. In contrast, the OCP state $|\psi \vee \phi\rangle$ is located at the intersection of the two down cones and can be viewed as the most entangled state that both $|\psi\rangle$ and $|\phi\rangle$ are capable to produce deterministically. We highlight the greedy (red) and thrifty (green) protocols as discussed in Section 8.4 for converting $|\psi\rangle$ into $|\phi\rangle$ (bold arrows correspond to deterministic LOCC transformations and dashed arrows to probabilistic LOCC transformations). The greedy protocol (in red) starts with a deterministic LOCC towards the OCP state and postpones the probabilistic LOCC. In contrast, the thrifty protocol (in green) starts with a probabilistic LOCC towards the OCR state before making the deterministic LOCC.

Thus, the optimal (maximum) probability with which Alice and Bob can perform the desired transformation is given by

$$p_{\max} = \min_{l \in \{1, \dots, d\}} \frac{E_l(\psi)}{E_l(\phi)}. \quad (8.5)$$

Note that $p_{\max} \leq 1$ because when $l = 1$, we have $E_1(\psi) = E_1(\phi) = 1$. Vidal described a protocol achieving this optimal probability in Ref. [233], which we review hereafter.

First, Alice and Bob apply a deterministic LOCC in order to transform $|\psi\rangle$ into an intermediate state $|\chi\rangle$ which maximizes the fidelity with respect to the target state $|\phi\rangle$ while being deterministically reachable from $|\psi\rangle$ using only LOCC [235], namely,

$$|\chi\rangle = \underset{|\alpha\rangle : |\psi\rangle \xrightarrow{\text{LOCC}} |\alpha\rangle}{\operatorname{argmax}} |\langle\alpha|\phi\rangle|^2. \quad (8.6)$$

Second, Alice performs a two-outcome measurement on her share of the system, leading, with some probability $p_{\max}$, to the target state $|\phi\rangle$ (in which case we say that the protocol has succeeded) or, with probability $1 - p_{\max}$, to a state denoted as $|\xi\rangle$ (in which case we say that the protocol has failed). We show below that this residual state $|\xi\rangle$ always possesses a strictly smaller number of non-vanishing Schmidt coefficients than $|\phi\rangle$, hence the conversion from $|\xi\rangle$ to $|\phi\rangle$ is fully impossible, even probabilistically. Note that Vidal's conversion protocol, which we simply refer to as a probabilistic LOCC transformation in the following, involves both a deterministic step and a probabilistic step.

The construction of the intermediate state $|\chi\rangle$ works as follows. We make use of the d entanglement monotones defined in Eq. (8.4) in order to define a sequence of ratios r_j 's. First, we define r_1 , which corresponds to the maximum conversion probability from $|\psi\rangle$ to $|\phi\rangle$, as

$$r_1 = \min_{l \in \{1, \dots, d\}} \frac{E_l(\psi)}{E_l(\phi)} := \frac{E_{l_1}(\psi)}{E_{l_1}(\phi)}, \quad (8.7)$$

where l_1 is the value of l that reaches the minimum (it is chosen as the smallest value of l in case of several minima). Then, we define the next ratios r_j 's with $j = 2, 3, \dots, d$ as follows (each r_j is associated with a corresponding l_j)

$$r_j = \min_{l \in \{1, \dots, l_{j-1}-1\}} \frac{E_l(\psi) - E_{l_{j-1}}(\psi)}{E_l(\phi) - E_{l_{j-1}}(\phi)} := \frac{E_{l_j}(\psi) - E_{l_{j-1}}(\psi)}{E_{l_j}(\phi) - E_{l_{j-1}}(\phi)}, \quad (8.8)$$

until we find some value of j , which we call k , satisfying $l_k = 1$. Finally, we define $l_0 = d + 1$. On a side note, it can easily be shown that the sequences of r_j 's and l_j 's satisfy $0 < r_1 < \dots < r_k$ and $l_0 > l_1 > \dots > l_k = 1$ [233].

From the ratios r_j 's, we define the Schmidt coefficients of the intermediate state

$$\lambda_\chi^{(i)} = r_j \lambda_\phi^{(i)}, \quad \text{if } i \in \{l_j, \dots, l_{j-1}-1\}, \quad \text{with } j \in \{1, \dots, k\}, \quad (8.9)$$

which, by construction, satisfies $\lambda_\psi < \lambda_\chi$. Hence, $|\psi\rangle$ is convertible to $|\chi\rangle$ by using a deterministic LOCC.

The generalized measurement performed by Alice is described by the two Kraus operators

$$\hat{M} = \begin{pmatrix} \hat{M}_k & & \\ & \ddots & \\ & & \hat{M}_1 \end{pmatrix}, \quad (8.10)$$

$$\hat{N} = \begin{pmatrix} \sqrt{\hat{\mathbb{I}}_{[l_{k-1}-l_k]} - \hat{M}_k^2} & & \\ & \ddots & \\ & & \sqrt{\hat{\mathbb{I}}_{[l_0-l_1]} - \hat{M}_1^2} \end{pmatrix}, \quad (8.11)$$

with

$$\hat{M}_j = \sqrt{\frac{r_1}{r_j}} \hat{\mathbb{I}}_{[l_{j-1}-l_j]}, \quad \text{for } j \in \{1, \dots, k\}, \quad (8.12)$$

where $\hat{\mathbb{I}}_{[l_{j-1}-l_j]}$ is the identity operator in a $(l_{j-1} - l_j)$ -dimensional Hilbert space. We have, of course, the completeness relation $\hat{M}^\dagger \hat{M} + \hat{N}^\dagger \hat{N} = \hat{\mathbb{I}}_{[d]}$. If Alice obtains the measurement outcome linked to $\hat{M}$, she obtains the target state $|\phi\rangle$. This happens with probability r_1 since we have

$$(\hat{M} \otimes \hat{\mathbb{I}}_{[d]}) |\chi\rangle = \sqrt{r_1} |\phi\rangle. \quad (8.13)$$

Otherwise, if she obtains the outcome linked to $\hat{N}$, she gets the residual state $|\xi\rangle$ such that

$$(\hat{N} \otimes \hat{\mathbb{I}}_{[d]}) |\chi\rangle = \sqrt{1 - r_1} |\xi\rangle. \quad (8.14)$$

It is clear from Eq. (8.11) that $\hat{N}$ possesses at least one vanishing diagonal element, hence $|\xi\rangle$ has strictly less non-vanishing Schmidt coefficients than $|\phi\rangle$. Therefore, it is impossible for Alice and Bob to obtain state $|\phi\rangle$ from $|\xi\rangle$ by using (even probabilistic) LOCC transformations.

Note that in the special case where $\lambda_\psi < \lambda_\phi$, Vidal's protocol reduces to a deterministic ($p_{\max} = 1$) protocol for converting $|\psi\rangle$ into $|\phi\rangle$, which exists as a consequence of Nielsen's theorem.

8.4 Greedy and thrifty protocols on the majorization lattice

We now introduce the two optimal probabilistic state conversion protocols which take advantage of the notions of OCP and OCR states in the majorization lattice (see Figure 8.2 for a diagrammatic representation of both protocols). As usual, Alice and Bob initially share a pure bipartite state $|\psi\rangle$ that they wish to transform into a target state $|\phi\rangle$ by using only LOCC operations. Further, we assume that $|\psi\rangle$ and $|\phi\rangle$ are incomparable.

The protocol passing through the OCP state works as follows. First, Alice and Bob transform $|\psi\rangle$ into the OCP of the initial and target states, denoted as $|\psi \vee \phi\rangle$, via a deterministic LOCC transformation. This conversion is possible as a consequence of Nielsen's theorem since, by definition, $\lambda_\psi < \lambda_{\psi \vee \phi}$. Then, Alice and Bob apply to $|\psi \vee \phi\rangle$ the probabilistic LOCC transformation⁴ presented in Section 8.3 in order to obtain the target state $|\phi\rangle$. This clearly cannot be done deterministically since $\lambda_{\psi \vee \phi} \not\leq \lambda_\phi$ (actually, $\lambda_{\psi \vee \phi} > \lambda_\phi$). We name this protocol *greedy* because the deterministic (rewarding) LOCC transformation is given priority and precedes the probabilistic (less rewarding) LOCC transformation.

Following a similar logic, we define another protocol using the OCR state instead. In this protocol, Alice and Bob first transform $|\psi\rangle$ into the OCR of the initial and target states, denoted as $|\psi \wedge \phi\rangle$, using the probabilistic LOCC transformation⁵ presented in Section 8.3. This transformation can only be made probabilistically because $\lambda_\psi \not\leq \lambda_{\psi \wedge \phi}$ (actually, $\lambda_\psi > \lambda_{\psi \wedge \phi}$). Then, the second phase consists in Alice and Bob transforming $|\psi \wedge \phi\rangle$ into $|\phi\rangle$ via a deterministic LOCC operation, which is obviously possible because $\lambda_{\psi \wedge \phi} < \lambda_\phi$. We name this protocol *thrifty* because the deterministic (rewarding) LOCC transformation is postponed after the probabilistic (less rewarding) LOCC transformation.

Interestingly, we will prove that the greedy and thrifty protocols both reach the optimal conversion probability [the same as that of Vidal's protocol, see Eq. (8.5)]. Furthermore, we will show that the thrifty protocol better preserves entanglement, on average, than the greedy protocol. More precisely, if the thrifty protocol fails, the residual state, which we call $|\nu\rangle$, is more entangled than the residual state of the greedy protocol, called $|\xi\rangle$, in the sense that $\lambda_\nu < \lambda_\xi$, or equivalently $|\nu\rangle \xrightarrow{\text{LOCC}} |\xi\rangle$.

Note that if $|\psi\rangle$ and $|\phi\rangle$ were comparable instead of incomparable states, then the greedy and thrifty protocols would be equal and would coincide with Vidal's conversion protocol from $|\psi\rangle$ to $|\phi\rangle$ (which is probabilistic when $\lambda_\phi < \lambda_\psi$, or even deterministic when $\lambda_\psi < \lambda_\phi$). Hence, we will only be interested in the case of incomparable Schmidt vectors λ_ψ and λ_ϕ , i.e., situations where a deterministic LOCC conversion is impossible in both directions, from $|\psi\rangle$ to $|\phi\rangle$ and from $|\phi\rangle$ to $|\psi\rangle$.

Let us first consider the easiest case of the greedy protocol, which can be brought to Vidal's protocol simply by merging the sequence of two deterministic steps $|\psi\rangle \rightarrow |\psi \vee \phi\rangle \rightarrow |\chi\rangle$ into a single deterministic step $|\psi\rangle \rightarrow |\chi\rangle$. Indeed, in the second phase of the greedy protocol (i.e., the probabilistic LOCC transformation from $|\psi \vee \phi\rangle$ to $|\phi\rangle$), the intermediate state happens to be the same as the intermediate state $|\chi\rangle$ of the transformation from $|\psi\rangle$ to $|\phi\rangle$ in Vidal's protocol, defined in Eq. (8.9).

⁴It consists of a deterministic step from $|\psi \vee \phi\rangle$ to the intermediate state $|\chi\rangle$, followed by a probabilistic step from $|\chi\rangle$ to $|\phi\rangle$.

⁵It consists of a deterministic step from $|\psi\rangle$ to the intermediate state $|\zeta\rangle$, followed by a probabilistic step from $|\zeta\rangle$ to $|\psi \wedge \phi\rangle$.

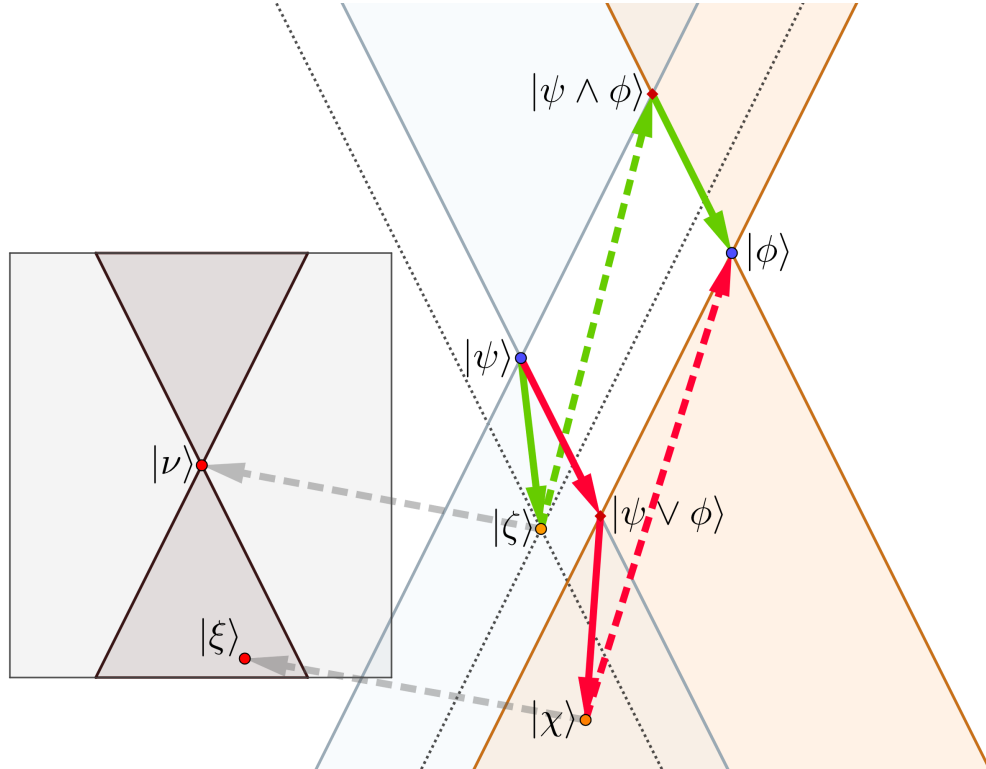


Figure 8.2: Diagrammatic representation of the greedy (red) and thrifty (green) protocols converting $|\psi\rangle$ into $|\phi\rangle$ on the majorization lattice. The probabilistic LOCC transformation of both protocols is itself split into two steps: a deterministic step (represented by a bold arrow) towards an intermediate state (either $|\chi\rangle$ or $|\zeta\rangle$) followed by a purely probabilistic step (represented by a dashed arrow) involving a local two-outcome measurement. If the conversion is successful, the resulting state is majorized by (is more entangled than) the initial state ($|\phi\rangle \prec |\psi \vee \phi\rangle$ and $|\psi \wedge \phi\rangle \prec |\psi\rangle$). The left inset represents the residual states if the conversion has failed, emphasizing that the residual state $|\nu\rangle$ of the thrifty protocol is majorized by (hence, is more entangled than) the residual state $|\xi\rangle$ of the greedy protocol. Note also that $|\zeta\rangle \prec |\chi\rangle$.

This is true because this intermediate state $|\chi\rangle$ has been shown to belong to the down cone of $|\psi \vee \phi\rangle$ in Ref. [26], and is thus deterministically reachable from $|\psi\rangle$. Since, by definition, the intermediate state is the closest state to $|\phi\rangle$ that is deterministically attainable from $|\psi\rangle$ [see Eq. (8.6)], both intermediate states coincide. Hence, the greedy protocol is also optimal.

The rest of this section is devoted to the analysis of the thrifty protocol, which is more involved. Before proving the optimality of the thrifty protocol, let us prove the following rather intuitive lemma.

Lemma 8.1. *The entanglement monotone for the OCR of two bipartite pure states $|\psi\rangle$ and $|\phi\rangle$, $E_l(\psi \wedge \phi)$, is equal to the maximum between the entanglement monotones of the initial state $E_l(\psi)$ and target state $E_l(\phi)$, i.e.,*

$$E_l(\psi \wedge \phi) = \max\{E_l(\psi), E_l(\phi)\}, \quad \forall l \in \{1, \dots, d\}. \quad (8.15)$$

Proof. We start from the definition of the l th entanglement monotone for the OCR state,

$$E_l(\psi \wedge \phi) = \sum_{i=l}^d \lambda_{\psi \wedge \phi}^{(i)}, \quad (8.16)$$

which can be rewritten, by means of the expression of the meet in Eq. (7.49), as

$$E_l(\psi \wedge \phi) = \sum_{i=l}^d \left(\min \left\{ \sum_{j=1}^i \lambda_{\psi}^{(j)}, \sum_{j=1}^i \lambda_{\phi}^{(j)} \right\} - \min \left\{ \sum_{j=1}^{i-1} \lambda_{\psi}^{(j)}, \sum_{j=1}^{i-1} \lambda_{\phi}^{(j)} \right\} \right). \quad (8.17)$$

Since all terms cancel out two by two, except for the first and last terms, we obtain

$$E_l(\psi \wedge \phi) = 1 - \min \left\{ \sum_{j=1}^{l-1} \lambda_{\psi}^{(j)}, \sum_{j=1}^{l-1} \lambda_{\phi}^{(j)} \right\}, \quad (8.18)$$

or equivalently,

$$E_l(\psi \wedge \phi) = \max \left\{ \sum_{j=l}^d \lambda_{\psi}^{(j)}, \sum_{j=l}^d \lambda_{\phi}^{(j)} \right\}, \quad (8.19)$$

$$= \max\{E_l(\psi), E_l(\phi)\}, \quad (8.20)$$

hence completing the proof, which holds for every $l \in \{1, \dots, d\}$. $\square$

We can now move to our first main result, namely the optimality of the thrifty protocol. Since the second phase of the thrifty protocol, converting $|\psi \wedge \phi\rangle$ into $|\phi\rangle$ (i.e., the third green arrow in Figure 8.2), is deterministic, all we need to consider is the first phase, namely the probabilistic LOCC transformation from $|\psi\rangle$ to $|\psi \wedge \phi\rangle$ (i.e., the two first green arrows in Figure 8.2).

In the following theorem, we prove that the maximum probability of achieving such a transformation from $|\psi\rangle$ to $|\psi \wedge \phi\rangle$ is equivalent to that of realizing the transformation from $|\psi\rangle$ to $|\phi\rangle$. In some sense, with this specific probability, one may as well stop the thrifty protocol halfway in order to produce the more entangled state $|\psi \wedge \phi\rangle$ rather than $|\phi\rangle$.

Theorem 8.1. *Let $|\psi\rangle$ and $|\phi\rangle$ be two bipartite pure states. The optimal probability of conversion from $|\psi\rangle$ to $|\psi \wedge \phi\rangle$, which we name τ_1 , is equal to that of the optimal conversion from $|\psi\rangle$ to $|\phi\rangle$, denoted as r_1 in Eq. (8.7).*

Proof. We need to show that $\tau_1 = r_1$, where r_1 is given by Eq. (8.7) and τ_1 is the optimal probability of conversion from $|\psi\rangle$ to $|\psi \wedge \phi\rangle$. We know that this probabilistic transformation can be performed optimally using Vidal's protocol, so τ_1 can be calculated with Eq. (8.7), namely,

$$\tau_1 = \min_{l \in \{1, \dots, d\}} \frac{E_l(\psi)}{E_l(\psi \wedge \phi)} := \frac{E_{l_1}(\psi)}{E_{l_1}(\psi \wedge \phi)}, \quad (8.21)$$

where l_1 is defined as the smallest $l \in \{1, \dots, d\}$ minimizing the ratio $E_l(\psi)/E_l(\psi \wedge \phi)$. With the use of Lemma 8.1, we can write

$$\tau_1 = \min_{l \in \{1, \dots, d\}} \frac{E_l(\psi)}{\max\{E_l(\psi), E_l(\phi)\}}, \quad (8.22)$$

which is to be compared with r_1 . We obviously have

$$\frac{E_l(\psi)}{\max\{E_l(\psi), E_l(\phi)\}} \leq \frac{E_l(\psi)}{E_l(\phi)}, \quad \forall l \in \{1, \dots, d\}. \quad (8.23)$$

For any value of l such that $E_l(\phi) < E_l(\psi)$, we have

$$1 = \frac{E_l(\psi)}{\max\{E_l(\psi), E_l(\phi)\}} < \frac{E_l(\psi)}{E_l(\phi)}, \quad (8.24)$$

implying that such a value of l can be disregarded in the minimization over l that yields r_1 in Eq. (8.7) or τ_1 in Eq. (8.22) since, when $l = 1$, $E_1(\psi) = E_1(\phi) = 1$ so the minimum is guaranteed not to exceed 1. Thus, we may restrict the minimization over the values of l such that $E_l(\phi) \geq E_l(\psi)$, for which

$$\frac{E_l(\psi)}{\max\{E_l(\psi), E_l(\phi)\}} = \frac{E_l(\psi)}{E_l(\phi)}. \quad (8.25)$$

This implies that we have equal conversion probabilities,

$$\tau_1 = r_1, \quad (8.26)$$

hence we have proven that the thrifty protocol is also an optimal probabilistic protocol. $\square$

We now turn to our second main result, namely that entanglement is better preserved, on average, by the thrifty protocol rather than the greedy protocol. Let us first prove the following lemma.

Lemma 8.2. *For all x and $y \in \mathcal{P}_d^1$ such that $x < y$, and for all non-increasingly ordered vectors $a \in \mathbb{R}_+^d$ such that $\sum_{i=1}^d a_i x_i = \sum_{i=1}^d a_i y_i = 1$, we have*

$$a \odot x < a \odot y, \quad (8.27)$$

where $\odot$ denotes the Hadamard (element-wise) product.

Proof. The equality [see Eq. (7.2)] in the majorization relation of Eq. (8.27) is trivially fulfilled by hypothesis, so it remains to prove the following $d - 1$ inequalities [see Eq. (7.2)],

$$\sum_{i=1}^k a_i (y_i - x_i) \geq 0, \quad \forall k \in \{1, \dots, d-1\}, \quad (8.28)$$

given the majorization hypothesis $x < y$, which reads as

$$\sum_{i=1}^k (y_i - x_i) \geq 0, \quad \forall k \in \{1, \dots, d-1\}. \quad (8.29)$$

The non-increasing rearrangement of vector a allows us to write its components in the form

$$a_i = \sum_{j=i}^d \mu_j, \quad \forall i \in \{1, \dots, d\}, \quad (8.30)$$

where $\mu_j \geq 0$, for all $j \in \{1, \dots, d\}$. By plugging Eq. (8.30) into Eq. (8.28), we have to prove

$$\sum_{i=1}^k (y_i - x_i) \sum_{j=i}^d \mu_j \geq 0, \quad \forall k \in \{1, \dots, d-1\}, \quad (8.31)$$

which can be rewritten by interchanging the two summation signs as

$$\sum_{j=1}^d \mu_j \sum_{i=1}^{\min(k,j)} (y_i - x_i) \geq 0, \quad \forall k \in \{1, \dots, d-1\}. \quad (8.32)$$

These inequalities are clearly satisfied given $\mu_j \geq 0$ and the majorization hypothesis, Eq. (8.29), thereby completing the proof. $\square$

With this in mind, we can prove our second main result, namely that the residual state $|\nu\rangle$ of the thrifty protocol is majorized by the residual state $|\xi\rangle$ of the greedy protocol. This is the content of the following theorem.

Theorem 8.2. *Let $|\psi\rangle$ and $|\phi\rangle$ be two bipartite pure states and let $|\xi\rangle$ be the residual state if the two-outcome probabilistic protocol for converting $|\psi\rangle$ into $|\phi\rangle$ fails and $|\nu\rangle$ be the residual state if the two-outcome probabilistic protocol for converting $|\psi\rangle$ into $|\psi \wedge \phi\rangle$ fails. Both residual states are related via the majorization relation*

$$\lambda_\nu < \lambda_\xi. \quad (8.33)$$

Proof. To show this, we proceed in three steps. First, using the definitions from Eq. (8.8), we characterize the sequence of ratios τ_j 's (and corresponding values l_j 's) associated with the probabilistic conversion from $|\psi\rangle$ to $|\psi \wedge \phi\rangle$ and prove that $\tau_j = r_j$ and $l_j = l_j$, for all j . Then, we express the intermediate state $|\zeta\rangle$ of this conversion and show that it satisfies $\lambda_\zeta < \lambda_\chi$, where $|\chi\rangle$ corresponds to the intermediate state of the probabilistic conversion from $|\psi\rangle$ (or from $|\psi \vee \phi\rangle$) to $|\phi\rangle$, as shown in Figure 8.2. Finally, we use $\lambda_\zeta < \lambda_\chi$ in order to prove the desired majorization relation $\lambda_\nu < \lambda_\xi$.

From Theorem 8.1, we already know that for $j = 1$, $\tau_1 = r_1$ and $l_1 = l_1$. In order to prove the same relations for higher values of j , we proceed by iteration. We can write τ_2 , by means of Eq. (8.8), as

$$\begin{aligned} \tau_2 &= \min_{l \in \{1, \dots, l_1 - 1\}} \frac{E_l(\psi) - E_{l_1}(\psi)}{E_l(\psi \wedge \phi) - E_{l_1}(\psi \wedge \phi)}, \\ &:= \frac{E_{l_2}(\psi) - E_{l_1}(\psi)}{E_{l_2}(\psi \wedge \phi) - E_{l_1}(\psi \wedge \phi)}, \end{aligned} \quad (8.34)$$

which can be reexpressed, using $l_1 = l_1$ and $E_{l_1}(\psi \wedge \phi) = \max\{E_{l_1}(\psi), E_{l_1}(\phi)\} = E_{l_1}(\phi)$, as

$$\tau_2 = \min_{l \in \{1, \dots, l_1 - 1\}} \frac{E_l(\psi) - E_{l_1}(\psi)}{\max\{E_l(\psi), E_l(\phi)\} - E_{l_1}(\phi)}. \quad (8.35)$$

We choose to rewrite this last expression as

$$\tau_2 = \min_{l \in \{1, \dots, l_1 - 1\}} \tau_{2,l}. \quad (8.36)$$

We now divide the values of l in two categories. Either $l \in \mathcal{L}^+$ when $\max\{E_l(\psi), E_l(\phi)\} = E_l(\psi)$, or $l \in \mathcal{L}^-$ when $\max\{E_l(\psi), E_l(\phi)\} = E_l(\phi)$. It is possible to prove that there always exists one value $l_{\min} \in \mathcal{L}^-$ such that $\tau_{2,l_{\min}} \leq \tau_{2,l}$, for every $l \in \mathcal{L}^+ \cup \mathcal{L}^-$. Hence,

$$\tau_2 = r_2. \quad (8.37)$$

Proceeding equivalently for each j , we show that

$$\tau_j = r_j \text{ and } l_j = l_j, \quad \forall j \in \{1, \dots, k\}. \quad (8.38)$$

This allows us to write the vector of Schmidt coefficients of the intermediate state $|\zeta\rangle$ between $|\psi\rangle$ and $|\psi \wedge \phi\rangle$ as

$$\lambda_\zeta^{(i)} = r_j \lambda_{\psi \wedge \phi}^{(i)}, \quad \text{if } i \in \{l_j, \dots, l_{j-1} - 1\}, \quad \text{with } j \in \{1, \dots, k\}. \quad (8.39)$$

Thus, this vector can be expressed as a Hadamard product of two vectors, namely,

$$\lambda_\zeta = \mathbf{r} \odot \lambda_{\psi \wedge \phi}, \quad (8.40)$$

where $(\mathbf{r})_i = r_j$, if $i \in \{l_j, \dots, l_{j-1} - 1\}$, for every $i \in \{1, \dots, d\}$. Similarly, in view of Eq. (8.9), the vector of Schmidt coefficients of the intermediate state $|\chi\rangle$ between $|\psi\rangle$ (or $|\psi \vee \phi\rangle$) and $|\phi\rangle$ can be rewritten using this notation as

$$\lambda_\chi = \mathbf{r} \odot \lambda_\phi. \quad (8.41)$$

Now, we simply make use of Lemma 8.2. Since $\lambda_{\psi \wedge \phi} < \lambda_\phi$, the majorization relation

$$\lambda_\zeta = \mathbf{r} \odot \lambda_{\psi \wedge \phi} < \mathbf{r} \odot \lambda_\phi = \lambda_\chi, \quad (8.42)$$

holds true provided the vector $\mathbf{r}$ is non-increasingly ordered (by increasing index i). This can be easily checked from the structure of this vector, namely,

$$\mathbf{r} = (r_k \dots r_k | \dots | r_2 \dots r_2 | r_1 \dots r_1)^T, \quad (8.43)$$

$\uparrow \qquad \qquad \qquad \uparrow \qquad \qquad \uparrow$
 $l_k \qquad \qquad \qquad l_2 \qquad \qquad l_1$

where each l_j variable points to the index of the corresponding element in $\mathbf{r}$. Hence, $\lambda_\zeta < \lambda_\chi$, which means that the intermediate state $|\zeta\rangle$ of the thrifty protocol is more entangled than the intermediate state $|\chi\rangle$ of the greedy protocol, as can also be seen in Figure 8.2.

Finally, we deduce from Eq. (8.42) the fact that $\lambda_\nu < \lambda_\zeta$, i.e., that if the protocol fails, the residual state of the thrifty protocol is more entangled than the residual state of the greedy protocol (see inset of Figure 8.2). To do this, first notice that

$$(\hat{N} \otimes \hat{1}_{[d]}) |\chi\rangle = \sqrt{1 - r_1} |\zeta\rangle, \quad (8.44)$$

and

$$(\hat{N} \otimes \hat{1}_{[d]}) |\zeta\rangle = \sqrt{1 - r_1} |\nu\rangle. \quad (8.45)$$

Given that $\hat{N}$ is diagonal [see Eq. (8.11)], we can rewrite Eqs. (8.44) and (8.45) using vectors of Schmidt coefficients and the Hadamard product as

$$\lambda_\xi = \mathbf{n} \odot \lambda_\chi, \quad \lambda_\nu = \mathbf{n} \odot \lambda_\zeta, \quad (8.46)$$

where $(\mathbf{n})_i = \langle i | \hat{N}^\dagger \hat{N} | i \rangle / (1 - r_1)$, for every $i \in \{1, \dots, d\}$. Finally, using again Lemma 8.2, valid because of the fact that $\mathbf{n} \in \mathbb{R}_+^d$ is a non-increasingly ordered vector and the fact that $\lambda_\zeta < \lambda_\chi$, we complete the proof that $\lambda_\nu < \lambda_\xi$. $\square$

8.5 Generalization to a collection of initial or final states

We can first generalize Theorem 8.1 to a scenario involving multiple final states. Imagine that Alice and Bob possess a state $|\psi\rangle$ and that, instead of a single target state, there is a collection of m possible target states $\{|\phi_j\rangle\}_{j=1}^m$ from which they will have to produce a single one (without knowing which one beforehand). The idea is that they can first perform a probabilistic LOCC transformation to the OCR state of the $m+1$ states, denoted as

$$\left| \psi \bigwedge_{j=1}^m \phi_j \right\rangle := |\psi \wedge \phi_1 \wedge \phi_2 \wedge \dots \wedge \phi_m\rangle, \quad (8.47)$$

extending the essence of the thrifty protocol. Once they possess the OCR state, they can wait until they learn which state $|\phi_j\rangle$ they have to produce before performing the corresponding deterministic LOCC transformation from the OCR state to $|\phi_j\rangle$. Interestingly, the probabilistic LOCC from $|\psi\rangle$ to the OCR state has an optimal probability that is equal to the minimum optimal probability of all individual conversions from $|\psi\rangle$ to $|\phi_j\rangle$. In some sense, with this minimum probability corresponding to the hardest-to-reach target state $|\phi_j\rangle$, we may as well stop the protocol halfway to produce the more entangled OCR state. This is the content of the following theorem.

Theorem 8.3. *Let $|\psi\rangle$ be a bipartite pure state and $\{|\phi_j\rangle\}_{j=1}^m$ be a collection of m bipartite pure states. The optimal probability $p_{\max}^{[1]}$ of conversion from $|\psi\rangle$ to the OCR state of the $m+1$ states $\left| \psi \bigwedge_{j=1}^m \phi_j \right\rangle$ is equal to the minimum optimal probability among the m possible individual conversions from $|\psi\rangle$ to $|\phi_j\rangle$, i.e.,*

$$p_{\max}^{[1]} = \min_{j \in [1, m]} p_{\max, j}^{[2]}, \quad (8.48)$$

where $p_{\max, j}^{[2]}$ denotes the optimal probability of conversion from $|\psi\rangle$ to $|\phi_j\rangle$.

Proof. From Eq. (8.5), we have

$$p_{\max}^{[1]} = \min_{l \in [1, d]} \frac{E_l(\psi)}{E_l\left(\psi \bigwedge_{j=1}^m \phi_j\right)}, \quad (8.49)$$

as well as

$$p_{\max, j}^{[2]} = \min_{l \in [1, d]} \frac{E_l(\psi)}{E_l(\phi_j)}, \quad (8.50)$$

and we want to prove that

$$p_{\max}^{[1]} = \min_{j \in [1, m]} p_{\max, j}^{[2]} =: p_{\max}^{[2]}. \quad (8.51)$$

In order to do so, we can use the straightforward generalization of Lemma 8.1, namely,

$$E_l\left(\psi \bigwedge_{j=1}^m \phi_j\right) = \max\{E_l(\psi), E_l(\phi_1), \dots, E_l(\phi_m)\}. \quad (8.52)$$

Thus, we have $p_{\max}^{[1]} = \min_{l \in [1, d]} p_{\max, l}^{[1]}$, where

$$p_{\max, l}^{[1]} = \frac{E_l(\psi)}{\max\{E_l(\psi), E_l(\phi_1), \dots, E_l(\phi_m)\}}, \quad (8.53)$$

and, by interchanging the minima, we can also write

$$p_{\max}^{[2]} = \min_{l \in [1, d]} p_{\max, l}^{[2]}, \quad (8.54)$$

where

$$p_{\max, l}^{[2]} = \min_{j \in [1, m]} \frac{E_l(\psi)}{E_l(\phi_j)}. \quad (8.55)$$

Therefore, using Eqs. (8.53) and (8.55), it is easy to see that $p_{\max, l}^{[2]} \geq p_{\max, l}^{[1]} = 1$ for any value of l such that $\max\{E_l(\psi), E_l(\phi_1), \dots, E_l(\phi_m)\} = E_l(\psi)$. These values of l can thus be disregarded in the minimization over l since the minimum cannot exceed 1 (indeed, $p_{\max, 1}^{[2]} = p_{\max, 1}^{[1]} = 1$ when $l = 1$). Thus, we only have to consider the values of l such that $\max\{E_l(\psi), E_l(\phi_1), \dots, E_l(\phi_m)\} = E_l(\phi_k)$, with $k \in \{1, \dots, m\}$, implying $p_{\max, l}^{[2]} = p_{\max, l}^{[1]} = E_l(\psi)/E_l(\phi_k)$. Hence, the minimum over l yields $p_{\max}^{[1]} = p_{\max}^{[2]}$, which completes the proof. $\square$

In other words, as long as Alice and Bob do not know which target state $|\phi_j\rangle$ they will have to reach, they may anticipate the probabilistic step and move to the OCR, but this can only be done with the minimum probability (corresponding to the worst-case target state). The benefit of this procedure is that once the target $|\phi_j\rangle$ is disclosed, the final step is deterministic. Otherwise, they may wait until the target $|\phi_j\rangle$ is disclosed and only then perform a probabilistic transformation from $|\psi\rangle$ to $|\phi_j\rangle$, whose probability may be higher (depending on which is the target state).

In a nutshell, a compromise can be made between doing the probabilistic transformation at first (making the rest fully deterministic regardless of the target state) or waiting to know which target state is wanted before doing the actual probabilistic transformation (possibly with a higher success probability).

It is worth noting that the Procrustean method introduced in Ref. [19] can be viewed as a special case of our thrifty protocol in a scenario where the target state is a priori totally arbitrary. The Procrustean method is a local filtering which converts a single copy of a given partially entangled pure state $|\psi\rangle$ into the maximally entangled state $|\Phi\rangle$ of the same dimension. Indeed, when considering the thrifty protocol with the set of target states $\{|\phi_j\rangle\}$ being the whole set of states of the same dimension as the initial state, the OCR then becomes the maximally entangled state $|\Phi\rangle$. Hence, the thrifty protocol starts with the probabilistic conversion of $|\psi\rangle$ into $|\Phi\rangle$, which is thus equivalent to the Procrustean method⁶. Once the actual target state $|\phi_j\rangle$ is revealed, one may then deterministically convert $|\Phi\rangle$ into $|\phi_j\rangle$ (which is possible since $|\Phi\rangle \prec |\phi_j\rangle$, for every j). Of course, this Procrustean procedure becomes suboptimal in situations where the set of target states is limited so the OCR is not maximally entangled, in which case the thrifty protocol comes with a higher success probability.

It is also tempting to compare our multitarget-state thrifty protocol to the protocol introduced in Ref. [143]. There, the conversion into an ensemble of pure states with associated probabilities is considered. Specifically, the conversion of $|\psi\rangle$ into the ensemble $\{q_j, |\phi_j\rangle\}_{j=1}^m$ is shown to be possible via a LOCC transformation if and only if

$$\lambda_\psi \prec \sum_{j=1}^m q_j \lambda_{\phi_j}, \quad (8.56)$$

which extends Nielsen's theorem. The conversion starts with a deterministic step towards an "average" pure state whose Schmidt vector is equal to $\sum_{j=1}^m q_j \lambda_{\phi_j}$, followed by a probabilistic step yielding each state $|\phi_j\rangle$ with probability q_j . Despite the apparent similarity between such a protocol and our multi-target thrifty protocol, the two have quite different purposes. Indeed, our protocol has a specific probability of succeeding, i.e., outputting the OCR, which is then deterministically converted into one of the desired target states $|\phi_j\rangle$ chosen at will. According to Theorem 8.3, the success probability is the minimum of the probabilities associated with each individual conversion (Vidal's bound for each pair of initial and target states). In contrast, the protocol of Ref. [143] is designed as a source that randomly outputs a state $|\phi_j\rangle$ drawn from a target ensemble. The weights q_j in this ensemble can be chosen freely but must satisfy Eq. (8.56) for any physically realizable ensemble. In particular, each q_j must still be upper bounded by Vidal's corresponding optimal conversion probability from $|\psi\rangle$ to $|\phi_j\rangle$.

⁶This is a consequence of the fact that Vidal's protocol itself can be seen as a generalization of the Procrustean method when the target state differs from the maximally entangled state of the same dimension as the initial state.

To complete the picture, we may also consider a scenario involving multiple initial states $\{|\psi_j\rangle\}_{j=1}^m$ and a single target state $|\phi\rangle$, generalizing the greedy protocol. This time, the optimal probability of conversion $p_{\max}^{[1]}$ from the OCP state of the $m+1$ states $\left|\phi \bigvee_{j=1}^m \psi_j\right\rangle^7$ to the target state $|\phi\rangle$ is equivalent to the minimum optimal probability between any of the m possible individual conversions from $|\psi_j\rangle$ to $|\phi\rangle$. This is immediately understood because, as proven in Ref. [26], the intermediate state for the least probable of the m possible conversions is located in the down cone of $\left|\phi \bigvee_{j=1}^m \psi_j\right\rangle$. Therefore, the transformation from the OCP state $\left|\phi \bigvee_{j=1}^m \psi_j\right\rangle$ to $|\phi\rangle$ cannot be more probable than the least probable of the m possible conversions from $|\psi_j\rangle$ to $|\phi\rangle$. Unlike the multi-target thrifty protocol, this multi-initial-state generalization of the greedy protocol does not have a straightforward practical meaning since the deterministic transformation from some of the initial states $|\psi_j\rangle$ to the OCP state cannot be performed until the actual $|\psi_j\rangle$ is disclosed. However, it is relevant in a scenario involving a catalyst state [142]. If one postpones the deterministic step and, instead, borrows the OCP state, then it is possible to implement the probabilistic step yielding $|\phi\rangle$ from the OCP state. Later on, once the actual $|\psi_j\rangle$ is disclosed, it can be deterministically converted into the OCP state, which is returned and plays therefore the role of a catalyst. Here again, there is a compromise between doing the probabilistic step at first (postponing the subsequent deterministic step) or waiting until the identity of the initial state is known before doing the probabilistic transformation (possibly with a higher success probability).

Finally, we discuss the following scenario. Imagine that two parties, Alice and Bob, are spatially separated and that they each have access to a branch of a so-called bank of bipartite pure states (see Figure 8.3). A branch of this bank has the particularity to rent a qudit of a pair of bipartite entangled pure state to a nearby user. However, the rental price is calculated as the difference between the entanglement of formation of the initial state and the one of the final state. Hence, Alice and Bob prefer to rent the least entangled state needed for whatever protocol they want to perform and return the most entangled state possible after their protocol has ended. Note that, in full generality, the bank possesses branches at many locations and if the entangled state needed for Alice and Bob is shared between two other branches than theirs, the bank can perform entanglement swapping [255], in order to bring the qudits to Alice and Bob's branches. Now, imagine that Alice and Bob wish to perform a series of protocols involving different bipartite pure states (we suppose that they know in advance what state they recover at the end of each protocol) and that they only have access to quantum communication with their local branch before knowing what state they need in their set of states.

⁷In analogy with the OCR of multiple states of Eq. (8.47), the OCP of multiple states can be understood as $\left|\phi \bigvee_{j=1}^m \psi_j\right\rangle := |\phi \vee \psi_1 \vee \psi_2 \vee \cdots \vee \psi_m\rangle$.

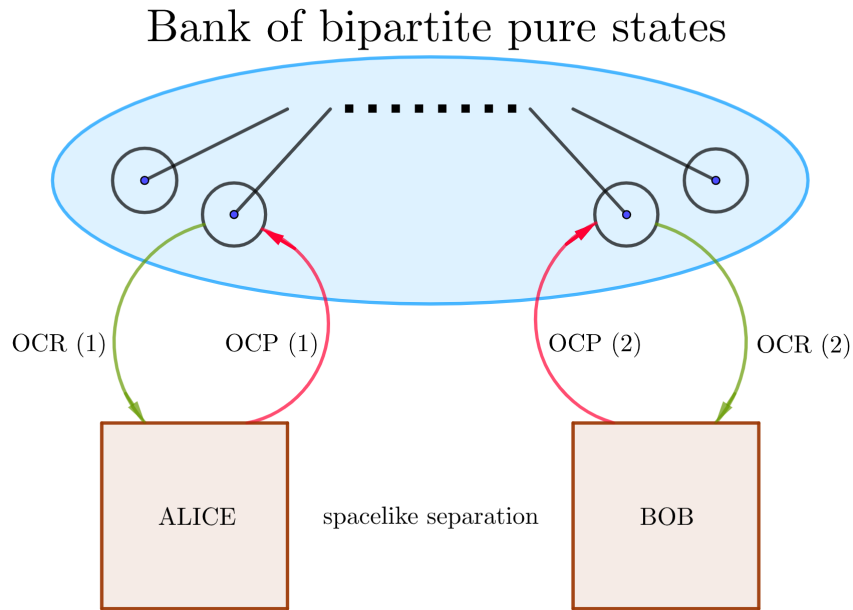


Figure 8.3: Representation of the bank of bipartite pure states. The bank possesses several branches spatially separated from one another, they are represented by the circles in the blue ellipse. Alice and Bob each have access to one branch of the bank, in order to request (resp. return) one qudit of an entangled pair corresponding to the OCR (resp. OCP) state of the states involved in their series of protocols.

The optimal rent they should perform is to ask the bank for the OCR state of all the different states they may need and to tell the bank that they will return the OCP state of all the possible states they can have as outcomes of each protocol, because we suppose that the bank needs to know in advance what state they will retrieve.

8.6 Conclusion

The lattice structure of majorization uncovers two essential states in the context of entanglement transformations between incomparable bipartite pure states $|\psi\rangle$ and $|\phi\rangle$, namely the optimal common resource state (i.e., the *meet* $|\psi \wedge \phi\rangle$ of the two states) and the optimal common product state (i.e., the *join* $|\psi \vee \phi\rangle$ of the two states). We have shown that both states naturally appear when considering (single-copy) probabilistic LOCC transformations from $|\psi\rangle$ to $|\phi\rangle$. We have indeed defined two antipodal protocols, namely the greedy protocol, passing through $|\psi \vee \phi\rangle$, and the thrifty protocol, passing through $|\psi \wedge \phi\rangle$. Both protocols can be proven to be optimal (their success probability is maximum). However, while the greedy protocol is very similar to Vidal's protocol [233], the thrifty protocol is superior in that the entanglement resource is better preserved on average (in case of failure, its residual state is majorized by – hence, is more entangled than – the residual state of the greedy protocol).

Note that in case $|\psi\rangle$ and $|\phi\rangle$ are comparable, both the greedy and thrifty protocols reduce to Vidal's protocol, which reflects that the incomparability between the states is an essential ingredient here. Finally, we have shown that the greedy and thrifty protocols can be generalized to scenarios involving an arbitrary number of initial or final states. This underlines the operational relevance of the majorization lattice in the scope of quantum entanglement theory.

Overall, this chapter sheds a new light on the resource theory of entanglement. Such a theory has proven to be essential in numerous areas within quantum information sciences. For example, the rapidly developing field of quantum networks [159, 179] relies on the ability to convert and manipulate entanglement at the local scale so to establish useful entanglement between distant nodes. Another notable example is quantum thermodynamics [27, 130, 217], where the resource-theoretical approach is well suited to describe the allowed state transformations (here, thermo-majorization is the condition that reflects the existence of thermal operations) and where the notion of majorization lattice has already proven its relevance [59, 149]. Thus, plugging the resource theory of entanglement into a majorization lattice, as we have sketched here, can be expected to be a very promising research avenue. We may conceive cryptographic or thermodynamic scenarios where the convertibility (or non-convertibility) from or towards the meet or join state is crucial.

Finally, it would be interesting to extend the current work to multipartite settings [41, 170]. To our knowledge, the notions of optimal common resource or optimal common product have never been explored for more than two parties. For example, we know that true tripartite entanglement can be split into two categories, namely GHZ-type and W-type states [74], but the determination of an optimal common resource state or optimal common product state within each class constitutes an open problem. In another direction, extending the present analysis to the conversion of mixed states and exploring the role of incomparability in this context would, if possible, even further broaden the applicability of this framework.

9 Geometric aspects of majorization: Relating distances and entropies in the majorization lattice

The content of this chapter is based on the following work: Serge Deside, Michael G. Jabbour, and Nicolas J. Cerf. “Geometric aspects of majorization: Relating distances and entropies in the majorization lattice” [64].

In this chapter, we investigate the geometry of the majorization lattice by means of the total variation distance. Namely, we prove several identities relating the total variation distances between probability vectors in the majorization lattice, their greatest lower bound (the meet) and their least upper bound (the join). For this purpose, we also characterize convex sets of interest in the lattice. We then prove an inequality that goes beyond the Schur-concavity of the Shannon entropy by exploiting the lattice structure of majorization. In doing so, we generalize an inequality introduced in Ref. [124]. Building on these results, we prove a uniform continuity bound for the Shannon entropy under the constraint that the probability vectors lie in a majorization diamond bounded from below and above by two fixed comparable vectors. Finally, this yields a uniform continuity bound for the von Neumann entropy of density matrices, with a constraint requiring the states to be majorized by a fixed upper state and to majorize a fixed lower state. We discuss applications to quantum communications, in particular to the derivation of upper bounds on capacities of quantum channels.

9.1 Introduction

When transformations or tasks associated with quantum resources are applied to a state, we often lack precise knowledge of the state we obtain at the end of the task, e.g., because of imperfections or noise. Another issue is that the resources are often associated with quantifiers that can be difficult to evaluate. One way to circumvent this problem is by estimating errors using metric distances, such as the total variation distance or the trace distance. This raises a natural question: How do such distances behave in the context of the majorization lattice? Asked differently: What is the geometry of the majorization lattice induced by such distances?

An area where the study of these questions would offer useful insights is that of continuity bounds for entropic quantities. Such bounds estimate how much an entropy can vary when the underlying probability distribution, or quantum state, is perturbed by a prescribed amount. They are fundamental stability tools in information theory and play an important role in tackling hard problems in the field, e.g., the study of quantum channel capacities. In the classical setting, continuity bounds have been extensively studied for the Shannon entropy, starting with the tight uniform continuity bound often attributed to Csiszár and derived from Fano's inequality [180], and later for conditional and generalized entropies [5, 137, 198]. They have also been extended to the quantum setting, including bounds for the von Neumann entropy, the conditional von Neumann entropy, Rényi-type entropies, and in particular for infinite-dimensional systems [6, 12, 13, 16, 17, 82, 113, 138, 197, 247].

Relevant notation and definitions are introduced and recalled in Section 9.2. We start in Section 9.3 by revealing the geometry of the majorization lattice through the proof of several lemmas involving the total variation distance (Lemmas 9.1 through 9.5). In Section 9.4, we turn to the characterization of what we coin the *majorization diamond*, a convex set comprising all probability vectors that majorize a given vector, called the ante-vector, and are majorized by another given vector, called the post-vector. We explain why this majorization diamond is of interest, in particular in the context of quantum communications. Section 9.5 is devoted to the proof of an inequality that goes beyond the Schur-concavity of the Shannon entropy in the lattice (Theorem 9.1), and reduces to a result of Ho and Verdú [124] for comparable vectors. We also obtain an inequality that is symmetric to the supermodularity of the Shannon entropy in the majorization lattice and involves the Jensen-Shannon divergence (Theorem 9.2). Section 9.6 concerns the proof of a new uniform continuity bound for the Shannon entropy of probability vectors under the additional constraint that the vectors belong to a fixed but arbitrary majorization diamond (Theorem 9.3). In Section 9.7, we exploit the aforementioned continuity bound to prove one for the von Neumann entropy of density matrices (quantum states), with majorization constraints given by a lower (majorized) state and an upper (majorizing) state (Theorem 9.4). Finally, Section 9.8 is devoted to a discussion and some perspectives. In particular, we discuss the usefulness of our results for quantum communications and for deriving bounds on capacities of quantum channels, hinting at the case of the quantum depolarizing channel, which we leave for future work.

9.2 Mathematical preliminaries

Classical systems. Let d be an integer greater than 1, and denote the set of probability vectors of size d by $\mathcal{P}_d$. The total variation (TV) distance between vectors $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d$ is defined as

$$d(\mathbf{u}, \mathbf{v}) := \frac{1}{2} \sum_{i=1}^d |u_i - v_i|. \quad (9.1)$$

Denote the set of probability vectors whose elements are arranged in non-increasing order by $\mathcal{P}_d^\downarrow$. Then, Definition 7.18 can be reformulated as follows.

Definition 9.1 (Majorization lattice [46]). The majorization lattice is the quadruple $(\mathcal{P}_d^\downarrow, <, \wedge, \vee)$, where the operations $\wedge$ and $\vee$ are defined as follows:

- (i) The *meet* of two elements $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$, denoted $\mathbf{u} \wedge \mathbf{v}$, is the greatest lower bound of $\mathbf{u}$ and $\mathbf{v}$. That is, $\mathbf{u} \wedge \mathbf{v} < \mathbf{u}$, $\mathbf{u} \wedge \mathbf{v} < \mathbf{v}$, and for any $\mathbf{w} \in \mathcal{P}_d^\downarrow$ such that $\mathbf{w} < \mathbf{u}$ and $\mathbf{w} < \mathbf{v}$, we have $\mathbf{w} < \mathbf{u} \wedge \mathbf{v}$;
- (ii) The *join* of two elements $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$, denoted $\mathbf{u} \vee \mathbf{v}$, is the least upper bound of $\mathbf{u}$ and $\mathbf{v}$. That is, $\mathbf{u} < \mathbf{u} \vee \mathbf{v}$, $\mathbf{v} < \mathbf{u} \vee \mathbf{v}$, and for any $\mathbf{w} \in \mathcal{P}_d^\downarrow$ such that $\mathbf{u} < \mathbf{w}$ and $\mathbf{v} < \mathbf{w}$, we have $\mathbf{u} \vee \mathbf{v} < \mathbf{w}$.

Recall that both vectors $\mathbf{u} \wedge \mathbf{v}$ and $\mathbf{u} \vee \mathbf{v}$ can be deduced from the components of $\mathbf{u}$ and $\mathbf{v}$ as follows [46]. For the meet,

$$(\mathbf{u} \wedge \mathbf{v})_k = \min(s_k(\mathbf{u}), s_k(\mathbf{v})) - \min(s_{k-1}(\mathbf{u}), s_{k-1}(\mathbf{v})), \quad (9.2)$$

for all $k = 1, \dots, d$. For the join, we begin by writing the intermediate vector $\beta(\mathbf{u}, \mathbf{v})$, whose components are defined as follows:

$$\beta(\mathbf{u}, \mathbf{v})_k := \max(s_k(\mathbf{u}), s_k(\mathbf{v})) - \max(s_{k-1}(\mathbf{u}), s_{k-1}(\mathbf{v})), \quad (9.3)$$

for all $k = 1, \dots, d$. However, $\beta(\mathbf{u}, \mathbf{v})$ does not belong to $\mathcal{P}_d^\downarrow$ in general. In fact, two cases can occur: either $\beta(\mathbf{u}, \mathbf{v})$ is a non-increasingly-ordered vector and $\mathbf{u} \vee \mathbf{v} = \beta(\mathbf{u}, \mathbf{v})$, or $\beta(\mathbf{u}, \mathbf{v})$ is not an ordered vector and a pool-adjacent-violators algorithm (PAVA) is used to obtain $\mathbf{u} \vee \mathbf{v}$ from $\beta(\mathbf{u}, \mathbf{v})$ (see Section 7.3.2 for the detailed PAVA).

Quantum systems. Let $\mathcal{H}$ denote the Hilbert space associated with a finite-dimensional quantum system; $\dim \mathcal{H} = d < \infty$. Let $\mathcal{L}(\mathcal{H})$ denote the algebra of linear operators acting on $\mathcal{H}$, and $\mathcal{D}(\mathcal{H}) \subset \mathcal{L}(\mathcal{H})$ denote the set of quantum states (density matrices) of the system, i.e., positive semidefinite operators of unit trace. Recall that the trace norm of any operator $\hat{A} \in \mathcal{L}(\mathcal{H})$ is defined as $\|\hat{A}\|_1 = \text{Tr}|\hat{A}|$, with $|\hat{A}| = \sqrt{\hat{A}^\dagger \hat{A}}$. The trace distance T between two states $\hat{\rho}, \hat{\sigma} \in \mathcal{D}(\mathcal{H})$ is then given by

$$T(\hat{\rho}, \hat{\sigma}) := \frac{1}{2} \|\hat{\rho} - \hat{\sigma}\|_1. \quad (9.4)$$

The von Neumann entropy S of a state $\hat{\rho}$ is given by [245]

$$S(\hat{\rho}) := -\text{Tr}(\hat{\rho} \log \hat{\rho}). \quad (9.5)$$

For a state $\hat{\rho} \in \mathcal{D}(\mathcal{H})$, we denote its vector of eigenvalues by $\lambda(\hat{\rho})$ in what follows. For two states $\hat{\rho}, \hat{\sigma} \in \mathcal{D}(\mathcal{H})$, we write $\hat{\rho} < \hat{\sigma}$ whenever $\lambda(\hat{\rho}) < \lambda(\hat{\sigma})$.

9.3 Geometry of the majorization lattice

We will be interested in different sets in the majorization lattice, which we now define:

- The *majorized set* of u in $\mathcal{P}_d^\downarrow$ is

$$\mathcal{S}_u^< := \{w \in \mathcal{P}_d^\downarrow \mid w < u\}; \quad (9.6)$$

- The *majorizing set* of u in $\mathcal{P}_d^\downarrow$ is

$$\mathcal{S}_u^> := \{w \in \mathcal{P}_d^\downarrow \mid u < w\}; \quad (9.7)$$

- The *majorization diamond* associated with $a, b \in \mathcal{P}_d^\downarrow$ satisfying $a < b$ is defined as

$$\mathcal{S}_{a,b}^\diamond := \mathcal{S}_a^> \cap \mathcal{S}_b^< = \{x \in \mathcal{P}_d^\downarrow : a < x < b\}. \quad (9.8)$$

We call a the *ante-vector* and b the *post-vector*, emphasizing that the diamond consists of all vectors lying between them in the majorization order.

In particular, if $u, v \in \mathcal{P}_d^\downarrow$, then the interval between their meet and their join is a majorization diamond, namely

$$\mathcal{S}_{u \wedge v, u \vee v}^\diamond = \{x \in \mathcal{P}_d^\downarrow : u \wedge v < x < u \vee v\}. \quad (9.9)$$

The different sets can be viewed in Figure 7.2. A first notion of distance in the majorization lattice was introduced in Ref. [45] using the Shannon entropy, in the form of Eq. (7.70). Later, a similar notion of distance was proposed in Ref. [249] by replacing the Shannon entropy with Rényi entropies in Eq. (7.70) for a certain range of parameters. Indeed, one recovers a proper metric as Rényi entropies also satisfy the property of supermodularity (see Proposition 7.3) for that range of parameters (see also Ref. [223]). Rényi entropies (the Shannon entropy being a particular case) are all Schur-concave. Thus, they play a particular role with respect to majorization. They are, however, not the only Schur-concave functions; thus, there is no particular reason why a distance in the majorization lattice should be defined in terms of these functions.

We argue that another notion of distance, that is, the TV distance defined in Eq. (9.1), allows us to recover more natural mathematical statements. In particular, we begin by showing that Eq. (7.74) holds with equality under the TV distance.

Lemma 9.1. *Consider two elements u and v in $\mathcal{P}_d^{\downarrow}$. Then,*

$$d(u \wedge v, u) = d(v, u \vee v). \quad (9.10)$$

Proof. We first prove the result in the case where $u \vee v = \beta(u, v)$. By definition of the meet and $\beta(u, v)$ (see Eqs. (9.2) and (9.3)), we have

$$(u \wedge v)_i + \beta(u, v)_i = u_i + v_i, \quad \forall i = 1, \dots, d. \quad (9.11)$$

Therefore,

$$(u \wedge v)_i - u_i = v_i - \beta(u, v)_i, \quad \forall i = 1, \dots, d. \quad (9.12)$$

Taking absolute values and summing over i gives

$$d(u \wedge v, u) = d(v, \beta(u, v)). \quad (9.13)$$

If $u \vee v = \beta(u, v)$, this concludes the proof.

In the general case, $u \vee v$ is obtained from $\beta(u, v)$ by the PAVA of Section 7.3.2. From Lemma C.1, we have that this procedure preserves the ℓ_1 -distance to v , namely

$$d(v, \beta(u, v)) = d(v, u \vee v). \quad (9.14)$$

Hence,

$$d(u \wedge v, u) = d(v, u \vee v), \quad (9.15)$$

as claimed. $\square$

Interestingly, we can prove that any convex combination of any two elements in $\{\mathbf{u}, \mathbf{v}, \mathbf{u} \wedge \mathbf{v}, \mathbf{u} \vee \mathbf{v}\}$ saturates the triangle inequality for the TV distance $d(\mathbf{u}, \mathbf{v})$. This is the content of Lemmas 9.2, 9.3, and 9.4. We begin by showing that $\mathbf{u} \wedge \mathbf{v}$ and $\mathbf{u} \vee \mathbf{v}$ saturate it.

Lemma 9.2. *Consider two elements $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{P}_d^\downarrow$. Then,*

$$\begin{aligned} \min_{\mathbf{w} \in \mathcal{P}_d^\downarrow} [d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})] &= d(\mathbf{u}, \mathbf{u} \wedge \mathbf{v}) + d(\mathbf{u} \wedge \mathbf{v}, \mathbf{v}) \\ &= d(\mathbf{u}, \mathbf{u} \vee \mathbf{v}) + d(\mathbf{u} \vee \mathbf{v}, \mathbf{v}) \\ &= d(\mathbf{u}, \mathbf{v}). \end{aligned} \tag{9.16}$$

Proof. From the triangle inequality, we have that for every $\mathbf{w} \in \mathcal{P}_d^\downarrow$,

$$d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v}), \tag{9.17}$$

while equality is attained by choosing, for instance, $\mathbf{w} = \mathbf{u}$ or $\mathbf{w} = \mathbf{v}$. Hence,

$$\min_{\mathbf{w} \in \mathcal{P}_d^\downarrow} [d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})] = d(\mathbf{u}, \mathbf{v}). \tag{9.18}$$

We now show that

$$d(\mathbf{u}, \mathbf{u} \wedge \mathbf{v}) + d(\mathbf{u} \wedge \mathbf{v}, \mathbf{v}) = d(\mathbf{u}, \mathbf{v}). \tag{9.19}$$

To do so, we first claim that,

$$(\mathbf{u} \wedge \mathbf{v})_i \in [\min(u_i, v_i), \max(u_i, v_i)], \tag{9.20}$$

for all $i = 1, \dots, d$. Indeed, in the expression of the meet (see Eq. (9.2)), if the same vector ($\mathbf{u}$ or $\mathbf{v}$) attains the minimum at both the i^{th} cumulative sum and at the $(i-1)^{\text{th}}$ cumulative sum, then $(\mathbf{u} \wedge \mathbf{v})_i$ is either u_i or v_i . Otherwise, suppose for instance that $s_i(\mathbf{u}) \leq s_i(\mathbf{v})$ and $s_{i-1}(\mathbf{u}) \geq s_{i-1}(\mathbf{v})$. Then,

$$(\mathbf{u} \wedge \mathbf{v})_i = s_i(\mathbf{u}) - s_{i-1}(\mathbf{v}). \tag{9.21}$$

Since $s_{i-1}(\mathbf{u}) \geq s_{i-1}(\mathbf{v})$, we have

$$s_i(\mathbf{u}) - s_{i-1}(\mathbf{v}) \geq s_i(\mathbf{u}) - s_{i-1}(\mathbf{u}) = u_i, \tag{9.22}$$

and since $s_i(\mathbf{u}) \leq s_i(\mathbf{v})$, we have

$$s_i(\mathbf{u}) - s_{i-1}(\mathbf{v}) \leq s_i(\mathbf{v}) - s_{i-1}(\mathbf{v}) = v_i. \tag{9.23}$$

Thus, $(u \wedge v)_i$ lies between u_i and v_i . The opposite crossing case is identical. We thus obtain Eq. (9.20). Using the latter, we get

$$|u_i - (u \wedge v)_i| + |(u \wedge v)_i - v_i| = |u_i - v_i|, \quad (9.24)$$

for all $i = 1, \dots, d$. Summing over i gives

$$d(u, u \wedge v) + d(u \wedge v, v) = d(u, v). \quad (9.25)$$

It remains to prove the corresponding statement for the join. By Lemma 9.1, applied to u, v and also with u and v exchanged, we have

$$d(u \wedge v, u) = d(v, u \vee v), \quad (9.26)$$

and

$$d(u \wedge v, v) = d(u, u \vee v). \quad (9.27)$$

Hence, summing both equations leads to

$$\begin{aligned} d(u, u \vee v) + d(u \vee v, v) &= d(u \wedge v, v) + d(u, u \wedge v) \\ &= d(u, v). \end{aligned} \quad (9.28)$$

This concludes the proof. $\square$

Notice that proving any convex combination of u and v saturates the triangle inequality is straightforward (see Proposition C.1 in Appendix C.2). We prove now that any convex combination of u or v with $u \wedge v$ or $u \vee v$ saturates the triangle inequality.

Lemma 9.3. Consider two elements $u, v \in \mathcal{P}_d^\downarrow$ and $\alpha \in [0, 1]$, and define

$$\begin{cases} u^* := \alpha u + (1 - \alpha)(u \wedge v), \\ u_* := \alpha u + (1 - \alpha)(u \vee v). \end{cases} \quad (9.29)$$

Then,

$$d(u, v) = d(u, u^*) + d(u^*, v), \quad (9.30)$$

and,

$$d(u, v) = d(u, u_*) + d(u_*, v). \quad (9.31)$$

Proof. Immediate by combining Lemmas 9.2 and C.2. $\square$

To complete the picture, we prove that any convex combination of $u \wedge v$ and $u \vee v$ saturates the triangle inequality.

Lemma 9.4. *Let $u, v \in \mathcal{P}_d^1$ and $\alpha \in [0, 1]$ and define*

$$w_\alpha := \alpha(u \wedge v) + (1 - \alpha)(u \vee v). \quad (9.32)$$

We have

$$d(u, v) = d(u, w_\alpha) + d(w_\alpha, v). \quad (9.33)$$

Proof. By the convexity of the TV distance in one of its arguments, we can write

$$d(u, w_\alpha) \leq \alpha d(u, (u \wedge v)) + (1 - \alpha) d(u, (u \vee v)), \quad (9.34)$$

as well as,

$$d(w_\alpha, v) \leq \alpha d((u \wedge v), v) + (1 - \alpha) d((u \vee v), v). \quad (9.35)$$

Summing both gives

$$d(u, w_\alpha) + d(w_\alpha, v) \leq \alpha [d(u, u \wedge v) + d(u \wedge v, v)] + (1 - \alpha) [d(u, u \vee v) + d(u \vee v, v)]. \quad (9.36)$$

By Lemma 9.2, each bracket is equal to $d(u, v)$. Hence, it simplifies to

$$d(u, w_\alpha) + d(w_\alpha, v) \leq d(u, v). \quad (9.37)$$

Therefore, together with the triangle inequality for the TV distance that reads

$$d(u, v) \leq d(u, w_\alpha) + d(w_\alpha, v), \quad (9.38)$$

Eq. (9.33) is obtained. □

Lemmas 9.2, 9.3, and 9.4 can be combined to express the following proposition, summarizing the corresponding results.

Proposition 9.1. *Let $u, v \in \mathcal{P}_d^1$, and define $x \in \mathcal{P}_d^1$ to be any convex combination of any two elements of the set $\{u, v, u \wedge v, u \vee v\}$. Then,*

$$d(u, v) = d(u, x) + d(x, v). \quad (9.39)$$

Another nontrivial distance inequality on the majorization lattice is the following.

Lemma 9.5. *Given $u, v \in \mathcal{P}_d^\downarrow$, we have*

$$d(u, v) \geq d(u \wedge v, u \vee v). \quad (9.40)$$

Note that $d(u, v)$ is not the maximal distance in the diamond, i.e., there exist¹ $p, q \in \mathcal{S}_{u \wedge v, u \vee v}^\diamond$ with $d(p, q) > d(u, v)$.

Proof of Lemma 9.5. This lemma is a direct consequence of Lemmas 9.1 and 9.2. Indeed, the triangle inequality allows us to write,

$$d(u \wedge v, u \vee v) \leq d(u \wedge v, u) + d(u, u \vee v). \quad (9.41)$$

Then, by Lemma 9.1, $d(u, u \vee v) = d(u \wedge v, v)$, giving

$$d(u \wedge v, u \vee v) \leq d(u \wedge v, u) + d(u \wedge v, v). \quad (9.42)$$

Finally, by Lemma 9.2, we have $d(u, v) = d(u, u \wedge v) + d(u \wedge v, v)$, leading to the desired bound, that is,

$$d(u \wedge v, u \vee v) \leq d(u, v). \quad (9.43)$$

□

One may ask when the inequality in Eq. (9.40) is saturated. It is simply when we saturate the triangle inequality from Eq. (9.41). For l_1 -distances, it is achieved when the inequality is saturated coordinate-wise. Indeed,

$$d(u \wedge v, u) + d(u, u \vee v) - d(u \wedge v, u \vee v) \quad (9.44)$$

is equal to

$$\frac{1}{2} \sum_i \left(|(u \wedge v)_i - u_i| + |u_i - (u \vee v)_i| - |(u \wedge v)_i - (u \vee v)_i| \right). \quad (9.45)$$

But, for every a, b, x , $|a - b| \leq |a - x| + |x - b|$. Hence,

$$|(u \wedge v)_i - u_i| + |u_i - (u \vee v)_i| - |(u \wedge v)_i - (u \vee v)_i| \geq 0, \forall i. \quad (9.46)$$

Therefore, for the whole sum to vanish, we need the triangle inequalities of each i to vanish.

¹Take, for example, $u = (\frac{1}{2}, \frac{1}{3}, \frac{1}{12}, \frac{1}{12})$, $v = (\frac{5}{12}, \frac{1}{3}, \frac{1}{4}, 0)$. By choosing $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0)$, $q = (\frac{5}{12}, \frac{5}{12}, \frac{1}{12}, \frac{1}{12}) \in \mathcal{S}_{u \wedge v, u \vee v}^\diamond$, we get $d(u, v) = 1/6$ and $d(p, q) = 1/4$.

Now, for real numbers, $|a - b| = |a - x| + |x - b|$ if and only if $x \in [a, b]$ (if $a \leq b$, but the same applies if $a > b$). Indeed, then $|a - x| + |x - b| = (x - a) + (b - x) = b - a = |a - b|$. Taking $a = (\mathbf{u} \wedge \mathbf{v})_i$, $b = (\mathbf{u} \vee \mathbf{v})_i$, $x = u_i$, we get that

$$|(\mathbf{u} \wedge \mathbf{v})_i - u_i| + |u_i - (\mathbf{u} \vee \mathbf{v})_i| - |(\mathbf{u} \wedge \mathbf{v})_i - (\mathbf{u} \vee \mathbf{v})_i| = 0, \quad (9.47)$$

which is equivalent to

$$u_i \in [(\mathbf{u} \wedge \mathbf{v})_i, (\mathbf{u} \vee \mathbf{v})_i]. \quad (9.48)$$

So, $d(\mathbf{u} \wedge \mathbf{v}, \mathbf{u} \vee \mathbf{v}) = d(\mathbf{u}, \mathbf{v})$ if and only if

$$u_i \in [\min((\mathbf{u} \wedge \mathbf{v})_i, (\mathbf{u} \vee \mathbf{v})_i), \max((\mathbf{u} \wedge \mathbf{v})_i, (\mathbf{u} \vee \mathbf{v})_i)], \quad \forall i. \quad (9.49)$$

The same obviously holds for $\mathbf{v}$ as well.

9.4 Characterization of the majorization diamond

We now turn our focus to the majorization diamond $\mathcal{S}_{a,b}^\diamond$, where $\mathbf{a}, \mathbf{b} \in \mathcal{P}_d^\downarrow$ satisfy $\mathbf{a} < \mathbf{b}$. As we will show later, it is useful in practically motivated situations, for instance, in the context of information processing and communication. In particular, our goal will be to prove specific entropic inequalities in $\mathcal{S}_{a,b}^\diamond$. We thus begin by characterizing the latter.

Proposition 9.2. *The majorization diamond $\mathcal{S}_{a,b}^\diamond$ is convex.*

Proof. Let $\mathbf{p}, \mathbf{q} \in \mathcal{S}_{a,b}^\diamond$. By definition of $\mathcal{S}_{a,b}^\diamond$ (see Eq. (9.8)), we have

$$\begin{aligned} s_k(\mathbf{a}) &\leq s_k(\mathbf{p}) \leq s_k(\mathbf{b}), \\ s_k(\mathbf{a}) &\leq s_k(\mathbf{q}) \leq s_k(\mathbf{b}). \end{aligned} \quad (9.50)$$

for every $k = 1, \dots, d-1$. Let us define

$$\mathbf{r} = \alpha \mathbf{p} + (1 - \alpha) \mathbf{q}, \quad \alpha \in [0, 1]. \quad (9.51)$$

Since $\mathcal{P}_d^\downarrow$ is convex, $\mathbf{r} \in \mathcal{P}_d^\downarrow$. We can write

$$s_k(\mathbf{r}) = \alpha s_k(\mathbf{p}) + (1 - \alpha) s_k(\mathbf{q}), \quad (9.52)$$

Therefore, using Eq. (9.50), we have

$$s_k(\mathbf{r}) \geq \alpha s_k(\mathbf{a}) + (1 - \alpha) s_k(\mathbf{a}) = s_k(\mathbf{a}), \quad (9.53)$$

and similarly,

$$s_k(\mathbf{r}) \leq \alpha s_k(\mathbf{b}) + (1 - \alpha) s_k(\mathbf{b}) = s_k(\mathbf{b}). \quad (9.54)$$

Thus, $\mathbf{a} < \mathbf{r} < \mathbf{b}$. Since also $\mathbf{r} \in \mathcal{P}_d^\downarrow$, we conclude that

$$\mathbf{r} \in \mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond, \quad (9.55)$$

which proves $\mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$ is convex.

Alternatively, one can simply use the fact that the majorization diamond is the intersection of a majorizing set and a majorized set (see Eq. (9.8)), both of which are convex. $\square$

Let us now discuss how to find the extreme points of the diamond set $\mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$, where $\mathbf{a}, \mathbf{b} \in \mathcal{P}_d^\downarrow$ and $\mathbf{a} < \mathbf{b}$. They are found by saturating $d - 1$ independent constraints that we detail now. An element $\mathbf{w} \in \mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$ has to satisfy the Lorenz bounds

$$s_k(\mathbf{a}) \leq s_k(\mathbf{w}) \leq s_k(\mathbf{b}), \quad k = 1, \dots, d - 1, \quad (9.56)$$

where $s_k(\mathbf{x})$ (defined in Eq. (7.1)) denotes the sum of the first k components of $\mathbf{x}$. However, these inequalities alone do not characterize the elements of $\mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$. Indeed, $\mathbf{w}$ must also remain non-increasingly ordered. In terms of the cumulative sums, this amounts to the following constraints

$$2s_k(\mathbf{w}) \geq s_{k-1}(\mathbf{w}) + s_{k+1}(\mathbf{w}), \quad k = 1, \dots, d - 1, \quad (9.57)$$

with $s_0(\mathbf{w}) = 0$ and $s_d(\mathbf{w}) = 1$. We call Eq. (9.57) “concavity constraints” because they are equivalent to constraining the associated Lorenz curve to remain concave.

Thus, one cannot obtain all extreme points simply by choosing, for each k , whether $s_k(\mathbf{w})$ saturates the lower bound $s_k(\mathbf{a})$ or the upper bound $s_k(\mathbf{b})$. Such a choice gives only candidate cumulative curves. Some of them may fail to be concave, and therefore may not correspond to non-increasingly ordered probability vectors. Moreover, some extreme points arise from saturated ordering constraints, namely from equalities of the form $w_i = w_{i+1}$, and not only from saturated Lorenz bounds.

Consequently, the extreme points can be obtained through an active-constraint procedure. In cumulative coordinates, one writes the saturated constraints in Eqs. (9.56) and (9.57) as

$$s_k(\mathbf{w}) = s_k(\mathbf{a}), \quad s_k(\mathbf{w}) = s_k(\mathbf{b}), \quad 2s_k(\mathbf{w}) = s_{k-1}(\mathbf{w}) + s_{k+1}(\mathbf{w}). \quad (9.58)$$

An extreme point is obtained by selecting $d - 1$ independent saturated constraints among the above equations, solving the corresponding linear system for $s_1(\mathbf{w}), \dots, s_{d-1}(\mathbf{w})$, and then checking whether the resulting cumulative curve lies between the Lorenz curves of $\mathbf{a}$ and $\mathbf{b}$ and is concave. The associated probability vector is then recovered from

$$w_k = s_k(\mathbf{w}) - s_{k-1}(\mathbf{w}), \quad k = 1, \dots, d. \quad (9.59)$$

Repeating this procedure over all independent active sets and removing duplicates gives the extreme points of $\mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$. Hence, as in the construction of the join, an algorithm is in general required.

Let us illustrate this in dimension $d = 3$, the first dimension in which incomparability appears. We look at the diamond generated by a meet $\mathbf{m}$ and a join $\mathbf{j}$, that is $\mathcal{S}_{\mathbf{m}, \mathbf{j}}^\diamond$. We choose $\mathbf{u}, \mathbf{v} \in \mathcal{S}_{\mathbf{m}, \mathbf{j}}^\diamond$ such that

$$\mathbf{m} := \mathbf{u} \wedge \mathbf{v}, \quad \mathbf{j} := \mathbf{u} \vee \mathbf{v}. \quad (9.60)$$

In this case, an element $\mathbf{w}$ is determined by the two cumulative sums $s_1(\mathbf{w})$ and $s_2(\mathbf{w})$. The Lorenz bounds read

$$s_1(\mathbf{m}) \leq s_1(\mathbf{w}) \leq s_1(\mathbf{j}), \quad s_2(\mathbf{m}) \leq s_2(\mathbf{w}) \leq s_2(\mathbf{j}), \quad (9.61)$$

and the ordering constraints are

$$s_2(\mathbf{w}) \leq 2s_1(\mathbf{w}), \quad s_2(\mathbf{w}) \geq \frac{1 + s_1(\mathbf{w})}{2}. \quad (9.62)$$

If $\mathbf{u}$ and $\mathbf{v}$ are incomparable, then, up to exchanging them,

$$s_1(\mathbf{u}) < s_1(\mathbf{v}), \quad s_2(\mathbf{u}) > s_2(\mathbf{v}). \quad (9.63)$$

Hence, for this choice, $s_1(\mathbf{m}) = s_1(\mathbf{u})$, $s_2(\mathbf{m}) = s_2(\mathbf{v})$, and $s_1(\mathbf{j}) = s_1(\mathbf{v})$, $s_2(\mathbf{j}) = s_2(\mathbf{u})$.

Thus, the four possible endpoint choices for $(s_1(\mathbf{w}), s_2(\mathbf{w}))$ are

$$(s_1(\mathbf{u}), s_2(\mathbf{v})), \quad (s_1(\mathbf{u}), s_2(\mathbf{u})), \quad (s_1(\mathbf{v}), s_2(\mathbf{v})), \quad (s_1(\mathbf{v}), s_2(\mathbf{u})). \quad (9.64)$$

These correspond respectively to

$$\mathbf{m}, \quad \mathbf{u}, \quad \mathbf{v}, \quad \mathbf{j}. \quad (9.65)$$

Therefore, when $\mathbf{u}$ and $\mathbf{v}$ are incomparable, the four extreme points are $\mathbf{m}$, $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{j}$.

If u and v are comparable, then m and j coincide with u and v , up to their order in the majorization relation. The two other choices of cumulative bounds may fail to define non-increasingly ordered vectors. In that case, the remaining extreme points, if any, come from saturating one of the ordering constraints, namely $w_1 = w_2$ or $w_2 = w_3$. Equivalently, they are obtained by solving the corresponding equalities together with one of the Lorenz bounds, and then checking feasibility.

We can also look for the extreme points of the two order intervals generated by a fixed element u , namely $\mathcal{S}_u^< := \{w \in \mathcal{P}_d^{\downarrow} : w < u\}$ and $\mathcal{S}_u^> := \{w \in \mathcal{P}_d^{\downarrow} : u < w\}$. The reasoning is the same as for the diamond set, except that only one Lorenz boundary is present. More precisely, for $\mathcal{S}_u^<$ one has

$$s_k(w) \leq s_k(u), \quad k = 1, \dots, d-1, \quad (9.66)$$

whereas for $\mathcal{S}_u^>$ one has

$$s_k(u) \leq s_k(w), \quad k = 1, \dots, d-1. \quad (9.67)$$

In both cases, these Lorenz constraints must be supplemented with the ordering constraints ensuring that $w \in \mathcal{P}_d^{\downarrow}$. In cumulative coordinates $s_k(w)$, these read

$$2s_k(w) \geq s_{k-1}(w) + s_{k+1}(w), \quad k = 1, \dots, d-1, \quad (9.68)$$

with $s_0(w) = 0$ and $s_d(w) = 1$.

For $\mathcal{S}_u^>$, one must also explicitly keep the simplex constraint

$$s_{d-1}(w) \leq 1, \quad (9.69)$$

or equivalently $w_d \geq 0$. In the case of $\mathcal{S}_u^<$, this constraint is automatically implied by $s_{d-1}(w) \leq s_{d-1}(u) \leq 1$.

Explicit constructions of the extreme elements can be described for $\mathcal{S}_u^<$ and $\mathcal{S}_u^>$. In Ref. [59], the extreme elements of $\mathcal{S}_u^>$ have been described using the notion of tangent vectors. Similarly, the extreme points of $\mathcal{S}_u^<$ can be described as follows. To construct one extreme point e_u , fix a subset of indices $k \in \{1, \dots, d-1\}$ at which the cumulative sums of u and e_u are chosen to coincide (they necessarily coincide at 0 and d since $s_0(u) = s_0(e_u) = 0$ and $s_d(u) = s_d(e_u) = 1$). Divide $\{1, \dots, d\}$ into intervals $\{I_1, \dots, I_r\}$, where $I_s = \{k_{s-1} + 1, \dots, k_s\}$ and $0 = k_0 < k_1 < \dots < k_r = d$, with $k_1, \dots, k_{r-1}$ the chosen intermediate coincidence indices. Then, e_u is a block-average of u , i.e., the i^{th} element of vector e_u is given by

$$(e_u)_i = \frac{1}{|I_s|} \sum_{j \in I_s} u_j, \quad i \in I_s. \quad (9.70)$$

This construction determines all extreme points of $\mathcal{S}_u^<$. Different choices of coincidence indices may yield the same point. Hence, the lower cone has at most 2^{d-1} extreme points, since each subset of $\{1, \dots, d-1\}$ specifies a possible set of intermediate coincidence indices.

As an example, for vectors of size 3, $\mathbf{u} = (u_1, u_2, u_3)^T$, we find the following extreme points of $\mathcal{S}_u^<$,

$$\begin{aligned} \{\{1\}, \{2\}, \{3\}\} &\longrightarrow \mathbf{u}, \\ \{\{1, 2\}, \{3\}\} &\longrightarrow \left(\frac{u_1+u_2}{2}, \frac{u_1+u_2}{2}, 1-u_1-u_2\right)^T, \\ \{\{1\}, \{2, 3\}\} &\longrightarrow \left(u_1, \frac{1-u_1}{2}, \frac{1-u_1}{2}\right)^T, \\ \{\{1, 2, 3\}\} &\longrightarrow \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^T. \end{aligned} \tag{9.71}$$

9.5 Beyond the Schur-concavity of the Shannon entropy in the majorization lattice

Schur-concavity plays a central role in the majorization lattice, as it is closely tied to majorization. In particular, the Shannon entropy being Schur-concave, given two vectors $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^1$, the relation $\mathbf{u} < \mathbf{v}$ directly implies $H(\mathbf{u}) \geq H(\mathbf{v})$. This relation was strengthened by Ho and Verdú, who proved that (see Theorem 7.4) if $\mathbf{u} < \mathbf{v}$, then

$$H(\mathbf{u}) - H(\mathbf{v}) \geq D(\mathbf{v} \parallel \mathbf{u}) \geq 0. \tag{9.72}$$

In the above, the Kullback–Leibler divergence (or relative entropy) D is defined as

$$D(\mathbf{v} \parallel \mathbf{u}) := \sum_{i: v_i > 0} v_i \log \frac{v_i}{u_i}, \tag{9.73}$$

whenever $\text{supp}(\mathbf{v}) \subseteq \text{supp}(\mathbf{u})$. If this support condition is not satisfied, we set $D(\mathbf{v} \parallel \mathbf{u}) := +\infty$. In the cases considered below, this support condition is always satisfied. Here, we exploit the majorization lattice to extend Eq. (9.72) to incomparable vectors, going beyond Schur-concavity of entropy.

Theorem 9.1. *Given $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^1$, we have*

$$d_H(\mathbf{u}, \mathbf{v}) \geq D(\mathbf{u} \vee \mathbf{v} \parallel \mathbf{u} \wedge \mathbf{v}), \tag{9.74}$$

where $d_H(\mathbf{u}, \mathbf{v}) = H(\mathbf{u}) + H(\mathbf{v}) - 2H(\mathbf{u} \vee \mathbf{v})$.

When u and v are comparable, e.g., if $u < v$, it is easy to see that Eq. (9.74) reduces to Eq. (9.72). It obviously becomes more interesting when u and v are incomparable, in which case Theorem 9.1 provides us with a relation that can be understood as going beyond the Schur-concavity of the Shannon entropy.

Proof of Theorem 9.1. For ease of notation, write $m := u \wedge v$, $j := u \vee v$, and $\beta := \beta(u, v)$. We shall prove a stronger bound that implies Eq. (9.74), namely,

$$D(j \parallel u) + D(j \parallel v) \geq D(j \parallel m). \quad (9.75)$$

Indeed, Eq. (9.72) implies that

$$\begin{aligned} H(u) - H(j) &\geq D(j \parallel u), \\ H(v) - H(j) &\geq D(j \parallel v), \end{aligned} \quad (9.76)$$

and summing the above two inequalities before making use of Eq. (9.75) directly implies Eq. (9.74).

Upon defining $f_j(x) := -D(j \parallel x)$ for any $x \in \mathcal{S}_j^<$, Eq. (9.75) could essentially be termed the supermodularity property of $f_j(x)$ on the majorization lattice (since $D(j \parallel j) = 0$ trivially). Equivalently, after expanding Eq. (9.75), it actually amounts to proving supermodularity of $g_j(x) := \sum_{i:j_i>0} j_i \log x_i$ (defined for any $x \in \mathcal{S}_j^<$), i.e.,

$$g_j(m) + g_j(j) - g_j(u) - g_j(v) \geq 0, \quad (9.77)$$

or after a straightforward manipulation,

$$\sum_{i:j_i>0} j_i \log \frac{m_i j_i}{u_i v_i} \geq 0. \quad (9.78)$$

We will actually prove that

$$m_i j_i \geq u_i v_i, \quad \forall i = 1, \dots, d, \quad (9.79)$$

implying Eq. (9.78).

We first prove it when (i) $j = \beta$ and then extend it to the case where the PAVA (see Section 7.3.2) is necessary to find the join, i.e., when (ii) $j \neq \beta$.

(i) $j = \beta$. We introduce the k^{th} majorization deficit,

$$\Delta_k := s_k(\mathbf{v}) - s_k(\mathbf{u}). \quad (9.80)$$

Then, we recall the expressions of $\mathbf{m}$ and β , that is, $m_i = \min(s_i(\mathbf{u}), s_i(\mathbf{v})) - \min(s_{i-1}(\mathbf{u}), s_{i-1}(\mathbf{v}))$ (see Eq. (9.2)) and $\beta_i = \max(s_i(\mathbf{u}), s_i(\mathbf{v})) - \max(s_{i-1}(\mathbf{u}), s_{i-1}(\mathbf{v}))$ (see Eq. (9.3)). Obviously,

$$m_i + \beta_i = u_i + v_i, \quad \forall i = 1, \dots, d. \quad (9.81)$$

Four cases can occur. If $\Delta_i \geq 0$ and $\Delta_{i-1} \geq 0$, then $m_i = u_i$ and $\beta_i = v_i$, and incidentally $m_i \beta_i = u_i v_i$. Similarly, if $\Delta_i \leq 0$ and $\Delta_{i-1} \leq 0$, then $m_i = v_i$ and $\beta_i = u_i$, and incidentally $m_i \beta_i = u_i v_i$. If $\Delta_{i-1} \leq 0 \leq \Delta_i$, then $u_i \leq \beta_i \leq v_i$ and $u_i \leq m_i \leq v_i$. Because of Eq. (9.81), we thus get $m_i \beta_i \geq u_i v_i$ from Lemma C.3 (see Appendix C.2). Finally, if $\Delta_{i-1} \geq 0 \geq \Delta_i$, then $v_i \leq \beta_i \leq u_i$ and $v_i \leq m_i \leq u_i$. Because of Eq. (9.81), we get $m_i \beta_i \geq u_i v_i$ from Lemma C.3 again.

(ii) $j \neq \beta$. Denote by $\cup_{\alpha=1}^r I_\alpha$ the union of the r final blocks of the PAVA, that is, j has a constant value on each block (see Section 7.3.2). We need to check that

$$\sum_{\alpha=1}^r \sum_{i \in I_\alpha} j_i \log \left(\frac{m_i j_i}{u_i v_i} \right) \geq 0. \quad (9.82)$$

Since on each block the values of j are constant on the block, we write $j_i = c_\alpha$ for all $i \in I_\alpha$. Hence,

$$\begin{aligned} \sum_{i \in I_\alpha} c_\alpha \log \left(\frac{c_\alpha m_i}{u_i v_i} \right) &= c_\alpha \log \prod_{i \in I_\alpha} \left(\frac{c_\alpha m_i}{u_i v_i} \right) \\ &= c_\alpha \log \left(\frac{c_\alpha^{|I_\alpha|} \prod_{i \in I_\alpha} m_i}{\prod_{i \in I_\alpha} (u_i v_i)} \right), \end{aligned} \quad (9.83)$$

for all $\alpha = 1, \dots, r$. We shall show now that indeed,

$$c_\alpha^{|I_\alpha|} \prod_{i \in I_\alpha} m_i \geq \prod_{i \in I_\alpha} u_i v_i, \quad \forall \alpha = 1, \dots, r. \quad (9.84)$$

By definition, $c_\alpha = \frac{1}{|I_\alpha|} \sum_{i \in I_\alpha} \beta_i$. Then, using the property that the arithmetic mean is greater than or equal to the geometric mean, i.e., $\sum_{i=1}^n x_i \geq n \sqrt[n]{\prod_{i=1}^n x_i}$, we have $c_\alpha^{|I_\alpha|} \geq \prod_{i \in I_\alpha} \beta_i$. Multiplying by $\prod_{i \in I_\alpha} m_i$ on both sides, we get

$$c_\alpha^{|I_\alpha|} \prod_{i \in I_\alpha} m_i \geq \prod_{i \in I_\alpha} m_i \beta_i, \quad \forall \alpha = 1, \dots, r. \quad (9.85)$$

Using what we proved in (i), that is $m_i \beta_i \geq u_i v_i$, for all $i = 1, \dots, d$, Eq. (9.84) is obtained, which completes our proof. $\square$

Making use of Pinsker's inequality [182], i.e.,

$$D(x \parallel y) \geq \frac{2}{\ln 2} d(x, y)^2 \quad (9.86)$$

for probability vectors x and y , together with Theorem 9.1, we can write the following link between the entropic distance d_H and the TV distance d ,

$$d_H(u, v) \geq \frac{2}{\ln 2} d(u \vee v, u \wedge v)^2, \quad (9.87)$$

for any pair of elements $u, v \in \mathcal{P}_d^\downarrow$.

We end this section with a new relation involving the Shannon entropies of u , v , their meet, and their join. Namely, while the supermodularity of the Shannon entropy on the majorization lattice is expressed through the lower bound of Eq. (7.66), we find a symmetric relation by upper bounding the left-hand side of the equation. To do so, we begin by introducing the following lemma.

Lemma 9.6. *Consider two elements u and v in $\mathcal{P}_d^\downarrow$. Then,*

$$u + v < (u \wedge v) + (u \vee v). \quad (9.88)$$

Proof. We know that all cumulative sums of the meet are such that $s_k(u \wedge v) = \min(s_k(u), s_k(v))$, while all cumulative sums of the join satisfy $s_k(u \vee v) \geq \max(s_k(u), s_k(v))$ (to better understand this last relation, see Section 7.3.2 for a graphical visualization of the PAVA applied to $\beta(u, v)$ in order to retrieve $u \vee v$). Hence, $s_k(u \wedge v) + s_k(u \vee v) \geq s_k(u) + s_k(v)$, for all $k = 1, \dots, d$, that is the majorization relation stated. $\square$

We are now in position to state our upper bound on the term in the left-hand side of the supermodularity property (Eq. (7.66)). First, for any $u, v \in \mathcal{P}_d^\downarrow$, define their Jensen-Shannon divergence [154] JS by

$$\text{JS}(u, v) := H\left(\frac{u+v}{2}\right) - \frac{1}{2}H(u) - \frac{1}{2}H(v). \quad (9.89)$$

Theorem 9.2. *Given $u, v \in \mathcal{P}_d^\downarrow$, we have*

$$H(u \wedge v) + H(u \vee v) - H(u) - H(v) \leq 2(\text{JS}(u, v) - \text{JS}(u \wedge v, u \vee v)) - 2D\left(\frac{(u \wedge v) + (u \vee v)}{2} \parallel \frac{u+v}{2}\right). \quad (9.90)$$

Proof. Dividing both terms in Eq. (9.88) by 2 (which does not affect the majorization relation), and applying the tightened version of the Schur-concavity of the Shannon entropy, that is Eq. (9.72), we have

$$H\left(\frac{u+v}{2}\right) - H\left(\frac{(u \wedge v) + (u \vee v)}{2}\right) \geq D\left(\frac{(u \wedge v) + (u \vee v)}{2} \parallel \frac{u+v}{2}\right). \quad (9.91)$$

which implies Eq. (9.90) using the definitions of $\text{JS}(u, v)$ and $\text{JS}(u \wedge v, u \vee v)$. $\square$

Notice that using the following upper bounds to the Jensen-Shannon divergence [48]

$$\text{JS}(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{v}) \leq 1, \quad (9.92)$$

gives a looser bound than Eq. (9.90) as

$$H(\mathbf{u} \wedge \mathbf{v}) + H(\mathbf{u} \vee \mathbf{v}) - H(\mathbf{u}) - H(\mathbf{v}) \leq 2d(\mathbf{u}, \mathbf{v}) \leq 2, \quad (9.93)$$

that is clearly tighter² than the trivial upper bound

$$H(\mathbf{u} \wedge \mathbf{v}) + H(\mathbf{u} \vee \mathbf{v}) - H(\mathbf{u}) - H(\mathbf{v}) \leq 2 \log d. \quad (9.94)$$

9.6 A uniform continuity bound for the Shannon entropy in the majorization diamond

Suppose we obtain a probability distribution by performing some prescribed information-theoretic task. In general, we may lack precise knowledge of that distribution, for instance due to the noise affecting the task. In that case, it is essential to have tight bounds on the error incurred in the task when the actual distribution is replaced by an approximate one. Continuity bounds for entropies are the key ingredient to obtain such stability bounds. More precisely, given two distributions that are at a distance of at most ε , a continuity bound for an entropy is a bound on the difference between their entropies in terms of ε . Continuity bounds are not only of fundamental interest, but also have useful applications in information theory, e.g., in the study of channel capacities.

Consider two probability vectors $\mathbf{p}, \mathbf{q} \in \mathcal{P}_d$ satisfying $d(\mathbf{p}, \mathbf{q}) \leq \varepsilon$, with $\varepsilon \in [0, 1]$. Csiszár proved the following continuity bound relating the Shannon entropies of the two vectors [180]:

$$|H(\mathbf{p}) - H(\mathbf{q})| \leq \begin{cases} \varepsilon \log(d-1) + h(\varepsilon), & \text{if } \varepsilon \leq 1 - \frac{1}{d}, \\ \log d, & \text{otherwise,} \end{cases} \quad (9.95)$$

where $h(\varepsilon) := -\varepsilon \log(\varepsilon) - (1 - \varepsilon) \log(1 - \varepsilon)$ is the binary entropy. The bound is uniform in the sense that it does not depend on the specific vectors $\mathbf{p}$ and $\mathbf{q}$, but only on the parameters ε and d . Moreover, it is tight, since one can find two vectors in $\mathcal{P}_d$ that saturate it. Since Csiszár's work, such continuity bounds have been proven for various information-theoretic quantities, such as the conditional Shannon entropy [5], the Rényi entropy [12], or the Arimoto-Rényi conditional entropy [137]. These bounds have also been extended to quantum systems [6, 82, 137, 138, 247], as well as infinite-dimensional systems, both classical and quantum [13, 16, 17, 247].

²For probability vectors of size greater than 2.

A relevant situation where continuity bounds are exploited is the following: Suppose we have a set of probability distributions $\mathcal{R} \subseteq \mathcal{P}_d$, and we want to compute the quantity

$$\sup_{p \in \mathcal{R}} |H(\mathcal{T}_1(p)) - H(\mathcal{T}_2(p))|, \quad (9.96)$$

where $\mathcal{T}_1$ and $\mathcal{T}_2$ are some stochastic processes which map $\mathcal{R}$ to itself. When the above quantity is difficult to compute, what one can do instead is calculate

$$\sup_{p \in \mathcal{R}} d(\mathcal{T}_1(p), \mathcal{T}_2(p)), \quad (9.97)$$

having in mind that distances are “easier to deal with” than entropies. The quantity in Eq. (9.96) can then be *estimated* with an upper bound by first bounding the distance in Eq. (9.97) and then applying Eq. (9.95), since a continuity bound controls the absolute value of the entropy difference. However, in general, there is no reason for this upper bound to coincide with the quantity in Eq. (9.96). Furthermore, the bound can be loose. Such a situation can be encountered when we have $\mathcal{R} = \mathcal{S}_{a,b}^\diamond$, where $a, b \in \mathcal{P}_d^\downarrow$ satisfy $a < b$. In fact, this is precisely the type of situation encountered in quantum communications, as discussed below in Section 9.8.

With this in mind, we ask the following question: Can we sharpen Eq. (9.95) under the additional constraint that the two vectors p, q belong to a fixed majorization diamond $\mathcal{S}_{a,b}^\diamond$? The resulting bound would still be uniform on the fixed diamond, in the sense that it would not depend on the specific pair p, q , but only on ε, a and b . If $p, q \in \mathcal{S}_{a,b}^\diamond$, Schur-concavity of the Shannon entropy gives the trivial bound

$$|H(p) - H(q)| \leq H(a) - H(b). \quad (9.98)$$

This bound, however, does not use the additional information that $d(p, q) \leq \varepsilon$. It can therefore be loose, especially when $d(a, b) > \varepsilon$. Indeed, we show that we can prove a better bound than that of Eq. (9.98), depending on a and b , but crucially, also on ε .

Let us define the ball of all vectors in $\mathcal{P}_d^\downarrow$ at a distance of at most ε from a vector $x \in \mathcal{P}_d^\downarrow$ as

$$\mathcal{B}_\varepsilon(x) := \{y \in \mathcal{P}_d^\downarrow \mid d(x, y) \leq \varepsilon\}. \quad (9.99)$$

We denote by $M_\varepsilon(x) \in \mathcal{B}_\varepsilon(x)$ the majorization minimum of the ball [114, 115, 131] (see Appendix C.1 for its explicit construction). It is the unique vector satisfying

$$M_\varepsilon(x) < y, \quad (9.100)$$

for every $y \in \mathcal{B}_\varepsilon(x)$.

Theorem 9.3 (Uniform continuity bound for the Shannon entropy inside a majorization diamond). *Let $\mathbf{a}, \mathbf{b} \in \mathcal{P}_d^\downarrow$ satisfy $\mathbf{a} < \mathbf{b}$, and let $\mathbf{p}, \mathbf{q} \in \mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$ satisfy $d(\mathbf{p}, \mathbf{q}) \leq \varepsilon$, with $\varepsilon \in [0, 1]$. Then, the following inequality holds:*

$$|H(\mathbf{p}) - H(\mathbf{q})| \leq H(\mathbf{a} \vee M_\varepsilon(\mathbf{b})) - H(\mathbf{b}). \quad (9.101)$$

Moreover, the inequality is tight when $d(\mathbf{a} \vee M_\varepsilon(\mathbf{b}), \mathbf{b}) \leq \varepsilon$.

Before proving the above theorem, we display an example where our continuity bound $f_{\text{ours}}(\varepsilon, \mathbf{a}, \mathbf{b})$ is strictly better than Csiszár's bound $f_{\text{Csiszár}}(\varepsilon, d)$ (see Eq. (9.95)) and the trivial lattice bound $f_{\text{lattice}}(\mathbf{a}, \mathbf{b})$ (see Eq. (9.98)). We define them by

$$\begin{aligned} f_{\text{ours}}(\varepsilon, \mathbf{a}, \mathbf{b}) &:= H(\mathbf{a} \vee M_\varepsilon(\mathbf{b})) - H(\mathbf{b}), \\ f_{\text{Csiszár}}(\varepsilon, d) &:= h(\varepsilon) + \varepsilon \log(d-1), \\ f_{\text{lattice}}(\mathbf{a}, \mathbf{b}) &:= H(\mathbf{a}) - H(\mathbf{b}), \end{aligned} \quad (9.102)$$

where the displayed expression of $f_{\text{Csiszár}}(\varepsilon, d)$ is the first branch of Csiszár's bound, which applies in the parameter regime considered below. We take $\mathbf{a}, \mathbf{b} \in \mathcal{P}_d^\downarrow$ such that

$$\mathbf{a} := \left(\frac{1}{d}, \dots, \frac{1}{d} \right), \quad (9.103)$$

and

$$\mathbf{b} := \left(1-t, \frac{t}{d-1}, \dots, \frac{t}{d-1} \right), \quad (9.104)$$

where $0 < t < 1 - 1/d$. The condition $t > 0$ excludes the peaked vector, while $t < 1 - 1/d$ ensures that $\mathbf{a} < \mathbf{b}$ with $\mathbf{a} \neq \mathbf{b}$.

Now, choose ε such that

$$0 < \varepsilon < 1 - \frac{1}{d} - t. \quad (9.105)$$

It remains to compute the majorization minimum of the ball of radius ε around $\mathbf{b}$, namely $M_\varepsilon(\mathbf{b})$, whose explicit construction is given in Appendix C.1. Thus, we obtain

$$M_\varepsilon(\mathbf{b}) = \left(1-t-\varepsilon, \frac{t+\varepsilon}{d-1}, \dots, \frac{t+\varepsilon}{d-1} \right). \quad (9.106)$$

Since $\mathbf{a}$ is the uniform vector in $\mathcal{P}_d^\downarrow$, we have $\mathbf{a} < M_\varepsilon(\mathbf{b})$. Therefore, $\mathbf{a} \vee M_\varepsilon(\mathbf{b}) = M_\varepsilon(\mathbf{b})$. Moreover, Eq. (9.105) gives

$$d(\mathbf{a}, \mathbf{b}) = 1 - \frac{1}{d} - t > \varepsilon, \quad (9.107)$$

so $M_\varepsilon(\mathbf{b}) \neq \mathbf{a}$. Hence, by strict Schur-concavity of the entropy,

$$H(M_\varepsilon(\mathbf{b})) < H(\mathbf{a}). \quad (9.108)$$

Consequently,

$$f_{\text{ours}}(\varepsilon, \mathbf{a}, \mathbf{b}) < f_{\text{lattice}}(\mathbf{a}, \mathbf{b}). \quad (9.109)$$

Furthermore, our bound $f_{\text{ours}}(\varepsilon, \mathbf{a}, \mathbf{b})$ can be expressed as

$$f_{\text{ours}}(\varepsilon, \mathbf{a}, \mathbf{b}) = h(t + \varepsilon) - h(t) + \varepsilon \log(d - 1). \quad (9.110)$$

Moreover, Eq. (9.105) implies $\varepsilon < 1 - 1/d$, so the first branch of Csiszár's bound applies. In order to compare this with $f_{\text{Csiszár}}(\varepsilon, d)$, we write

$$\begin{aligned} f_{\text{Csiszár}}(\varepsilon, d) - f_{\text{ours}}(\varepsilon, \mathbf{a}, \mathbf{b}) &= h(\varepsilon) - (h(t + \varepsilon) - h(t)) \\ &> 0, \end{aligned} \quad (9.111)$$

which is true since the binary entropy h is strictly concave, so its increment over an interval of fixed length $\varepsilon > 0$ strictly decreases when the starting point is shifted from 0 to $t > 0$.

Proof of Theorem 9.3. The proof is divided into the steps described below. Starting from arbitrary $\mathbf{p}, \mathbf{q} \in \mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$ with $d(\mathbf{p}, \mathbf{q}) \leq \varepsilon$, we wish to find an upper bound on their entropy difference that depends only on $\mathbf{a}, \mathbf{b}$ and ε . The idea is first to reduce to the case of comparable vectors. We then alter the two vectors $\mathbf{p}$ and $\mathbf{q}$ by transfers between their components that do not increase the distance $d(\mathbf{p}, \mathbf{q})$ and do not decrease the entropy difference $|H(\mathbf{p}) - H(\mathbf{q})|$, while ensuring that the new vectors remain in $\mathcal{S}_{\mathbf{a}, \mathbf{b}}^\diamond$. After the series of such steps, we end up with two final vectors whose entropy difference is our desired upper bound.

Step 0: From (in general) incomparable vectors to comparable ones– The initial step of our proof consists of transforming $\mathbf{p}, \mathbf{q}$ into $\mathbf{p} \wedge \mathbf{q}, \mathbf{p} \vee \mathbf{q}$ in order to only deal with comparable vectors (since $\mathbf{a} < \mathbf{p} \wedge \mathbf{q} < \mathbf{p} \vee \mathbf{q} < \mathbf{b}$), which will prove useful in the following steps. This step is valid since, according to Lemma 9.5,

$$d(\mathbf{p}, \mathbf{q}) \geq d(\mathbf{p} \wedge \mathbf{q}, \mathbf{p} \vee \mathbf{q}), \quad (9.112)$$

so the distance between $\mathbf{p}$ and $\mathbf{q}$ does not increase. In addition, using that $\max\{H(\mathbf{p}), H(\mathbf{q})\} \leq H(\mathbf{p} \wedge \mathbf{q})$, and $\min\{H(\mathbf{p}), H(\mathbf{q})\} \geq H(\mathbf{p} \vee \mathbf{q})$, we get

$$|H(\mathbf{p}) - H(\mathbf{q})| \leq H(\mathbf{p} \wedge \mathbf{q}) - H(\mathbf{p} \vee \mathbf{q}), \quad (9.113)$$

so the entropy difference does not decrease. For ease of notation, we replace the original pair by $\mathbf{p} \wedge \mathbf{q}$ and $\mathbf{p} \vee \mathbf{q}$, and denote this new pair again by $\mathbf{p}$ and $\mathbf{q}$, respectively. Thus, from now on, $\mathbf{p} < \mathbf{q}$, and Schur-concavity gives $H(\mathbf{p}) \geq H(\mathbf{q})$.

Step 1: Move q to the most ordered vector of $\mathcal{S}_{a,b}^\diamond$ – The main idea of this step is to show that there exists a path that transforms any $q \in \mathcal{S}_{a,b}^\diamond$ to the most ordered vector of $\mathcal{S}_{a,b}^\diamond$, that is b , and that any correspondingly admissible p can simultaneously follow a valid path leading it to a vector $p' \in \mathcal{S}_{a,b}^\diamond$. In other words, we show that there exists a finite sequence of transfers, applied either to q alone or simultaneously to p and q , such that they become p' and b respectively, and satisfy $p' \in \mathcal{S}_{a,b}^\diamond$, $d(p', b) \leq d(p, q)$, and $H(p) - H(q) \leq H(p') - H(b)$.

By Lemma C.4 in Appendix C.2, since $q < b$, there exists a finite sequence of vectors $\bar{q}^{(r)}$ ($r = 0, \dots, N_0$) with

$$\bar{q}^{(0)} = q, \quad \bar{q}^{(N_0)} = b, \quad (9.114)$$

and such that each vector of the sequence is transformed into the next one by a so-called reverse T -transform, namely, for every $r = 1, \dots, N_0$,

$$\bar{q}^{(r)} = \bar{q}^{(r-1)} + \delta_r (e_{i_r} - e_{l_r}), \quad 1 \leq i_r < l_r \leq d, \quad (9.115)$$

with $\delta_r > 0$. Moreover, since

$$a < q = \bar{q}^{(0)} < \bar{q}^{(r)} < b, \quad (9.116)$$

all the vectors $\bar{q}^{(r)}$ belong to the diamond.

In general, we cannot apply each whole reverse T -transform to q alone while ensuring that the TV distance to the current vector p does not increase. Therefore, we show that every reverse T -transform in Eq. (9.115) can be realized by at most two substeps, indeed reaching the desired vector $\bar{q}^{(r)}$, but where one of the two substeps is simultaneously applied to p so that the TV distance between them does not increase. Consider one such reverse T -transform, and denote by p_{cur} the current vector p at the beginning of its treatment. For $t \in [0, \delta_r]$, define

$$\bar{q}^{(r)}(t) := \bar{q}^{(r-1)} + t(e_{i_r} - e_{l_r}). \quad (9.117)$$

Since

$$\bar{q}^{(r)}(t) = \left(1 - \frac{t}{\delta_r}\right) \bar{q}^{(r-1)} + \frac{t}{\delta_r} \bar{q}^{(r)}, \quad (9.118)$$

and since the diamond is convex (see Proposition 9.2), every intermediate vector $\bar{q}^{(r)}(t)$ remains in the diamond.

Only components i_r and l_r vary during the transfer of Eq. (9.117). Setting

$$x := \bar{q}_{i_r}^{(r-1)} - p_{\text{cur}, i_r}, \quad y := \bar{q}_{l_r}^{(r-1)} - p_{\text{cur}, l_r}, \quad (9.119)$$

the contribution of these two components (evolving with t) to the TV distance $d(p_{\text{cur}}, \bar{q}^{(r)}(t))$ is

$$\chi(t) := \frac{1}{2} (|x + t| + |y - t|). \quad (9.120)$$

At every t such that $x + t \neq 0$ and $y - t \neq 0$, we have

$$\chi'(t) = \frac{1}{2} (\operatorname{sgn}(x + t) - \operatorname{sgn}(y - t)). \quad (9.121)$$

Hence, $\chi(t)$ can strictly increase only when $x + t > 0$ and $y - t < 0$, and therefore only in the regime

$$(\bar{q}^{(r)}(t))_{i_r} \geq p_{\text{cur}, i_r}, \quad (\bar{q}^{(r)}(t))_{l_r} \leq p_{\text{cur}, l_r}. \quad (9.122)$$

Moreover, since $x + t$ is non-decreasing and $y - t$ is non-increasing, this regime, if reached, begins at most once. If this regime is never reached, the whole transfer is applied to q alone. Otherwise, let t^* be the first time at which the regime in Eq. (9.122) is reached. The part of the transfer on $[0, t^*]$ is applied to q alone, while the remaining part on $[t^*, \delta_r]$ is applied simultaneously to p and q . Thus, whether the transfer is performed in one or two steps, the endpoint $\bar{q}^{(r)}$ is unchanged, and the TV distance does not increase.

Therefore, each reverse T -transform of the initial finite sequence is replaced by at most two nonzero substeps. After relabelling, we obtain a finite sequence

$$q^{(0)} = q, \quad q^{(N)} = b, \quad N \leq 2N_0. \quad (9.123)$$

At each substep k , we write

$$q^{(k)} = q^{(k-1)} + \delta_k(e_i - e_l), \quad 1 \leq i < l \leq d, \quad (9.124)$$

where $\delta_k > 0$ denotes the amount transferred during the considered substep.

We now proceed recursively over these N substeps. At the beginning of substep k , we assume that $p^{(k-1)}$ and $q^{(k-1)}$ belong to the diamond, satisfy the majorization relation

$$p^{(k-1)} < q^{(k-1)}, \quad (9.125)$$

and that the TV distance has not increased and the entropy difference has not decreased during the preceding substeps. These properties hold before the first substep by Step 0. We now show that these recursive assumptions are preserved in each of the two possible cases (when transfers are applied on q alone, and when they are applied on both p and q).

The simple case corresponds to a substep applied to $q^{(k-1)}$ alone. By its construction above, this substep does not increase the TV distance. We set $p^{(k)} := p^{(k-1)}$. Since $q^{(k-1)} < q^{(k)}$, Schur-concavity of entropy gives $H(q^{(k-1)}) \geq H(q^{(k)})$. Therefore,

$$H(p^{(k-1)}) - H(q^{(k-1)}) \leq H(p^{(k)}) - H(q^{(k)}). \quad (9.126)$$

Moreover,

$$\mathbf{p}^{(k)} = \mathbf{p}^{(k-1)} < \mathbf{q}^{(k-1)} < \mathbf{q}^{(k)} < \mathbf{b}. \quad (9.127)$$

Since $\mathbf{p}^{(k)} = \mathbf{p}^{(k-1)}$ and $\mathbf{p}^{(k-1)}$ belongs to the diamond, we also have $\mathbf{a} < \mathbf{p}^{(k)}$. Hence, both vectors remain in the diamond, and the recursive assumptions are preserved in this case.

The more involved case corresponds to the remaining part of a split reverse T -transform. In this substep, we have

$$q_i^{(k-1)} \geq p_i^{(k-1)}, \quad q_l^{(k-1)} \leq p_l^{(k-1)}. \quad (9.128)$$

If so, we apply the remaining part of the reverse T -transform simultaneously to $\mathbf{p}^{(k-1)}$ and $\mathbf{q}^{(k-1)}$. In addition to Eq. (9.124), we thus define

$$\hat{\mathbf{p}}^{(k)} = \mathbf{p}^{(k-1)} + \delta_k(\mathbf{e}_i - \mathbf{e}_l), \quad 1 \leq i < l \leq d. \quad (9.129)$$

Since the same transfer is applied to both vectors, we have

$$\hat{\mathbf{p}}^{(k)} - \mathbf{q}^{(k)} = \mathbf{p}^{(k-1)} - \mathbf{q}^{(k-1)}, \quad (9.130)$$

and therefore

$$d(\hat{\mathbf{p}}^{(k)}, \mathbf{q}^{(k)}) = d(\mathbf{p}^{(k-1)}, \mathbf{q}^{(k-1)}). \quad (9.131)$$

Moreover, $\hat{\mathbf{p}}^{(k)}$ has nonnegative entries. Indeed,

$$\hat{p}_l^{(k)} = p_l^{(k-1)} - \delta_k \geq q_l^{(k-1)} - \delta_k = q_l^{(k)} \geq 0, \quad (9.132)$$

while the other components are clearly nonnegative. Hence, since the transfer also preserves the total mass, $\hat{\mathbf{p}}^{(k)}$ is a probability vector, although it may fail to be non-increasingly ordered. We set

$$\mathbf{p}^{(k)} := \hat{\mathbf{p}}^{(k)\downarrow}, \quad (9.133)$$

and now prove that (i) the TV distance does not increase after sorting, (ii) the entropy difference does not decrease, and (iii) the relevant majorization relations are preserved.

(i) We show that $d(\hat{\mathbf{p}}^{(k)}, \mathbf{q}^{(k)}) \geq d(\mathbf{p}^{(k)}, \mathbf{q}^{(k)})$. Indeed, suppose there exist integers x, y with $x < y$ such that $\hat{p}_x^{(k)} < \hat{p}_y^{(k)}$ (since $\mathbf{q}^{(k)}$ is already non-increasingly ordered, we have $q_x^{(k)} \geq q_y^{(k)}$). We immediately see that

$$|\hat{p}_x^{(k)} - q_x^{(k)}| + |\hat{p}_y^{(k)} - q_y^{(k)}| \geq |\hat{p}_y^{(k)} - q_x^{(k)}| + |\hat{p}_x^{(k)} - q_y^{(k)}| \quad (9.134)$$

because matching small (resp. big) values with small (resp. big) values can only reduce the total variation distance. We repeat those swaps between x and y components until the vector is non-increasingly ordered. Combining this with Eq. (9.131), we obtain

$$d(\mathbf{p}^{(k)}, \mathbf{q}^{(k)}) \leq d(\mathbf{p}^{(k-1)}, \mathbf{q}^{(k-1)}). \quad (9.135)$$

(ii) Rearranging does not affect entropy, i.e.,

$$H(\hat{\mathbf{p}}^{(k)}) = H(\mathbf{p}^{(k)}). \quad (9.136)$$

Therefore, $H(\hat{\mathbf{p}}^{(k)}) - H(\mathbf{q}^{(k)}) = H(\mathbf{p}^{(k)}) - H(\mathbf{q}^{(k)})$. Thus, it remains to prove that

$$H(\mathbf{p}^{(k-1)}) - H(\mathbf{q}^{(k-1)}) \leq H(\hat{\mathbf{p}}^{(k)}) - H(\mathbf{q}^{(k)}). \quad (9.137)$$

To do so, we introduce the function

$$f : [0, \delta_k] \rightarrow \mathbb{R}_{\geq 0} : t \mapsto H(\mathbf{x} + t(\mathbf{e}_i - \mathbf{e}_l)), \quad (9.138)$$

where $\mathbf{x} \in \mathcal{P}_d^1$ is such that $\mathbf{x} + t(\mathbf{e}_i - \mathbf{e}_l)$ remains a probability vector on $[0, \delta_k]$. Since only components i and l of $\mathbf{x}$ vary with t , $f'(t) = \log((x_l - t)/(x_i + t))$. Therefore,

$$H(\mathbf{x}) - H(\mathbf{x} + \delta_k(\mathbf{e}_i - \mathbf{e}_l)) = \int_0^{\delta_k} dt \log \frac{x_i + t}{x_l - t}. \quad (9.139)$$

Since $q_i^{(k-1)} \geq p_i^{(k-1)}$ and $q_l^{(k-1)} \leq p_l^{(k-1)}$, we have, for every $t \in [0, \delta_k]$,

$$\frac{q_i^{(k-1)} + t}{q_l^{(k-1)} - t} \geq \frac{p_i^{(k-1)} + t}{p_l^{(k-1)} - t}. \quad (9.140)$$

Combining this with Eq. (9.139), and treating the endpoint $t = \delta_k$ by continuity, the entropy loss at the level of $\mathbf{q}$'s is at least the entropy loss at the level of $\mathbf{p}$'s, i.e.,

$$H(\mathbf{p}^{(k-1)}) - H(\mathbf{p}^{(k-1)} + \delta_k(\mathbf{e}_i - \mathbf{e}_l)) \leq H(\mathbf{q}^{(k-1)}) - H(\mathbf{q}^{(k-1)} + \delta_k(\mathbf{e}_i - \mathbf{e}_l)). \quad (9.141)$$

(iii) We prove that $\mathbf{a} < \mathbf{p}^{(k)} < \mathbf{q}^{(k)} < \mathbf{b}$. First, by construction of the reverse T -transforms, we have $\mathbf{q}^{(k)} < \mathbf{b}$. Secondly, we show that $\mathbf{a} < \mathbf{p}^{(k)}$. The transfer applied to $\mathbf{p}^{(k-1)}$ (see Eq. (9.129)) moves mass from component l to component i , with $i < l$. Hence, it increases the cumulative sums labelled by indices going from i to $l-1$, by an amount δ_k , and leaves other cumulative sums unchanged. Therefore, for every $\sigma = 1, \dots, d-1$,

$$\sum_{r=1}^{\sigma} \hat{p}_r^{(k)} \geq \sum_{r=1}^{\sigma} p_r^{(k-1)}. \quad (9.142)$$

After rearranging $\hat{\mathbf{p}}^{(k)}$ in non-increasing order, the sum of the first σ components can only increase, so

$$\sum_{r=1}^{\sigma} p_r^{(k)} \geq \sum_{r=1}^{\sigma} \hat{p}_r^{(k)} \geq \sum_{r=1}^{\sigma} p_r^{(k-1)}. \quad (9.143)$$

Thus,

$$\mathbf{p}^{(k-1)} < \mathbf{p}^{(k)}. \quad (9.144)$$

Since $\mathbf{a} < \mathbf{p}^{(k-1)}$ (by the recursive assumption), the transitivity of majorization gives

$$\mathbf{a} < \mathbf{p}^{(k)}. \quad (9.145)$$

Finally, it remains to show that $\mathbf{p}^{(k)} < \mathbf{q}^{(k)}$. We use again the recursive hypothesis (having in mind that $\mathbf{p}^{(0)} < \mathbf{q}^{(0)}$), that is, at substep k , we suppose $\mathbf{p}^{(k-1)} < \mathbf{q}^{(k-1)}$. Then, for every convex function ϕ , we have [161]

$$\sum_{r=1}^d \phi(p_r^{(k-1)}) \leq \sum_{r=1}^d \phi(q_r^{(k-1)}). \quad (9.146)$$

Let us consider

$$\mathbf{p}(t) = \mathbf{p}^{(k-1)} + t(\mathbf{e}_i - \mathbf{e}_l), \quad \mathbf{q}(t) = \mathbf{q}^{(k-1)} + t(\mathbf{e}_i - \mathbf{e}_l), \quad (9.147)$$

with $t \in [0, \delta_k]$. They obey $\mathbf{p}(0) = \mathbf{p}^{(k-1)}$, $\mathbf{p}(\delta_k) = \hat{\mathbf{p}}^{(k)}$, and $\mathbf{q}(0) = \mathbf{q}^{(k-1)}$, $\mathbf{q}(\delta_k) = \mathbf{q}^{(k)}$. During the considered substep, we have

$$q_i^{(k-1)} \geq p_i^{(k-1)}, \quad q_l^{(k-1)} \leq p_l^{(k-1)}. \quad (9.148)$$

For a convex function ϕ , the map $x \mapsto \phi(x+t) - \phi(x)$ is non-decreasing. Hence, using Eq. (9.148),

$$\phi(q_i^{(k-1)} + t) - \phi(q_i^{(k-1)}) \geq \phi(p_i^{(k-1)} + t) - \phi(p_i^{(k-1)}) \quad (9.149)$$

Similarly, the map $x \mapsto \phi(x-t) - \phi(x)$ is non-increasing. Hence, using again Eq. (9.148),

$$\phi(q_l^{(k-1)} - t) - \phi(q_l^{(k-1)}) \geq \phi(p_l^{(k-1)} - t) - \phi(p_l^{(k-1)}) \quad (9.150)$$

Since all other components are left unchanged by the transfers of Eq. (9.147), we have

$$\sum_{r=1}^d \phi(p_r(t)) \leq \sum_{r=1}^d \phi(q_r(t)), \quad (9.151)$$

for all $t \in [0, \delta_k]$. Taking $t = \delta_k$, we get

$$\sum_{r=1}^d \phi(\hat{p}_r^{(k)}) \leq \sum_{r=1}^d \phi(q_r^{(k)}). \quad (9.152)$$

Since this is true for every convex function ϕ , it is equivalent to

$$\hat{p}^{(k)} < q^{(k)}, \quad (9.153)$$

which is itself equivalent to

$$p^{(k)} < q^{(k)}, \quad (9.154)$$

because $p^{(k)} = \hat{p}^{(k)\downarrow}$ and majorization is invariant under permutations.

Consequently, in both cases (when doing a transfer on q alone, or when doing a transfer on both p and q), the recursive assumptions are preserved from substep $k - 1$ to substep k . Since the refined sequence of Eq. (9.123) is finite and satisfies $q^{(N)} = b$, the recursive procedure terminates after N substeps. We denote by $p' := p^{(N)}$ the corresponding vector p at the end of the process.

Step 2: Bounding the remaining entropy difference– At the end of Step 1, the vector q has been transformed into the most ordered element of the diamond, i.e., b . Moreover, the construction provides a vector $p' \in \mathcal{S}_{a,b}^\diamond$ such that

$$d(p', b) \leq \varepsilon, \quad (9.155)$$

and

$$H(p) - H(q) \leq H(p') - H(b). \quad (9.156)$$

It remains to upper bound $H(p')$ by a quantity depending only on ε , a , and b .

Since $d(p', b) \leq \varepsilon$, we have

$$p' \in \mathcal{B}_\varepsilon(b). \quad (9.157)$$

By definition of the majorization minimum $M_\varepsilon(b)$ of the ball $\mathcal{B}_\varepsilon(b)$, this implies

$$M_\varepsilon(b) < p'. \quad (9.158)$$

Thus, p' is an upper bound, in majorization order, of both $M_\varepsilon(b)$ and a . By definition of the join, we obtain

$$a \vee M_\varepsilon(b) < p'. \quad (9.159)$$

Schur-concavity of the Shannon entropy then gives

$$H(p') \leq H(a \vee M_\varepsilon(b)). \quad (9.160)$$

Combining this with Eq. (9.156), we get

$$H(\mathbf{p}) - H(\mathbf{q}) \leq H(\mathbf{a} \vee M_\varepsilon(\mathbf{b})) - H(\mathbf{b}). \quad (9.161)$$

Together with Step 0, this proves the uniform continuity bound of Eq. (9.101).

It remains to justify the tightness statement. Set

$$\mathbf{z} := \mathbf{a} \vee M_\varepsilon(\mathbf{b}). \quad (9.162)$$

If $d(\mathbf{z}, \mathbf{b}) \leq \varepsilon$, then the admissible pair $\mathbf{p} = \mathbf{z}, \mathbf{q} = \mathbf{b}$ attains the bound, since

$$|H(\mathbf{p}) - H(\mathbf{q})| = H(\mathbf{z}) - H(\mathbf{b}) = H(\mathbf{a} \vee M_\varepsilon(\mathbf{b})) - H(\mathbf{b}). \quad (9.163)$$

□

Several relevant particular cases of Theorem 9.3 can be considered.

- If $\mathbf{a} = (1/d, \dots, 1/d)$ and $\mathbf{b} = (1, 0, \dots, 0)$: In this case, as expected, we recover Csiszár's uniform continuity bound expressed in Eq. (9.95). Indeed, $H(\mathbf{b}) = 0$ since $\mathbf{b}$ is the peaked vector. If $\varepsilon \in [0, 1 - 1/d]$, the majorization minimum of the ball of radius ε around $\mathbf{b}$ is

$$M_\varepsilon(\mathbf{b}) = \left(1 - \varepsilon, \frac{\varepsilon}{d-1}, \dots, \frac{\varepsilon}{d-1}\right). \quad (9.164)$$

Since $\mathbf{a}$ is the uniform vector in $\mathcal{P}_d^\downarrow$, we have $\mathbf{a} < M_\varepsilon(\mathbf{b})$ and therefore

$$\mathbf{a} \vee M_\varepsilon(\mathbf{b}) = M_\varepsilon(\mathbf{b}). \quad (9.165)$$

Hence, the bound of Theorem 9.3 gives

$$H(M_\varepsilon(\mathbf{b})) - H(\mathbf{b}) = h(\varepsilon) + \varepsilon \log(d-1), \quad (9.166)$$

which is precisely the first branch of Csiszár's bound. If $\varepsilon > 1 - 1/d$, then the uniform vector belongs to the ball around $\mathbf{b}$, so $M_\varepsilon(\mathbf{b}) = \mathbf{a}$, and the bound becomes $H(\mathbf{a}) - H(\mathbf{b}) = \log d$, corresponding to the second branch of Eq. (9.95).

- If $\mathbf{a} = (1/d, \dots, 1/d)$ and $\mathbf{b}$ is arbitrary: In this case, the lower constraint is inactive, since the uniform vector is majorized by every element of $\mathcal{P}_d^\downarrow$. Hence,

$$\mathbf{a} \vee M_\varepsilon(\mathbf{b}) = M_\varepsilon(\mathbf{b}). \quad (9.167)$$

If $\varepsilon < d(\mathbf{a}, \mathbf{b})$, then $M_\varepsilon(\mathbf{b}) \neq \mathbf{a}$. Otherwise, if $\varepsilon \geq d(\mathbf{a}, \mathbf{b})$, then $M_\varepsilon(\mathbf{b}) = \mathbf{a}$, and the bound reduces to

$$H(\mathbf{a}) - H(\mathbf{b}). \quad (9.168)$$

- If $\mathbf{a}$ is arbitrary and $\mathbf{b} = (1, 0, \dots, 0)$: In this case, the upper constraint $\mathbf{x} < \mathbf{b}$ is inactive, since every element of $\mathcal{P}_d^\perp$ is majorized by the peaked vector. We therefore need to find the expression of the vector $\mathbf{z} := \mathbf{a} \vee M_\varepsilon(\mathbf{b})$ appearing in the bound of Theorem 9.3. It can be described explicitly as follows. If $\mathbf{x} = (x_1, \dots, x_d)$ belongs to the ball of radius ε around $\mathbf{b}$ and is in the diamond, then the TV constraint with respect to $\mathbf{b}$ is expressed as

$$d(\mathbf{x}, \mathbf{b}) = 1 - x_1 \leq \varepsilon. \quad (9.169)$$

Combined with the lower majorization constraint, this gives

$$s_k(\mathbf{x}) \geq B_k, \quad k = 0, \dots, d, \quad (9.170)$$

where

$$B_k := \begin{cases} 0, & k = 0, \\ \max\{1 - \varepsilon, s_k(\mathbf{a})\}, & k = 1, \dots, d-1, \\ 1, & k = d. \end{cases} \quad (9.171)$$

Indeed, $s_k(\mathbf{x}) \geq s_k(\mathbf{a})$ comes from $\mathbf{a} < \mathbf{x}$, while $s_k(\mathbf{x}) \geq x_1 \geq 1 - \varepsilon$ follows from the TV constraint.

The cumulative curve of $\mathbf{z} = \mathbf{a} \vee M_\varepsilon(\mathbf{b})$ is obtained by taking the least concave majorant of the points

$$(k, B_k), \quad k = 0, \dots, d. \quad (9.172)$$

Equivalently, this least concave majorant can be computed by the PAVA procedure described in Section 7.3.2, applied to the successive slopes $B_k - B_{k-1}$, by pooling adjacent slopes whenever they violate the non-increasing order. The slopes of this least concave majorant define a non-increasing probability vector, since the curve is concave and connects $(0, 0)$ to $(d, 1)$. Moreover, since the curve lies above all the points (k, B_k) , the resulting vector satisfies

$$\mathbf{a} < \mathbf{z}, \quad \text{and} \quad d(\mathbf{z}, \mathbf{b}) \leq \varepsilon. \quad (9.173)$$

Conversely, any vector satisfying these two constraints has a concave cumulative-sum curve lying above the same points. By minimality of the least concave majorant, the vector recovered from its slopes is therefore majorized by every vector satisfying Eq. (9.173). Schur-concavity of the entropy then shows that $\mathbf{z}$ is the entropy maximizer in this particular case.

9.7 A uniform continuity bound for the von Neumann entropy with majorization constraints

Let us exploit Theorem 9.3 to prove a continuity bound for the von Neumann entropy of density matrices, subject to majorization constraints. Instead of introducing a majorization diamond directly at the level of quantum states, we consider two reference states $\hat{\omega}$ and $\hat{\gamma}$, whose spectra play the roles of the ante-vector and the post-vector, respectively. Indeed, this scenario is sufficient for our purposes, namely applications for quantum communications.

Theorem 9.4 (Uniform continuity bound for the von Neumann entropy with majorization constraints). *Let $\hat{\rho}, \hat{\sigma}, \hat{\omega}, \hat{\gamma} \in \mathcal{D}(\mathcal{H})$, with $\dim \mathcal{H} = d < \infty$, be such that $\hat{\omega} \prec \hat{\rho} \prec \hat{\gamma}$ and $\hat{\omega} \prec \hat{\sigma} \prec \hat{\gamma}$. Set $\mathbf{a} := \boldsymbol{\lambda}^\downarrow(\hat{\omega})$, and $\mathbf{b} := \boldsymbol{\lambda}^\downarrow(\hat{\gamma})$. If $T(\hat{\rho}, \hat{\sigma}) \leq \varepsilon$, with $\varepsilon \in [0, 1]$, then*

$$|S(\hat{\rho}) - S(\hat{\sigma})| \leq H(\mathbf{a} \vee M_\varepsilon(\mathbf{b})) - H(\mathbf{b}). \quad (9.174)$$

Moreover, the inequality is tight when $d(\mathbf{a} \vee M_\varepsilon(\mathbf{b}), \mathbf{b}) \leq \varepsilon$.

Proof. Write $\mathbf{p} := \boldsymbol{\lambda}^\downarrow(\hat{\rho})$ and $\mathbf{q} := \boldsymbol{\lambda}^\downarrow(\hat{\sigma})$. We have

$$S(\hat{\rho}) = H(\mathbf{p}), \quad S(\hat{\sigma}) = H(\mathbf{q}). \quad (9.175)$$

Thus,

$$|S(\hat{\rho}) - S(\hat{\sigma})| = |H(\mathbf{p}) - H(\mathbf{q})|. \quad (9.176)$$

By definition of majorization for density operators, the conditions $\hat{\omega} \prec \hat{\rho} \prec \hat{\gamma}$ and $\hat{\omega} \prec \hat{\sigma} \prec \hat{\gamma}$ are equivalent to $\mathbf{a} \prec \mathbf{p} \prec \mathbf{b}$ and $\mathbf{a} \prec \mathbf{q} \prec \mathbf{b}$, respectively. Using the well-known relation [16, 215]

$$\frac{1}{2} \sum_{i=1}^d |\lambda_i^\downarrow(\hat{\rho}) - \lambda_i^\downarrow(\hat{\sigma})| \leq \frac{1}{2} \|\hat{\rho} - \hat{\sigma}\|_1, \quad (9.177)$$

we have

$$d(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \sum_{i=1}^d |p_i - q_i| \leq T(\hat{\rho}, \hat{\sigma}) \leq \varepsilon. \quad (9.178)$$

Therefore, we find $d(\mathbf{p}, \mathbf{q}) \leq \varepsilon$, i.e., the spectra of $\hat{\rho}$ and $\hat{\sigma}$ satisfy the hypotheses of the continuity bound of Theorem 9.3. Applying the bound gives

$$|H(\mathbf{p}) - H(\mathbf{q})| \leq H(\mathbf{a} \vee M_\varepsilon(\mathbf{b})) - H(\mathbf{b}). \quad (9.179)$$

Hence, by Eq. (9.176),

$$|S(\hat{\rho}) - S(\hat{\sigma})| \leq H(\mathbf{a} \vee M_\varepsilon(\mathbf{b})) - H(\mathbf{b}). \quad (9.180)$$

Finally, the bound is tight whenever $d(\mathbf{a} \vee M_\varepsilon(\mathbf{b}), \mathbf{b}) \leq \varepsilon$ since it suffices to take commuting states $\hat{\rho}^*$ and $\hat{\sigma}^*$, with eigenvalues $\mathbf{a} \vee M_\varepsilon(\mathbf{b})$ and $\mathbf{b}$, respectively. $\square$

9.8 Discussion and perspectives

We now discuss the relevance of our results in the context of quantum communications. We focus in particular on Theorem 9.4, which extends Theorem 9.3 to the quantum realm, and builds on our other main results.

One of the most challenging problems in quantum information theory is to derive closed-form or tractable expressions for the capacities of quantum channels, i.e., the highest rates at which they can reliably transmit information. One such quantity is the *quantum capacity*. Consider a quantum channel (i.e., a completely positive trace preserving map) $\Phi : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$, where $\mathcal{H}_A$ is the Hilbert space of the input system, while $\mathcal{H}_B$ is that of the output system. One can associate with Φ a so-called complementary channel $\Phi^c : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_E)$ of Φ , where $\mathcal{H}_E$ is viewed as the Hilbert space of the environment (see, for instance, Ref. [138] for details on the channel's definition). The quantum capacity Q of Φ is given by the so-called LSD theorem [68, 155, 214]:

$$Q(\Phi) = \lim_{k \rightarrow \infty} \frac{1}{k} Q^{(1)}(\Phi^{\otimes k}), \quad (9.181)$$

where $Q^{(1)}(\Phi)$ denotes the coherent information of channel Φ and is given by

$$Q^{(1)}(\Phi) := \max_{\hat{\rho} \in \mathcal{D}(\mathcal{H}_A)} (S(\Phi(\hat{\rho})) - S(\Phi^c(\hat{\rho}))). \quad (9.182)$$

The regularization (i.e., the limit) appearing in Eq. (9.181) makes the quantum capacity difficult to compute in general.

To circumvent this problem, various techniques have been proposed to compute upper bounds on capacities instead. One such successful method involves the use of continuity bounds for entropies [138, 153, 227]. In a nutshell, one looks for a so-called *approximate degrading* channel $\Lambda : \mathcal{L}(\mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_E)$ that minimizes the following *diamond norm* distance

$$\|\Phi^c - \Lambda \circ \Phi\|_{\diamond} := \|(\Phi^c - \Lambda \circ \Phi) \otimes \text{id}_{A'}\|_{1 \rightarrow 1}, \quad (9.183)$$

with, for $\Phi : \mathcal{L}(\mathcal{H}_A) \rightarrow \mathcal{L}(\mathcal{H}_B)$,

$$\|\Phi\|_{1 \rightarrow 1} := \max \{ \|\Phi(\hat{X})\|_1 : \hat{X} \in \mathcal{L}(\mathcal{H}_A), \|\hat{X}\|_1 \leq 1 \}, \quad (9.184)$$

and where $\text{id}_{A'}$ denotes the identity map on A' , a copy of system A . By plugging the value of the diamond norm (for the trace distance) into a continuity bound such as that of Eq. (9.95) (which can also be expressed for density matrices), one can derive an upper bound on the quantum capacity [138, 153, 227]. The core idea is thus that one chooses to compute trace distances between channel outputs, rather than entropy differences. In fact, this is exactly the situation described at the beginning of Section 9.6 (see Eqs. (9.96) and (9.97)), albeit in the context of quantum communication.

We now argue that Theorem 9.4 may lead to better upper bounds on quantum capacities (and other capacities, e.g., the *private classical capacity* [30, 68]) using this technique. Indeed, by characterizing in more detail the set of the output states of Φ and Φ^c , one could in principle exploit the constrained continuity bound of Theorem 9.4, thereby leading to better capacity bounds. This can be done by identifying a lower (majorized) state and an upper (majorizing) one at the output of the quantum channels. We now briefly illustrate our idea with the example of the *quantum depolarizing channel*, a ubiquitous noise model for which upper bounds on the quantum and classical private capacities were obtained using the technique described above [138, 227].

The depolarizing channel $\mathcal{E}_\mu : \mathcal{D}(\mathcal{H}_A) \rightarrow \mathcal{D}(\mathcal{H}_B)$ with $\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathbb{C}^d$, is defined as [172]

$$\mathcal{E}_\mu(\hat{\rho}) := \mu\hat{\rho} + \frac{1-\mu}{d}\hat{\mathbb{I}}_d, \quad (9.185)$$

where $\hat{\mathbb{I}}_d$ is the d -dimensional identity operator. The condition of complete positivity of the channel requires $-1/(d^2 - 1) \leq \mu \leq 1$. Now, for any $\hat{\rho}, \hat{\sigma} \in \mathcal{D}(\mathcal{H}_A)$, $\hat{\rho} < \hat{\sigma}$ implies $\mathcal{E}_\mu(\hat{\rho}) < \mathcal{E}_\mu(\hat{\sigma})$. Indeed, if $\hat{\rho} < \hat{\sigma}$, there exists a doubly stochastic matrix D such that $\lambda(\hat{\rho}) = D\lambda(\hat{\sigma})$ [161]. Thus, denoting by ν the d -dimensional uniform distribution,

$$\begin{aligned} \lambda(\mathcal{E}_\mu(\hat{\rho})) &= \mu\lambda(\hat{\rho}) + (1-\mu)\nu \\ &= \mu D\lambda(\hat{\sigma}) + (1-\mu)\nu \\ &= D\lambda(\mathcal{E}_\mu(\hat{\sigma})), \end{aligned} \quad (9.186)$$

so that $\mathcal{E}_\mu(\hat{\rho}) < \mathcal{E}_\mu(\hat{\sigma})$. From this, we see that we can choose any pure quantum state (whose density matrix is of rank 1) $|\psi\rangle \in \mathcal{H}_A$, and for any $\hat{\rho} \in \mathcal{D}(\mathcal{H}_A)$, we thus have

$$\mathcal{E}_\mu(\hat{\rho}) < \hat{\gamma}_\mu := \mu|\psi\rangle\langle\psi| + \frac{1-\mu}{d}\hat{\mathbb{I}}_d. \quad (9.187)$$

In fact, the spectrum of $\hat{\gamma}_\mu$ does not depend on the choice of $|\psi\rangle$.

Now, the case of the complementary channel $\mathcal{E}_\mu^c : \mathcal{D}(\mathcal{H}_A) \rightarrow \mathcal{D}(\mathcal{H}_E)$ of the depolarizing channel is more involved, since it does not admit a simple expression of the form of Eq. (9.185) [55, 125]. However, preliminary numerical tests suggest that a relation similar to that of Eq. (9.187) holds for the complementary channel $\mathcal{E}_\mu^c$. From these observations, we expect the diamond norm distance of Eq. (9.183) in the case of the depolarizing channel, namely the quantity $\|\mathcal{E}_\mu^c - \Lambda \circ \mathcal{E}_\mu\|_\diamond$, to be minimized when the channel $\Lambda \circ \mathcal{E}_\mu$ also satisfies a relation of the form of Eq. (9.187). If this is indeed the case as expected, we would be in the setting of Theorem 9.4, by taking the appropriate $\hat{\gamma}$ of the form of $\hat{\gamma}_\mu$ in the right-hand side of Eq. (9.187), and by simply taking $\hat{\omega}$ to be the maximally mixed state $\hat{\mathbb{I}}_{d_E}/d_E$ of the environment system, where $d_E = \dim \mathcal{H}_E$.

Applying Theorem 9.4 in the context of the method described above would then lead to improved bounds on the capacities of the depolarizing channel. In fact, we would be exactly in the setting of the example highlighted just below Theorem 9.3, i.e., with $\hat{\omega}$ and $\hat{\gamma}$ having spectra of the form of Eqs. (9.103) and (9.104), respectively, for which we have shown that our continuity bound can provide an improvement on Eq. (9.95). Since the focus on quantum channel capacities is out of the scope of the present work, we leave these questions for future investigation.

In view of the above discussion, we conclude with some open problems. The first obvious open question concerns what is discussed in the previous paragraph. Namely, the main question would be to exploit Theorem 9.4 to derive improved upper bounds for channel capacities. In particular, as already explained, a more detailed analysis of the output set of the complementary channel of the depolarizing channel would be required.

In the same context, it turns out that continuity bounds for the quantum *conditional* entropy are also useful in deriving capacity estimates [138, 227]. In relation to the present work, an interesting question would thus be to derive new continuity bounds for the Shannon conditional entropy $H(X|Y)$ of a pair of random variables (X, Y) (we consider here a different convention than the one taken in the definition of the Shannon entropy, see Eq. (7.23), for the sake of the discussion). As in Theorem 9.3, the joint probability distribution p_{XY} of the variables (X, Y) could be constrained to belong to a majorization diamond. However, one would need to identify which space should be constrained by such a diamond: the distributions associated with X , those associated with Y , or rather those associated with the pair (X, Y) ? The answer to this question would be guided by the role of continuity bounds for the conditional entropy in the derivation of capacity bounds (see Refs. [138, 227]).

10 Conclusion and perspectives

This thesis was centered around two main themes: entanglement detection in continuous-variable systems and the study of the majorization lattice in quantum information. The first part focused on moment-based entanglement criteria and their experimental implementation in continuous-variable systems. The second part investigated how the lattice structure of majorization can be exploited in problems related to entanglement transformations and entropic inequalities.

In Chapter 5, we first extended entanglement criteria based on moments of the partially transposed density operator to continuous-variable systems. The complete hierarchies of PT-moment conditions recover the PPT condition, while their low-order instances provide experimentally accessible tests that require only a few replicas of the state. As a central result of this chapter, it was shown that such low-order PT moments can remain sensitive to strongly non-Gaussian entanglement, since they may implicitly contain information about mode-operator correlations of arbitrarily high order. This feature was illustrated in particular for NOON states, for which standard criteria based on second moments or entropies fail to detect entanglement, while mode-operator-based criteria require moments of increasingly high order as the photon number increases. We further showed that the corresponding multicopy implementation, based on passive linear optics and photon-number-resolving detection, remains robust against losses, noisy state copies, and finite statistics. These results establish PT moments as practical tools for certifying non-Gaussian entanglement beyond the Gaussian regime.

In Chapter 6, we investigated a complementary approach based on the realignment criterion. Instead of working directly with the trace norm of the realigned density operator, which is not immediately accessible experimentally in the continuous-variable regime, we introduced criteria formulated in terms of moments of the realigned operator. In particular, the second-order conditions improve upon the weak realignment criterion and were shown to detect entanglement in Gaussian states, two-mode entangled cat states, and lossy NOON states. We also derived continuous-variable measurement schemes based on homodyne and parity detection. Although the realignment-based criteria considered here do not systematically outperform the PT-moment criteria in terms of detection capability, their experimental implementation has a different structure: coherence between fewer copies is required at a given time, while measurements are based on homodyne and parity detection. Thus, the two approaches developed in Chapters 5 and 6 provide complementary ways of accessing nonlinear separability criteria in continuous-variable systems.

The second theme of this thesis concerned the majorization lattice and the information carried by its specific notions of meet and join. In Chapter 8, we gave these two lattice operations an operational interpretation in the context of probabilistic transformations between incomparable bipartite pure states. The meet and join naturally define, respectively, an optimal common resource state and an optimal common product state, which in turn lead to the thrifty and greedy conversion protocols. Both protocols achieve the same maximal success probability as Vidal's original optimal protocol. Their behavior differs, however, when the conversion fails: the thrifty protocol leaves a more entangled residual state and therefore preserves the entanglement resource better on average. For comparable initial and target states, both constructions reduce to Vidal's protocol, emphasizing that the additional lattice structure becomes relevant precisely in the presence of incomparability. We also generalized these constructions to situations involving several possible initial or final states.

In Chapter 9, we explored a different aspect of the same lattice structure by studying its geometry through the total variation distance and relating it to entropic quantities defined on the lattice. This led us to introduce the notion of a majorization diamond, consisting of probability distributions constrained between fixed lower and upper bounds in the majorization order. We established several relations between distances, the meet, and the join, and used this structure to derive an entropic inequality going beyond the standard Schur-concavity property. Building on these results, we obtained a continuity bound for the Shannon entropy under majorization constraints and subsequently extended it to the von Neumann entropy of quantum states. These results show that prior information expressed through majorization can lead to more refined control of entropic quantities than bounds formulated over the entire probability simplex or state space. They also suggest possible applications in quantum communication, where continuity bounds play an important role in estimating quantum-channel capacities. Taken together, Chapters 8 and 9 illustrate two complementary roles of the majorization lattice. In the former, the meet and join identify states with an operational meaning in resource conversion, whereas in the latter they enter the construction of geometric constraints, entropic inequalities, and continuity bounds.

The results of this thesis also leave several questions open, both from an experimental and a theoretical point of view. Although an alternative single-copy homodyne implementation of the p_3 -PPT criteria presented in Chapter 5 has recently been proposed and analyzed in Ref. [224], the requirement of several copies remains a potential experimental challenge for the multicopy approach. Therefore, inspired notably by Ref. [35], it would be interesting to investigate whether the number of required copies could be reduced through the use of auxiliary states, potentially at the expense of a reduced entanglement-detection capability. Furthermore, our analysis of the impact of imperfections on NOON states could be extended to other non-Gaussian states relevant to current experimental settings [241]. The criteria

could also be generalized to multimode and multipartite systems, potentially leading to new multipartite entanglement witnesses based on multicopy measurements.

Similarly, it would be interesting to investigate the robustness of the realignment criteria presented in Chapter 6 under realistic experimental conditions. This includes studying their resilience to imperfect copies and finite statistics, as well as determining how many measurements are required to obtain sufficiently accurate estimates of the relevant moments. Another important question is whether physically relevant states exist that are detected by the proposed realignment criteria but remain undetected by the PT-moment criteria considered in this thesis. Of particular interest in this context would be PPT bound entangled states, which by construction are not detected by the PPT criterion but may potentially be accessible through realignment-based criteria.

In Chapter 8, we constructed alternative optimal probabilistic pure-state conversion protocols using the majorization lattice. It would be interesting to generalize this framework to mixed-state entanglement transformations [102, 235], although this problem remains considerably more complex. Multipartite entanglement transformations [74] would also provide an interesting setting in which to investigate whether the meet and join retain a useful operational interpretation. Such extensions could benefit from the deeper understanding of the geometry of the majorization lattice developed in Chapter 9. The continuity bounds obtained there could potentially be strengthened further, in particular by identifying broader conditions under which they are tight. Another direction would be to extend the same ideas to other information-theoretic quantities, such as conditional entropies, for which continuity bounds play a central role in quantum information. Finally, it would be interesting to determine under which conditions majorization-constrained continuity bounds can be used to estimate quantum-channel capacities within a fixed majorization diamond, beyond the preliminary discussion of the depolarizing channel given at the end of Chapter 9.

Declaration on the use of Artificial Intelligence

Artificial intelligence tools (GPT-5.5 and GPT-5.6 Sol) were used during the preparation of this thesis as supporting tools.

They were used to assist in the creation of certain figures. The resulting material required multiple attempts, manual modifications, and substantial refinement before being incorporated into the final document. All visual content was reviewed and adapted by the author, who remains responsible for its accuracy and relevance.

Artificial intelligence was also used to improve the phrasing of passages that appeared unclear or non-idiomatic, to identify typographical and grammatical errors, and, in some instances, to assist in checking calculations or intermediate results. AI-generated output was not treated as an authoritative source, and any mathematical or computational result for which AI was used as a checking aid was independently verified by the author.

The ideas, arguments, interpretations, and overall structure of the thesis were developed by the author. In a limited part of the appendix, AI was used to assist in the development of some mathematical ideas; this use is explicitly indicated at the relevant point. All such developments were independently checked and verified by the author.

All factual information and external claims are supported by identifiable sources and cited accordingly. Results that are not derived from external sources are established, calculated, or demonstrated by the author within the thesis.

The author takes full responsibility for the accuracy, integrity, and final content of this work.

Part IV

Appendices

Appendix to Chapter 5

We give here detailed calculations for Chapter 5, devoted to the detection of strongly non-Gaussian entanglement.

A.1 Gaussian states

A.1.1 Two-mode squeezed state at finite temperature

To illustrate the breakdown of the various PT-moment-based criteria for sufficiently mixed Gaussian states, we consider a two-mode squeezed state at finite temperature. Starting from a product of two identical thermal states $\hat{\rho}_{\bar{n}} = \sum_{n=0}^{\infty} \frac{\bar{n}^n}{(\bar{n}+1)^{(n+1)}} |n\rangle \langle n|$ with mean particle number $\bar{n} \geq 0$, this state is generated via the two-mode squeezer unitary $\hat{U}_r = \exp[r(\hat{a}\hat{b} - \hat{a}^\dagger\hat{b}^\dagger)/2]$ with squeezing parameter $r \in [0, \infty)$, resulting in

$$\hat{\rho}_{\bar{n},r} = \hat{U}_r(\hat{\rho}_{\bar{n}} \otimes \hat{\rho}_{\bar{n}})\hat{U}_r^\dagger. \quad (\text{A.1})$$

One can easily find that the symplectic eigenvalues of the partially transposed state are $\tilde{\nu}_1 = e^{-2r}(2\bar{n}+1)$ and $\tilde{\nu}_2 = e^{2r}(2\bar{n}+1)$, such that the state's purity depends on the mean particle number only, i.e., $p_2 = 1/(2\bar{n}+1)^2$.

When $\bar{n} = 0$, Eq. (A.1) reduces to the pure two-mode squeezed vacuum state [see dashed black curves in Figures 5.2b) and 5.2c)], which is entangled for any finite squeezing $r > 0$. Upon increasing $\bar{n}$, the state becomes more mixed, with purity $p_2 = 1/2$ (see dashed yellow curves) being attained for $\bar{n} = (\sqrt{2} - 1)/2 \approx 0.207$ particles per mode. In this regime, sufficient squeezing $r > \ln(2\bar{n}+1)/2$ is required for the state to be entangled at all. For the above case, this entails $r > \ln(2)/4 \approx 0.173$ or $p_3 < 4/13$ (see yellow squares). However, all third-order criteria require $r \gtrsim 0.363$ or $p_3 < 1/4$ to detect entanglement (see yellow diamonds). Note that physicality translates into $p_3 \leq 16/49$ for $p_2 = 1/2$ [see yellow triangle in Figure 5.2b)], showing that the range of values for p_3 compatible with a separable state is small compared to an entangled state [see the light gray region in the inset of Figure 5.2b)].

A.1.2 Higher-order criteria

We analyze higher-order p_n -PPT criteria from the Stieltjes approach [cf. Eq. (5.2)] for Gaussian states in Figure A.1. For the two-mode squeezed state at finite temperature considered in Appendix A.1.1, the p_n -PPT criteria for $n = 5$ (green curve) and $n = 7$ (red curve) flag entanglement already when $r \gtrsim 0.221$ (yellow squares) and $r \gtrsim 0.187$ (yellow triangles), respectively, compared to $r \gtrsim 0.363$ (yellow diamonds) for $n = 3$.

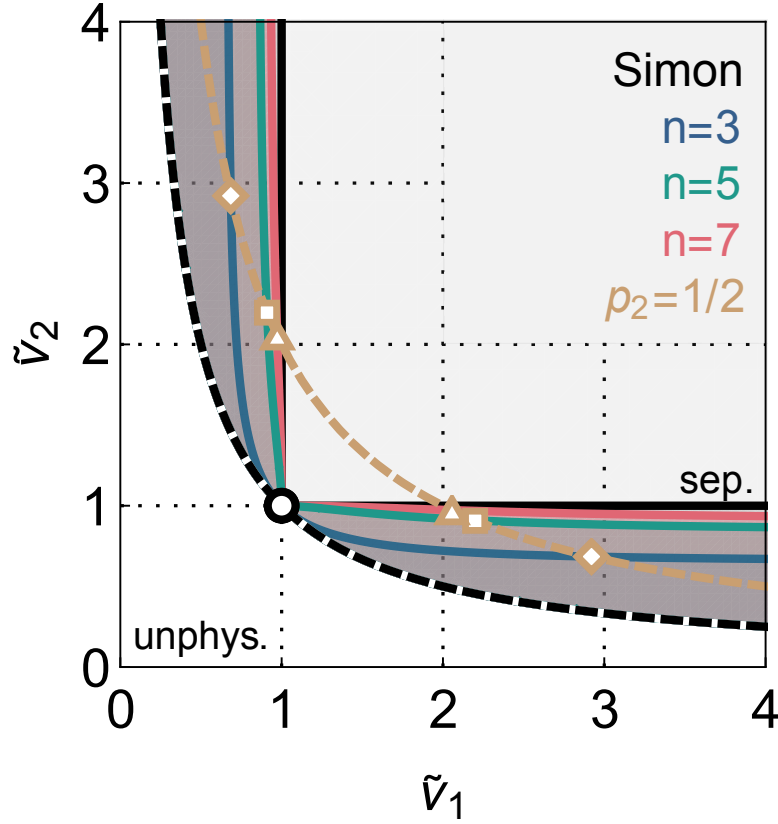


Figure A.1: Gaussian entanglement detected by the p_n -PPT criteria (5.2) of order $n = 3$ (blue), $n = 5$ (petrol), and $n = 7$ (red), analogous to Figure 2a). More mixed entangled states are detected for increasing n , indicating convergence toward Simon's criterion in the limit $n \rightarrow \infty$.

A.2 Mixed entangled cat states

To evaluate the PT moments for the class of states (5.42), we make abundantly use of the coherent-state inner product $\langle \alpha | \beta \rangle = e^{-\frac{1}{2}(|\alpha|^2 + |\beta|^2) + \alpha \beta^* - \alpha^* \beta}$, which leads to

$$p_2 = \frac{2}{C^2} \left[(1 + (1-z)^2) \left(1 + e^{-4(|\alpha|^2 + |\beta|^2)} \right) \pm 4(1-z) e^{-2(|\alpha|^2 + |\beta|^2)} \right] \quad (\text{A.2})$$

and

$$p_3 = \frac{2}{C^3} \left[1 + 3e^{-4(|\alpha|^2 + |\beta|^2)} \pm 3(1-z) e^{-2(|\alpha|^2 + |\beta|^2)} \left(2 + e^{-4|\alpha|^2} + e^{-4|\beta|^2} \right) \right. \\ \left. + 3(1-z)^2 e^{-4(|\alpha|^2 + |\beta|^2)} \left(2 + e^{4|\alpha|^2} + e^{4|\beta|^2} \right) \pm (1-z)^3 e^{-6(|\alpha|^2 + |\beta|^2)} \left(1 + 3e^{4(|\alpha|^2 + |\beta|^2)} \right) \right]. \quad (\text{A.3})$$

A.3 High Harmonic Generation (HHG)

The reduced state of the driving field and one harmonic mode for equal harmonic shifts $\chi := \chi_q = (2/\sqrt{N^2 + 2N - 3})\delta\alpha$ reads

$$\hat{\rho} = \frac{1}{C} \left[|\alpha - \delta\alpha, \chi\rangle \langle \alpha - \delta\alpha, \chi| - e^{-\frac{N^2+6N-11}{2(N^2+2N-3)}|\delta\alpha|^2} |\alpha - \delta\alpha, \chi\rangle \langle \alpha, 0| \right. \\ \left. - e^{-\frac{N^2+6N-11}{2(N^2+2N-3)}|\delta\alpha|^2} |\alpha, 0\rangle \langle \alpha - \delta\alpha, \chi| + e^{-\frac{N^2+4N-5}{N^2+2N-3}|\delta\alpha|^2} |\alpha, 0\rangle \langle \alpha, 0| \right], \quad (\text{A.4})$$

where we assumed that $\alpha, \delta\alpha$, and χ are real without loss of generality. The purity evaluates to

$$p_2 = \frac{1}{C^2} \left[1 - 2(1 - C) \left(2 - e^{-(|\delta\alpha|^2 + |\chi|^2)} \right) + (1 - C)^2 \left(2e^{|\delta\alpha|^2 + |\chi|^2} - 1 \right) \right], \quad (\text{A.5})$$

while the third PT moment is

$$p_3 = \frac{1}{C^3} \left[1 - 3(1 - C) \left(2 - e^{-(|\delta\alpha|^2 + |\chi|^2)} \right) \right. \\ \left. + 3(1 - C)^2 \left(2 + e^{-(|\delta\alpha|^2 + |\chi|^2)} + e^{|\delta\alpha|^2} - 2e^{-|\delta\alpha|^2} + e^{|\chi|^2} - 2e^{-|\chi|^2} \right) \right. \\ \left. - (1 - C)^3 \left(1 + 6e^{|\delta\alpha|^2 + |\chi|^2} - 3e^{|\delta\alpha|^2} - 3e^{|\chi|^2} \right) \right]. \quad (\text{A.6})$$

A.4 NOON states

A.4.1 Mode-operator moments

We compute the set of constraint mode-operator moments for the class of NOON states. To this end, we normal order Eq. (5.46), for which we use the identity [209]

$$|0\rangle \langle 0| =: e^{-\hat{a}^\dagger \hat{a}} = \sum_{j=0}^{\infty} \hat{a}^{\dagger j} \hat{a}^j \frac{(-1)^j}{j!}. \quad (\text{A.7})$$

We end up with

$$\rho_{ijkl} = |\alpha|^2 \frac{(-1)^{i+k-N}}{N!(i-N)!k!} \delta_{ij} \delta_{kl} \Theta(i-N) + \alpha^* \beta \frac{(-1)^{i+l}}{N!i!l!} \delta_{ij-N} \delta_{kl+N} \Theta(j-N) \Theta(k-N) \\ + \alpha \beta^* \frac{(-1)^{j+k}}{N!j!k!} \delta_{ij+N} \delta_{kl-N} \Theta(i-N) \Theta(l-N) + |\beta|^2 \frac{(-1)^{i+k-N}}{N!i!(k-N)!} \delta_{ij} \delta_{kl} \Theta(k-N), \quad (\text{A.8})$$

where $\Theta(x)$ denotes the Heaviside Θ function (with the convention $\Theta(0) = 1$ understood).

This leads to the expansion for the purity

$$\begin{aligned}
p_2 \sim & \frac{|\alpha|^2}{N!} \left[\langle \hat{a}^{\dagger N} \hat{a}^N \rangle + \langle \hat{a}^{\dagger(N+\dots)} \hat{a}^{N+\dots} \hat{b}^{\dagger(0+\dots)} \hat{b}^{0+\dots} \rangle \right] \\
& + \frac{\alpha^* \beta}{N!} \left[\langle \hat{a}^N \hat{b}^{\dagger N} \rangle + \langle \hat{a}^{\dagger(0+\dots)} \hat{a}^{N+\dots} \hat{b}^{\dagger(N+\dots)} \hat{b}^{0+\dots} \rangle \right] \\
& + \frac{\alpha \beta^*}{N!} \left[\langle \hat{a}^{\dagger N} \hat{b}^N \rangle + \langle \hat{a}^{\dagger(N+\dots)} \hat{a}^{0+\dots} \hat{b}^{\dagger(0+\dots)} \hat{b}^{N+\dots} \rangle \right] \\
& + \frac{|\beta|^2}{N!} \left[\langle \hat{b}^{\dagger N} \hat{b}^N \rangle + \langle \hat{a}^{\dagger(0+\dots)} \hat{a}^{0+\dots} \hat{b}^{\dagger(N+\dots)} \hat{b}^{N+\dots} \rangle \right],
\end{aligned} \tag{A.9}$$

where center dots indicates higher-order contributions. Only the first term in each row survives, and the purity reduces to

$$p_2 = \frac{1}{N!} \left[|\alpha|^2 \langle \hat{a}^{\dagger N} \hat{a}^N \rangle + \alpha^* \beta \langle \hat{a}^N \hat{b}^{\dagger N} \rangle + \alpha \beta^* \langle \hat{a}^{\dagger N} \hat{b}^N \rangle + |\beta|^2 \langle \hat{b}^{\dagger N} \hat{b}^N \rangle \right]. \tag{A.10}$$

Similarly, we find for the third PT moment

$$p_3 = \frac{1}{N!} \left[|\alpha|^4 \langle \hat{a}^{\dagger N} \hat{a}^N \rangle + |\beta|^4 \langle \hat{b}^{\dagger N} \hat{b}^N \rangle \right]. \tag{A.11}$$

It is straightforward to verify consistency of the latter two equations with $p_2 = 1$ and Eq. (5.47) since $\langle \hat{a}^{\dagger N} \hat{a}^N \rangle = |\alpha|^2 N!$ (analogously for $\hat{b}$), and $\langle \hat{a}^N \hat{b}^{\dagger N} \rangle = \langle \hat{a}^{\dagger N} \hat{b}^N \rangle^* = \alpha \beta^* N!$ for NOON states.

Equations (A.10) and (A.11) nicely demonstrate the main strength of PT-moment-based entanglement criteria: In contrast to entropic and Shchukin-Vogel-type criteria, where higher-order mode-operator moments are exponentially suppressed or hard to access, respectively, PT moments often select the right (and in principle arbitrarily high) mode-operator moments that contain the necessary information about quantum correlations.

A.4.2 Lossy NOON states

Since we account for noisy copies by fluctuating state parameters, we consider in general different losses for Alice and Bob. Then, the lossy NOON state reads

$$\hat{\rho} = (\mathcal{L}_{\tau_A} \otimes \mathcal{L}_{\tau_B}) (|\psi\rangle \langle \psi|), \tag{A.12}$$

where $|\psi\rangle$ is the lossless NOON state as defined in Eq. (5.46). Direct calculation leads to

$$\begin{aligned}
\hat{\rho} = & |\alpha|^2 \sum_{k=0}^N \binom{N}{k} \tau_A^{(N-k)} (1 - \tau_A)^k |N-k, 0\rangle \langle N-k, 0| \\
& + \sqrt{\tau_A \tau_B}^N (\alpha \beta^* |N, 0\rangle \langle 0, N| + \alpha^* \beta |0, N\rangle \langle N, 0|) \\
& + |\beta|^2 \sum_{k=0}^N \binom{N}{k} \tau_B^{(N-k)} (1 - \tau_B)^k |0, N-k\rangle \langle 0, N-k|.
\end{aligned} \tag{A.13}$$

The PT moments evaluate to

$$\begin{aligned}
 p_2 = & |\alpha|^4 \sum_{k=0}^N \binom{N}{k}^2 \tau_A^{2(N-k)} (1-\tau_A)^{2k} \\
 & + 2|\alpha|^2 |\beta|^2 (\tau_A^N \tau_B^N + (1-\tau_A)^N (1-\tau_B)^N) \\
 & + |\beta|^4 \sum_{k=0}^N \binom{N}{k}^2 \tau_B^{2(N-k)} (1-\tau_B)^{2k}
 \end{aligned} \tag{A.14}$$

and

$$\begin{aligned}
 p_3 = & |\alpha|^6 \sum_{k=0}^N \binom{N}{k}^3 \tau_A^{3(N-k)} (1-\tau_A)^{3k} \\
 & + 3|\alpha|^2 |\beta|^2 \left(|\alpha|^2 \tau_A^N \tau_B^N (1-\tau_A)^N + |\beta|^2 \tau_A^N \tau_B^N (1-\tau_B)^N \right. \\
 & \left. + |\alpha|^2 (1-\tau_A)^{2N} (1-\tau_B)^N + |\beta|^2 (1-\tau_A)^N (1-\tau_B)^{2N} \right) \\
 & + |\beta|^6 \sum_{k=0}^N \binom{N}{k}^3 \tau_B^{3(N-k)} (1-\tau_B)^{3k},
 \end{aligned} \tag{A.15}$$

where we considered the copies to be identical for brevity. Please note that the expressions underlying Section 5.5.2 describe the more general case of unequal copies.

A.4.3 Particle-number output distributions

We show the particle-number output distributions $f_2(N_2^A, N_2^B)$ and $f_3(N_2^A, N_2^B, N_3^A, N_3^B)$ of the Fourier interferometers of orders 2 [see Figure 5.1b)] and 3 [see Figure 5.1c)] in Tables A.1 and A.2, respectively. Again, we present expressions for equal copies for brevity. Note that at most two (three) particles can be detected at the outputs for two (three) NOON input states. Note also that for $\tau = 1$, i.e., when the state is pure, contributions with odd total particle number $N_2^A + N_2^B$ vanish since all individual parities must be even to average to $p_2 = 1$. Additionally, we provide graphical visualizations of the output distributions at the typical loss levels considered in the main text, i.e., $\tau = 0.9, 0.75, 0.6$, together with the pure-state case $\tau = 1$ [see Figure A.2].

N_{tot}	N_2^A	N_2^B	f_2
0	0	0	$1 - \frac{\tau}{2}(2 - \tau)$
	1	0	$\tau(1 - \tau) \alpha ^2$
1	0	1	$\tau(1 - \tau) \beta ^2$
	1	1	$\tau^2 \alpha ^2 \beta ^2$
2	2	0	$\frac{\tau^2}{2} \alpha ^4$
	0	2	$\frac{\tau^2}{2} \beta ^4$

Table A.1: Particle-number output distribution f_2 of the circuit in Figure 5.1b) for a lossy NOON state ($N = 1$) over all possible particle numbers N_2^A, N_2^B grouped by total particle number $N_{\text{tot}} = N_2^A + N_2^B$.

A.4.4 Statistical error models

We validate the consistency between the statistical error models discussed in Section 5.5.1.3 and the numerical sampling of the third-order criteria for the noiseless and balanced ($\alpha = 1/\sqrt{2}$) NOON state of order $N = 1$ in Figure A.3. As a complementary check, we plot the standard deviations obtained in Eq. (5.58) relative to the loss τ in Figure A.4. The numerical data aligns well with the theoretical predictions for the linear and quadratic criteria. The optimal criterion matches the linear one when $\tau \lesssim 0.4$ and $\tau \gtrsim 0.6$ (which corresponds to $p_2 > 1/2$), but approaches the quadratic one towards $\tau \approx 1/2$ (for which $p_2 \approx 1/2$). As the linear criterion's error is an upper bound for the optimal criterion's error, it is justified to work with the linear error model for the optimal criterion when $\tau \gtrsim 0.6$, which comprises all cases considered in the main text, in particular, Figures 5.3c) and 5.4.

N_{tot}	N_2^A	N_2^B	N_3^A	N_3^B	f_3
0	0	0	0	0	$1 - \frac{1}{9}\tau[18 + \tau(-15 + 4\tau)]$
1	1	0	0	0	$\frac{1}{3} \alpha ^2(1 - \tau)\tau(3 - 2\tau)$
	0	0	1	0	
	0	1	0	0	$\frac{1}{3} \beta ^2(1 - \tau)\tau(3 - 2\tau)$
	0	0	0	1	
2	1	1	0	0	$\frac{4}{3} \alpha ^2 \beta ^2(1 - \tau)\tau^2$
	0	0	1	1	
	1	0	0	1	$\frac{1}{3} \alpha ^2 \beta ^2\tau^2$
	0	1	1	0	
	1	0	1	0	$\frac{1}{3} \alpha ^4\tau^2$
	0	1	0	1	$\frac{1}{3} \beta ^4\tau^2$
	2	0	0	0	$\frac{2}{3} \alpha ^4(1 - \tau)\tau^2$
	0	0	2	0	
	0	2	0	0	$\frac{2}{3} \beta ^4(1 - \tau)\tau^2$
	0	0	0	2	
	1	2	0	0	$\frac{2}{3} \alpha ^2 \beta ^4\tau^3$
	0	0	1	2	
3	2	1	0	0	$\frac{2}{3} \alpha ^4 \beta ^2\tau^3$
	0	0	2	1	
	1	0	2	0	0
	2	0	1	0	
	0	1	0	2	
	0	2	0	1	
	0	1	2	0	
	0	2	1	0	
	1	0	0	2	
	2	0	0	1	
	0	1	1	1	
	1	0	1	1	
	1	1	0	1	
	1	1	1	0	
	3	0	0	0	$\frac{2}{9} \alpha ^6\tau^3$
	0	0	3	0	
	0	3	0	0	$\frac{2}{9} \beta ^6\tau^3$
	0	0	0	3	

Table A.2: Particle-number output distribution f_3 of the circuit in Figure 5.1c) for a lossy NOON state ($N = 1$) over all possible particle numbers $N_2^A, N_2^B, N_3^A, N_3^B$ grouped by total particle number $N_{\text{tot}} = N_2^A + N_2^B + N_3^A + N_3^B$.

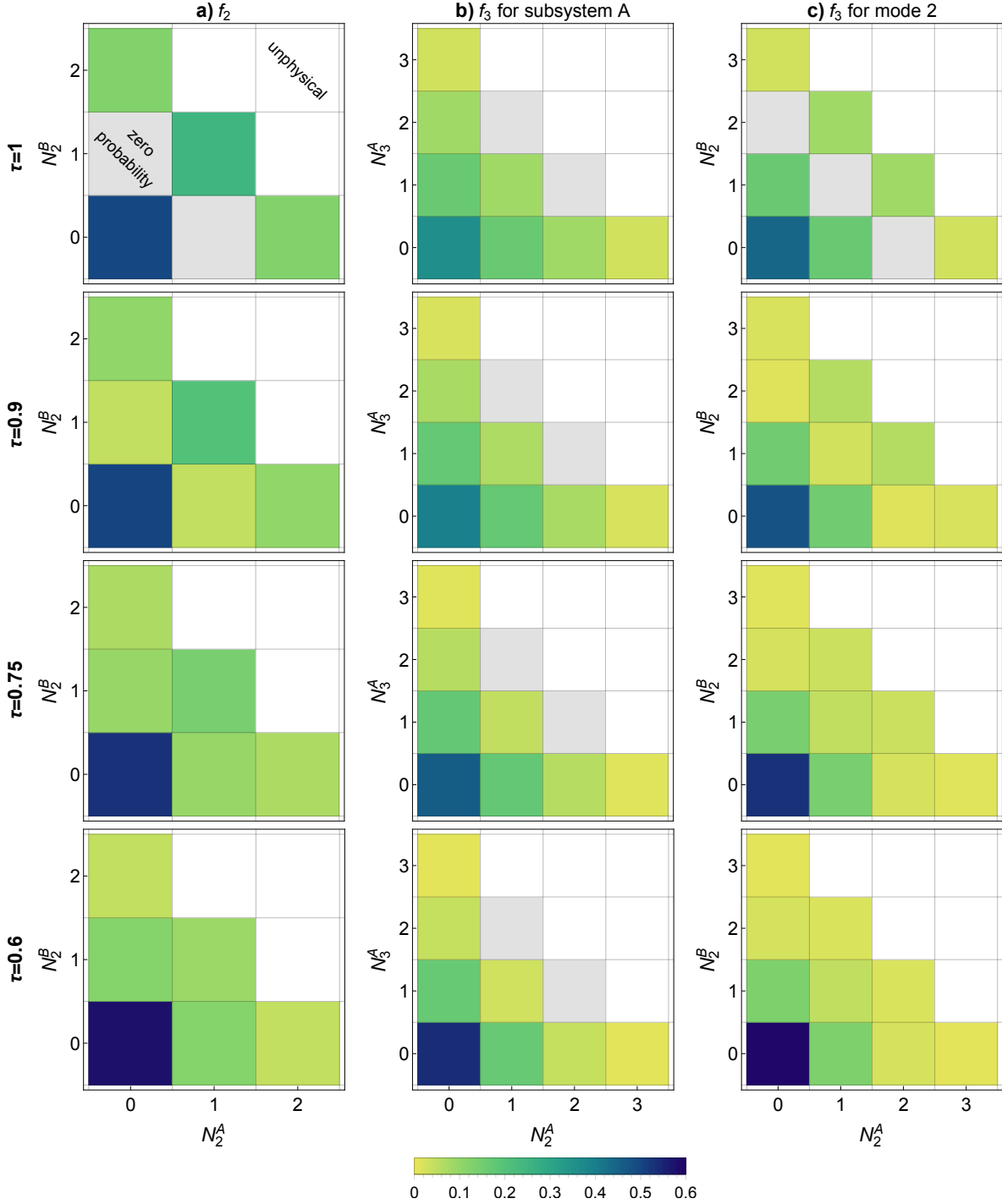


Figure A.2: Particle-number output distributions for the balanced ($\alpha = 1/\sqrt{2}$) NOON state of order $N = 1$ for various losses τ . **a)** shows f_2 from the purity circuit (see Table A.1 for analytic expressions). **b)** and **c)** depict f_3 from the p_3 circuit for subsystem A and in the second mode, respectively (see Table A.2). When the total particle number at the output exceeds its input value, any corresponding configuration is unphysical (white elements). In some cases, physical configurations attain zero probability (gray elements), e.g., when the state is pure ($\tau = 1$) and the total particle number is odd (see the top left panel).

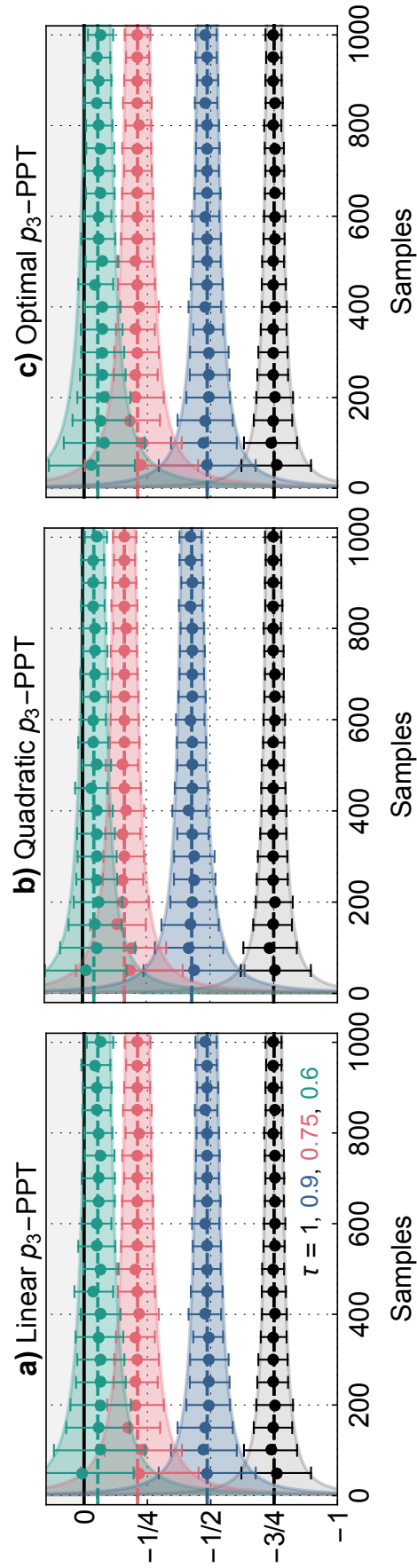


Figure A.3: Numerical validation of the statistical error models for the three p_3 -PPT criteria in the case of the noiseless and balanced ($\alpha = 1/\sqrt{2}$) NOON state with mode population $N = 1$, considered for the three typical losses $\tau = 0.9, 0.75, 0.6$ (blue, red, green), together with the lossless case $\tau = 1$ (black). See Figure 5.4 for reference.

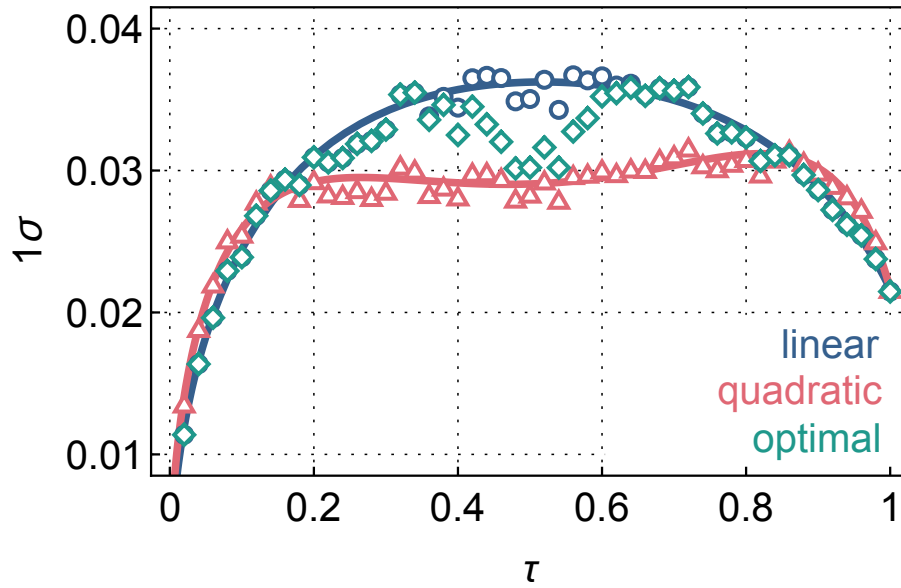


Figure A.4: Complementary analysis to Figure A.3 in terms of the standard deviation as a function of the loss τ . Analytic curves correspond to Eq. (5.58). The criteria are sampled 2×10^3 times, which we repeat 500 times to obtain their standard deviations (points). The optimal criterion's error mostly resembles the linear one's behavior as $p_2 \geq 1/2$ with equality if and only if $\tau = 1/2$.

Appendix to Chapter 6

We give here detailed calculations for Chapter 6, devoted to realignment moment separability criteria for continuous-variable quantum systems.

B.1 Properties of w_n and r_n

Here, we prove several properties that w_n and r_n obey for all $n \in \mathbb{N}_{>0}$. We first prove that

$$r_n := \text{Tr} \left[\left(R(\hat{\rho})^\dagger R(\hat{\rho}) \right)^{n/2} \right] \quad (\text{B.1})$$

is real and non-negative for all $n \in \mathbb{N}_{>0}$. To do so, we write the singular value decomposition of the realigned density matrix as

$$R(\hat{\rho}) = UDV^\dagger, \quad (\text{B.2})$$

where U and V are unitary matrices, and D is diagonal with non-negative entries. Then,

$$r_n = \text{Tr} \left[\left(VD^2V^\dagger \right)^{n/2} \right] \quad (\text{B.3})$$

$$= \text{Tr} \left[\left(VDV^\dagger \right)^n \right] \quad (\text{B.4})$$

$$= \text{Tr}[D^n], \quad (\text{B.5})$$

due to the unitarity of V . Thus, r_n is simply the sum of the n th powers of the singular values of $R(\hat{\rho})$. In particular, $r_n \geq 0$.

We now prove that $w_n \in \mathbb{R}$ for every $n \in \mathbb{N}_{>0}$. Let $\hat{S}$ denote the swap operator acting as follows on basis elements,

$$\hat{S}|i, j\rangle = |j, i\rangle. \quad (\text{B.6})$$

Using the realignment convention of Eq. (6.4),

$$\langle i, k | R(\hat{\rho}) | j, l \rangle = \langle i, j | \hat{\rho} | k, l \rangle, \quad (\text{B.7})$$

together with the Hermiticity of $\hat{\rho}$, we obtain

$$\begin{aligned} \langle i, k | R(\hat{\rho}) | j, l \rangle^* &= \langle i, j | \hat{\rho} | k, l \rangle^* \\ &= \langle k, l | \hat{\rho} | i, j \rangle = \langle k, i | R(\hat{\rho}) | l, j \rangle. \end{aligned} \quad (\text{B.8})$$

Therefore,

$$R(\hat{\rho})^* = \hat{S}R(\hat{\rho})\hat{S}. \quad (\text{B.9})$$

It follows that

$$\begin{aligned} w_n^* &= \text{Tr}[(R(\hat{\rho})^*)^n] \\ &= \text{Tr}[(\hat{S}R(\hat{\rho})\hat{S})^n] \\ &= \text{Tr}[\hat{S}R(\hat{\rho})^n\hat{S}] = w_n. \end{aligned} \quad (\text{B.10})$$

Hence, w_n is real. Moreover, as shown in Appendix B.6.1, its multicopy observable is positive semidefinite, which further implies $w_n \geq 0$.

We next show that

$$|w_n| \leq r_n, \quad (\text{B.11})$$

for all $n \in \mathbb{N}_{>0}$. Let y_i denote the eigenvalues of $R(\hat{\rho})$, and let s_i denote its singular values. Then,

$$w_n = \text{Tr}[R(\hat{\rho})^n] = \sum_i y_i^n, \quad (\text{B.12})$$

whereas

$$r_n = \sum_i s_i^n. \quad (\text{B.13})$$

Using the triangle inequality and the standard (weak) majorization relation between eigenvalues and singular values [161], we obtain

$$|w_n| \leq \sum_i |y_i|^n \leq \sum_i s_i^n = r_n. \quad (\text{B.14})$$

Finally, we prove that $r_n \geq r_{n+1}$ for all $n \in \mathbb{N}_{>0}$. Since r_2 corresponds to the purity of the state $\hat{\rho}$, we have

$$r_2 := \text{Tr}[R(\hat{\rho})^\dagger R(\hat{\rho})] = \text{Tr}[\hat{\rho}^2] \leq 1. \quad (\text{B.15})$$

It follows that each singular value s_i of $R(\hat{\rho})$ satisfies $0 \leq s_i \leq 1$. Therefore,

$$s_i^n \geq s_i^{n+1} \quad (\text{B.16})$$

for all $n \in \mathbb{N}_{>0}$, and summing over i gives

$$r_n \geq r_{n+1}. \quad (\text{B.17})$$

B.2 Properties of the operator Schmidt decomposition

Let $\{\hat{A}_i\}$ and $\{\hat{B}_i\}$ be orthonormal sets of operators with respect to the Hilbert–Schmidt inner product, i.e.,

$$\text{Tr}[\hat{A}_i^\dagger \hat{A}_j] = \text{Tr}[\hat{B}_i^\dagger \hat{B}_j] = \delta_{ij}. \quad (\text{B.18})$$

We assume that the local operator spaces are identified through a fixed basis, so that the operators $\hat{A}_i$ and $\hat{B}_i^*$ can be compared as elements of the same vector space.

We denote by $\{\hat{B}_j^*\}$ the set of entrywise complex conjugates of the operators $\{\hat{B}_j\}$ in the fixed basis. Since $\{\hat{B}_j^*\}$ is also an orthonormal set, each operator $\hat{A}_i$ can be decomposed into its projection onto the span of $\{\hat{B}_j^*\}$ and an orthogonal remainder, that is,

$$\hat{A}_i = \sum_j m_j^{(i)} \hat{B}_j^* + \hat{C}_i, \quad (\text{B.19})$$

where $m_j^{(i)} \in \mathbb{C}$ and $\hat{C}_i$ is orthogonal to the span of $\{\hat{B}_j^*\}$, namely,

$$\text{Tr}[(\hat{B}_j^*)^\dagger \hat{C}_i] = \text{Tr}[\hat{B}_j^T \hat{C}_i] = 0 \quad \forall i, j. \quad (\text{B.20})$$

Using the orthonormality of the operators in the sets $\{\hat{A}_i\}$ and $\{\hat{B}_i^*\}$, see Eq. (B.18), and Eq. (B.19), we get

$$\sum_k (m_k^{(i)})^* m_k^{(j)} + \text{Tr}[\hat{C}_i^\dagger \hat{C}_j] = \delta_{ij}. \quad (\text{B.21})$$

Introducing the matrix M with entries

$$M_{ij} = m_i^{(j)}, \quad (\text{B.22})$$

and the Gram matrix N_C of the residual operators,

$$(N_C)_{ij} = \text{Tr}[\hat{C}_i^\dagger \hat{C}_j], \quad (\text{B.23})$$

we may write

$$M^\dagger M + N_C = I. \quad (\text{B.24})$$

Since N_C is positive semidefinite, it follows that

$$M^\dagger M \leq I, \quad (\text{B.25})$$

or equivalently,

$$\|M\|_2 \leq 1, \quad (\text{B.26})$$

where $\|\cdot\|_2$ denotes the spectral norm. Thus, we call the matrix M a contraction.

As we deduce from Eq. (B.22), the coefficients appearing in Eq. (B.19) are the Hilbert–Schmidt overlaps between $\hat{A}_j$ and $\hat{B}_i^*$. Hence,

$$M_{ij} = \text{Tr} \left[(\hat{B}_i^*)^\dagger \hat{A}_j \right] = \text{Tr} \left[\hat{B}_i^T \hat{A}_j \right]. \quad (\text{B.27})$$

B.3 Second-order bound

Here, we prove that

$$|M_{ii}M_{jj} - M_{ij}M_{ji}| \leq 1. \quad (\text{B.28})$$

We use the result of Appendix B.2, namely the fact that M is a contraction. Let $M^{[ij]}$ denote the 2×2 submatrix of M obtained by keeping rows i, j and columns i, j , namely,

$$M^{[ij]} = \begin{pmatrix} M_{ii} & M_{ij} \\ M_{ji} & M_{jj} \end{pmatrix}. \quad (\text{B.29})$$

Equivalently, one may write

$$M^{[ij]} = P^{[ij]} M Q^{[ij]}, \quad (\text{B.30})$$

where $P^{[ij]}$ is the $2 \times r$ matrix selecting rows i, j , and $Q^{[ij]}$ is the $r \times 2$ matrix selecting columns i, j . Both are partial isometries, so

$$\|P^{[ij]}\|_2 = \|Q^{[ij]}\|_2 = 1. \quad (\text{B.31})$$

Then, by submultiplicativity of the spectral norm and the fact that M is a contraction,

$$\|M^{[ij]}\|_2 \leq \|P^{[ij]}\|_2 \|M\|_2 \|Q^{[ij]}\|_2 \leq 1. \quad (\text{B.32})$$

Let s_1 and s_2 be the singular values of $M^{[ij]}$. Since

$$\|M^{[ij]}\|_2 = \max\{s_1, s_2\} \leq 1, \quad (\text{B.33})$$

we have $s_1 \leq 1$ and $s_2 \leq 1$. Therefore,

$$|\det M^{[ij]}| = s_1 s_2 \leq 1. \quad (\text{B.34})$$

Finally,

$$\det M^{[ij]} = M_{ii}M_{jj} - M_{ij}M_{ji}. \quad (\text{B.35})$$

Hence,

$$|M_{ii}M_{jj} - M_{ij}M_{ji}| \leq 1. \quad (\text{B.36})$$

In Appendix B.4, we generalize this idea to higher-order determinants, which leads to higher-order criteria.

B.4 Higher-order inequalities

We extend here the argument of Appendix B.3 to higher-order minors of the matrix M . Recall that the realigned operator can be written as (see Eq. (6.6))

$$R(\hat{\rho}) = \sum_i \lambda_i |A_i\rangle\langle B_i^*|, \quad (\text{B.37})$$

where $\lambda_i \geq 0$, and define matrix M by

$$M_{ij} = \text{Tr}[\hat{B}_i^T \hat{A}_j]. \quad (\text{B.38})$$

From Appendix B.2, we have that M is a contraction, i.e.,

$$M^\dagger M \leq I. \quad (\text{B.39})$$

We introduce the moments

$$r_n = \sum_i \lambda_i^n, \quad (\text{B.40})$$

and

$$w_n = \sum_{i_1, \dots, i_n} \lambda_{i_1} \cdots \lambda_{i_n} M_{i_1 i_2} M_{i_2 i_3} \cdots M_{i_n i_1}, \quad (\text{B.41})$$

and refer to Appendix B.6 for the quantum-optical measurements of w_n and r_{2n} .

Let us first consider the third-order case. For three distinct indices i, j, k , define the 3×3 principal submatrix

$$M^{[ijk]} = \begin{pmatrix} M_{ii} & M_{ij} & M_{ik} \\ M_{ji} & M_{jj} & M_{jk} \\ M_{ki} & M_{kj} & M_{kk} \end{pmatrix}. \quad (\text{B.42})$$

Its determinant is

$$\begin{aligned} \det M^{[ijk]} &= M_{ii} M_{jj} M_{kk} + M_{ij} M_{jk} M_{ki} + M_{ik} M_{ji} M_{kj} \\ &\quad - M_{ik} M_{jj} M_{ki} - M_{ij} M_{ji} M_{kk} - M_{ii} M_{jk} M_{kj}. \end{aligned} \quad (\text{B.43})$$

Since M is a contraction, every principal submatrix is also a contraction. Therefore,

$$\|M^{[ijk]}\|_2 \leq 1. \quad (\text{B.44})$$

If s_1, s_2, s_3 are the singular values of $M^{[ijk]}$, then $s_\alpha \leq 1$ for all α , and hence

$$|\det M^{[ijk]}| = s_1 s_2 s_3 \leq 1. \quad (\text{B.45})$$

We now use this bound to derive a third-order inequality. Notice that

$$r_1^3 - 3r_1 r_2 + 2r_3 = \sum_{\substack{i,j,k \\ \text{all distinct}}} \lambda_i \lambda_j \lambda_k. \quad (\text{B.46})$$

Similarly, expanding the weak moments gives

$$w_1^3 - 3w_1 w_2 + 2w_3 = \sum_{\substack{i,j,k \\ \text{all distinct}}} \lambda_i \lambda_j \lambda_k \det M^{[ijk]}. \quad (\text{B.47})$$

Consequently,

$$r_1^3 - 3r_1 r_2 + 2r_3 \pm (w_1^3 - 3w_1 w_2 + 2w_3) = \sum_{\substack{i,j,k \\ \text{all distinct}}} \lambda_i \lambda_j \lambda_k \left(1 \pm \det M^{[ijk]}\right). \quad (\text{B.48})$$

Although the right-hand side of Eq. (B.48) is real after summation, the individual determinants $\det M^{[ijk]}$ need not be real. Taking the real part term by term gives

$$r_1^3 - 3r_1 r_2 + 2r_3 \pm (w_1^3 - 3w_1 w_2 + 2w_3) = \sum_{\substack{i,j,k \\ \text{all distinct}}} \lambda_i \lambda_j \lambda_k \left(1 \pm \operatorname{Re}(\det M^{[ijk]})\right). \quad (\text{B.49})$$

Since

$$\left| \operatorname{Re}(\det M^{[ijk]}) \right| \leq \left| \det M^{[ijk]} \right| \leq 1, \quad (\text{B.50})$$

each term in the last sum is non-negative. Hence, we obtain

$$r_1^3 - 3r_1 r_2 + 2r_3 \pm (w_1^3 - 3w_1 w_2 + 2w_3) \geq 0. \quad (\text{B.51})$$

Equivalently,

$$r_1^3 \geq 3r_1 r_2 - 2r_3 \mp (w_1^3 - 3w_1 w_2 + 2w_3). \quad (\text{B.52})$$

For separable states, the realignment criterion gives

$$r_1 = \|R(\hat{\rho})\|_1 \leq 1. \quad (\text{B.53})$$

Moreover, since $r_1 \geq w_1$ and $r_3 \leq r_2$, the previous inequalities imply the experimentally weaker but directly useful bound

$$3w_1 r_2 - 2r_2 + |w_1^3 - 3w_1 w_2 + 2w_3| \leq 1. \quad (\text{B.54})$$

Therefore, any violation,

$$3w_1r_2 - 2r_2 + |w_1^3 - 3w_1w_2 + 2w_3| > 1. \quad (\text{B.55})$$

certifies entanglement. We can now state the generalization¹ to order ℓ . Let

$$I = \{i_1, \dots, i_\ell\} \quad (\text{B.56})$$

be a set of ℓ distinct indices, and let $M^{[I]}$ denote the $\ell \times \ell$ principal submatrix of M obtained by keeping the rows and columns indexed by I . Since M is a contraction, every such principal submatrix is also a contraction. Hence,

$$\|M^{[I]}\|_2 \leq 1. \quad (\text{B.57})$$

If $s_1, \dots, s_\ell$ are the singular values of $M^{[I]}$, then $s_\alpha \leq 1$ for all α , and therefore

$$|\det M^{[I]}| = \prod_{\alpha=1}^{\ell} s_\alpha \leq 1. \quad (\text{B.58})$$

Since determinants of $\ell \times \ell$ submatrices are associated with products of ℓ distinct Schmidt coefficients,

$$\prod_{i \in I} \lambda_i, \quad (\text{B.59})$$

summing over all subsets I of cardinality ℓ gives the ℓ th elementary symmetric polynomial [158],

$$e_\ell(\lambda) := \sum_{\substack{I \\ |I|=\ell}} \prod_{i \in I} \lambda_i. \quad (\text{B.60})$$

However, the quantities we use in the construction of the criteria are the moments

$$r_k = \sum_i \lambda_i^k. \quad (\text{B.61})$$

These are the power sums of the Schmidt coefficients. Newton's identities allow one to express the elementary symmetric polynomials in terms of these power sums [158]. For this reason, we define

$$\mathcal{P}_\ell(x_1, \dots, x_\ell) := \ell! e_\ell, \quad (\text{B.62})$$

where e_ℓ is expressed in terms of the power sums $x_1, \dots, x_\ell$.

¹The connection with elementary symmetric polynomials was suggested by GPT-5.6 Sol. The subsequent development of the generalization was carried out with assistance from GPT-5.6 Sol and was independently verified in full. See [Declaration on the use of Artificial Intelligence](#) for further information on the use of AI in this thesis.

Equivalently,

$$\mathcal{P}_\ell(x_1, \dots, x_\ell) = \det \begin{pmatrix} x_1 & 1 & 0 & \cdots & 0 \\ x_2 & x_1 & 2 & \cdots & 0 \\ x_3 & x_2 & x_1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ell-1 \\ x_\ell & x_{\ell-1} & x_{\ell-2} & \cdots & x_1 \end{pmatrix}. \quad (\text{B.63})$$

Thus, for the Schmidt coefficients,

$$\mathcal{P}_\ell(r_1, \dots, r_\ell) = \ell! \sum_{\substack{I \\ |I|=\ell}} \prod_{i \in I} \lambda_i. \quad (\text{B.64})$$

For example, at order 2,

$$\mathcal{P}_2(x_1, x_2) = x_1^2 - x_2, \quad (\text{B.65})$$

so that

$$\mathcal{P}_2(r_1, r_2) = r_1^2 - r_2 = \sum_{i \neq j} \lambda_i \lambda_j = 2 \sum_{i < j} \lambda_i \lambda_j. \quad (\text{B.66})$$

At order 3,

$$\mathcal{P}_3(x_1, x_2, x_3) = x_1^3 - 3x_1x_2 + 2x_3, \quad (\text{B.67})$$

and therefore

$$\mathcal{P}_3(r_1, r_2, r_3) = r_1^3 - 3r_1r_2 + 2r_3 = \sum_{\substack{i,j,k \\ \text{all distinct}}} \lambda_i \lambda_j \lambda_k. \quad (\text{B.68})$$

This is precisely the sum over products of three distinct Schmidt coefficients.

The same structure appears for the weak moments. Since

$$w_n := \text{Tr} [R(\hat{\rho})^n], \quad (\text{B.69})$$

the quantities w_n are the power sums of the eigenvalues of $R(\hat{\rho})$. Applying Newton's identities to these moments gives

$$\mathcal{P}_\ell(w_1, \dots, w_\ell) = \ell! \sum_{\substack{I \\ |I|=\ell}} \left(\prod_{i \in I} \lambda_i \det M^{[I]} \right). \quad (\text{B.70})$$

Combining the two expressions, we obtain

$$\mathcal{P}_\ell(r_1, \dots, r_\ell) \pm \mathcal{P}_\ell(w_1, \dots, w_\ell) = \ell! \sum_{\substack{I \\ |I|=\ell}} \left(\prod_{i \in I} \lambda_i \right) (1 \pm \text{Re}(\det M^{[I]})). \quad (\text{B.71})$$

Since $|\det M^{[I]}| \leq 1$ and the moments w_n are real (see Appendix B.1), this reduces to

$$\mathcal{P}_\ell(r_1, \dots, r_\ell) \pm \mathcal{P}_\ell(w_1, \dots, w_\ell) \geq 0. \quad (\text{B.72})$$

Thus, each order ℓ yields a higher-order constraint that follows from the contraction property $M^\dagger M \leq I$. Together with the separability condition $r_1 \leq 1$, these inequalities generate a hierarchy of realignment-type entanglement criteria. To obtain a directly measurable separability criterion at a given order, the odd moments r_{2n+1} must additionally be eliminated or bounded using the experimentally accessible quantities w_n and r_{2n} (see Appendix B.6 for their quantum-optical measurement methods). The third-order criterion derived above provides one explicit example of such a reduction.

B.5 Werner state

We consider the exemplary Werner state of two qubits,

$$\hat{\rho}_W = \lambda |\psi^-\rangle \langle \psi^-| + \frac{1-\lambda}{4} \hat{\mathbb{1}}, \quad \lambda \in [0, 1], \quad (\text{B.73})$$

where $|\psi^-\rangle = 1/\sqrt{2}(|0, 1\rangle - |1, 0\rangle)$. Figure B.1 displays the detected regions using the different criteria.

B.6 General measurement techniques for w_n and r_{2n}

In this appendix, we derive multicopy observables for the moments w_n and the even moments r_{2n} , and describe their interferometric implementation. Recall that

$$w_n = \text{Tr} [R(\hat{\rho})^n], \quad (\text{B.74})$$

whereas

$$r_{2n} = \text{Tr} \left[\left(R(\hat{\rho})^\dagger R(\hat{\rho}) \right)^n \right]. \quad (\text{B.75})$$

The quantity w_n is a homogeneous polynomial of degree n in the matrix elements of $\hat{\rho}$, while r_{2n} is a homogeneous polynomial of degree $2n$. They can therefore be accessed using n and $2n$ copies of the state, respectively [29].

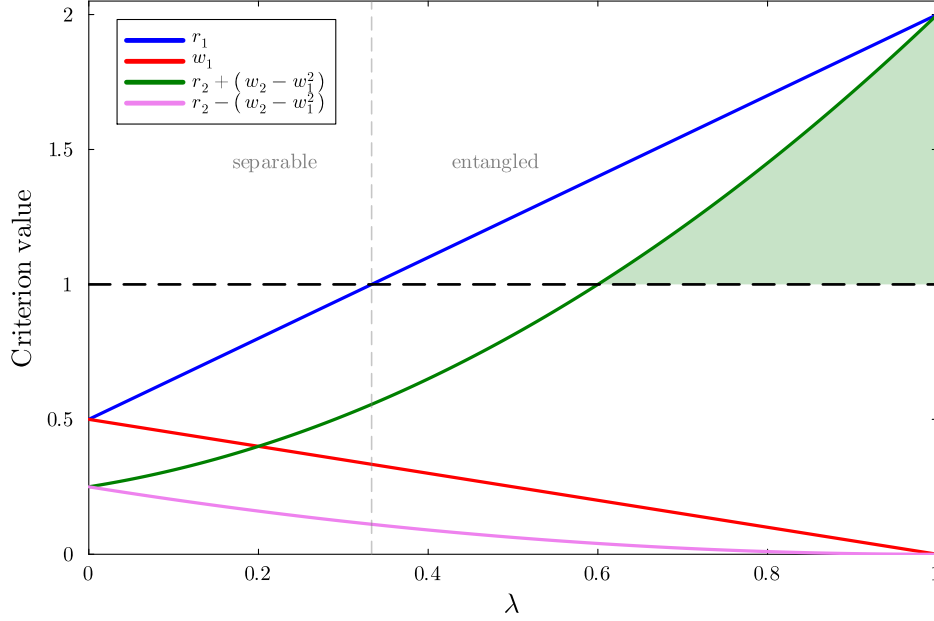


Figure B.1: The Werner state is entangled as soon as $\lambda > 1/3$, which is what detects the original realignment criterion (based on r_1). The weak realignment criterion (based on w_1) never detects Werner state entanglement, whereas the order 2 criterion detects in the region $\lambda > 0.6$, i.e., when the state is sufficiently pure.

More generally, the multicopy method [29] allows one to access homogeneous polynomial functions of the matrix elements of a density operator. Let a general bipartite density operator be written

$$\hat{\rho} = \sum_{ijkl} \rho_{ij,kl} |i, j\rangle \langle k, l|, \quad (\text{B.76})$$

where $\rho_{ij,kl} = \langle i, j | \hat{\rho} | k, l \rangle \in \mathbb{C}$. A homogeneous polynomial of degree n can be written as

$$f_n(\hat{\rho}) = \sum_{\substack{i_1 j_1 k_1 l_1 \\ \vdots \\ i_n j_n k_n l_n}} c_{i_1 j_1 k_1 l_1 \dots i_n j_n k_n l_n} \rho_{i_1 j_1, k_1 l_1} \dots \rho_{i_n j_n, k_n l_n}, \quad (\text{B.77})$$

where the coefficients $c_{i_1 j_1 k_1 l_1 \dots i_n j_n k_n l_n}$ are independent of $\hat{\rho}$. Such a polynomial can be expressed as

$$f_n(\hat{\rho}) = \text{Tr}[\hat{\rho}^{\otimes n} \hat{O}_n] \quad (\text{B.78})$$

for an appropriate multicopy operator $\hat{O}_n$.

For expressing the multicopy operators corresponding to w_n and r_{2n} , we begin by introducing the right cyclic-shift operator acting on n copies,

$$\hat{\Pi}_n |i_1, i_2, \dots, i_n\rangle = |i_n, i_1, \dots, i_{n-1}\rangle, \quad (\text{B.79})$$

and the corresponding left cyclic shift,

$$\hat{\Pi}_n = \left(\hat{\Pi}_n \right)^\dagger, \quad (\text{B.80})$$

which acts as

$$\hat{\Pi}_n |i_1, i_2, \dots, i_n\rangle = |i_2, \dots, i_n, i_1\rangle. \quad (\text{B.81})$$

In what follows, the superscripts A and B indicate whether the shift acts on Alice's or Bob's copies.

B.6.1 Multicopy observable for w_n

Using the density operator writing of Eq. (B.76), the trace of the n th power of the realigned operator reads

$$w_n = \sum_{\substack{i_1, \dots, i_n \\ k_1, \dots, k_n}} \rho_{i_1 i_2, k_1 k_2} \rho_{i_2 i_3, k_2 k_3} \cdots \rho_{i_n i_1, k_n k_1}. \quad (\text{B.82})$$

Then, since

$$\text{Tr} [\hat{\rho} |k, l\rangle \langle i, j|] = \rho_{ij, kl}, \quad (\text{B.83})$$

we obtain

$$w_n = \text{Tr} [\hat{\rho}^{\otimes n} \hat{O}_n^{(w)}], \quad (\text{B.84})$$

where

$$\hat{O}_n^{(w)} = \sum_{\substack{i_1, \dots, i_n \\ k_1, \dots, k_n}} \bigotimes_{t=1}^n |k_t, k_{t+1}\rangle \langle i_t, i_{t+1}|, \quad (\text{B.85})$$

where $i_{n+1} := i_1$ and $k_{n+1} := k_1$. To express this operator in terms of cyclic shifts, we use

$$R(|k_t, i_t\rangle \langle k_{t+1}, i_{t+1}|) = |k_t, k_{t+1}\rangle \langle i_t, i_{t+1}|. \quad (\text{B.86})$$

It follows that

$$\hat{O}_n^{(w)} = R^{\otimes n} \left[\sum_{\substack{i_1, \dots, i_n \\ k_1, \dots, k_n}} \bigotimes_{t=1}^n |k_t, i_t\rangle \langle k_{t+1}, i_{t+1}| \right]. \quad (\text{B.87})$$

After reordering the tensor factors from $(A_1 B_1) \cdots (A_n B_n)$ to $(A_1 \cdots A_n)(B_1 \cdots B_n)$, the operator inside the brackets factorizes as

$$\sum_{\substack{i_1, \dots, i_n \\ k_1, \dots, k_n}} \bigotimes_{t=1}^n |k_t, i_t\rangle \langle k_{t+1}, i_{t+1}| = \hat{\Pi}_n^A \otimes \hat{\Pi}_n^B. \quad (\text{B.88})$$

Consequently,

$$\hat{O}_n^{(w)} = R^{\otimes n} \left(\hat{\Pi}_n^A \otimes \hat{\Pi}_n^B \right). \quad (\text{B.89})$$

In order to be able to write a quantum-optics measurement setting for w_n , it is also useful to express this operator directly in terms of EPR projectors. Using the unnormalized EPR vector

$$|\Omega\rangle := \sum_{q=0}^{\infty} |q, q\rangle. \quad (\text{B.90})$$

Regrouping the factors of $\hat{O}_n^{(w)}$ (see Eq. (B.85)) according to the pairs (A_t, B_{t-1}) , with the cyclic convention $B_0 := B_n$, gives

$$\hat{O}_n^{(w)} = \bigotimes_{t=1}^n |\Omega\rangle \langle \Omega|_{A_t B_{t-1}}. \quad (\text{B.91})$$

Therefore,

$$w_n = \text{Tr} \left[\hat{\rho}^{\otimes n} \bigotimes_{t=1}^n |\Omega\rangle \langle \Omega|_{A_t B_{t-1}} \right]. \quad (\text{B.92})$$

Equation (B.91) also shows directly that $\hat{O}_n^{(w)}$ is Hermitian and positive semidefinite, and hence that w_n is real and non-negative (going beyond the initial proof that w_n is real in Appendix B.1).

It is instructive to compare the expression of the multicopy observable in Eq. (B.89) with the multicopy expression for the moments of the partially transposed state [63, 80],

$$\begin{aligned} p_n &:= \text{Tr} \left[(\hat{\rho}^{T_B})^n \right] \\ &= \text{Tr} \left[\hat{\rho}^{\otimes n} \left(\hat{\Pi}_n^A \otimes \hat{\Pi}_n^B \right) \right], \end{aligned} \quad (\text{B.93})$$

where T_B denotes the transposition on Bob's side. For the partial-transpose moments, the cyclic shifts on Alice's and Bob's copies have opposite orientations. By contrast, the expression for w_n involves shifts in the same direction, followed by an independent realignment map on each copy.

B.6.2 Multicopy observable for r_{2n}

Using the density operator of Eq. (B.76), we may write

$$r_{2n} = \sum_{\substack{i_1, \dots, i_n \\ j_1, \dots, j_n \\ k_1, \dots, k_n \\ l_1, \dots, l_n}} \prod_{t=1}^n \rho_{i_t j_t, k_t l_t}^* \rho_{i_t j_{t+1}, k_t l_{t+1}}, \quad (\text{B.94})$$

where $j_{n+1} := j_1$ and $l_{n+1} := l_1$. Then, using the Hermiticity of $\hat{\rho}$, i.e., $\rho_{i_t j_t, k_t l_t}^* = \rho_{k_t l_t, i_t j_t}$, we obtain

$$r_{2n} = \sum_{\substack{i_1, \dots, i_n \\ j_1, \dots, j_n \\ k_1, \dots, k_n \\ l_1, \dots, l_n}} \prod_{t=1}^n \rho_{k_t l_t, i_t j_t} \rho_{i_t j_{t+1}, k_t l_{t+1}}. \quad (\text{B.95})$$

Each term contains $2n$ matrix elements of $\hat{\rho}$, so the moment can be written as an expectation value on $2n$ copies, that is,

$$r_{2n} = \text{Tr} \left[\hat{\rho}^{\otimes 2n} \hat{O}_{2n}^{(r)} \right], \quad (\text{B.96})$$

where

$$\hat{O}_{2n}^{(r)} = \sum_{\substack{i_1, \dots, i_n \\ j_1, \dots, j_n \\ k_1, \dots, k_n \\ l_1, \dots, l_n}} \bigotimes_{t=1}^n \left((|i_t, j_t\rangle \langle k_t, l_t|)_{2t-1} \otimes (|k_t, l_{t+1}\rangle \langle i_t, j_{t+1}|)_{2t} \right). \quad (\text{B.97})$$

After regrouping the tensor factors from $(A_1 B_1) \cdots (A_{2n} B_{2n})$ to $(A_1 \cdots A_{2n})(B_1 \cdots B_{2n})$, the operator factorizes into independent permutations on Alice's and Bob's subsystems. Indeed, on Alice's side,

$$\sum_{\substack{i_1, \dots, i_n \\ k_1, \dots, k_n}} \bigotimes_{t=1}^n \left((|i_t\rangle \langle k_t|)_{A_{2t-1}} \otimes (|k_t\rangle \langle i_t|)_{A_{2t}} \right) = \bigotimes_{t=1}^n \hat{S}_{A_{2t-1}, A_{2t}}, \quad (\text{B.98})$$

where $\hat{S}_{X,Y}$ denotes the swap operator between subsystems X and Y , i.e.,

$$\hat{S}_{X,Y} |u\rangle_X |v\rangle_Y = |v\rangle_X |u\rangle_Y. \quad (\text{B.99})$$

Similarly, on Bob's side,

$$\sum_{\substack{j_1, \dots, j_n \\ l_1, \dots, l_n}} \bigotimes_{t=1}^n \left((|j_t\rangle \langle l_t|)_{B_{2t-1}} \otimes (|l_{t+1}\rangle \langle j_{t+1}|)_{B_{2t}} \right) = \bigotimes_{t=1}^n \hat{S}_{B_{2t}, B_{2t+1}}, \quad (\text{B.100})$$

with the convention $B_{2n+1} := B_1$. Thus, combining the two sides gives

$$\hat{O}_{2n}^{(r)} = \left(\bigotimes_{t=1}^n \hat{S}_{A_{2t-1}, A_{2t}} \right) \otimes \left(\bigotimes_{t=1}^n \hat{S}_{B_{2t}, B_{2t+1}} \right). \quad (\text{B.101})$$

In summary, Alice's subsystems are paired according to

$$(A_1, A_2), (A_3, A_4), \dots, (A_{2n-1}, A_{2n}), \quad (\text{B.102})$$

whereas Bob's subsystems follow the staggered cyclic pairing

$$(B_2, B_3), (B_4, B_5), \dots, (B_{2n}, B_1). \quad (\text{B.103})$$

The operator $\hat{O}_{2n}^{(r)}$ is Hermitian and unitary because it is a product of disjoint swap operators.

For $n = 1$, Eq. (B.101) reduces to

$$\hat{O}_2^{(r)} = \hat{S}_{A_1, A_2} \otimes \hat{S}_{B_1, B_2}, \quad (\text{B.104})$$

so that

$$r_2 = \text{Tr} [\hat{\rho}^{\otimes 2} (\hat{S}_{A_1, A_2} \otimes \hat{S}_{B_1, B_2})] = \text{Tr} [\hat{\rho}^2], \quad (\text{B.105})$$

as expected.

B.6.3 Interferometric implementation

Although the initial expressions of w_n and r_{2n} involve the realignment map, this map does not need to be directly implemented physically. As a matter of fact, the corresponding multicopy observables can be measured using balanced beam splitters, homodyne detection, and photon-number-parity measurements, as detailed below.

B.6.3.1 Measurement of w_n

Let us consider n copies of the bipartite state, namely,

$$\hat{\rho}_{A_1 B_1} \otimes \dots \otimes \hat{\rho}_{A_n B_n}. \quad (\text{B.106})$$

Equation (B.92) shows that an EPR projection must be performed between Alice's subsystem of copy t and Bob's subsystem of the preceding copy. The modes are therefore paired as

$$(A_1, B_n), (A_2, B_1), \dots, (A_n, B_{n-1}). \quad (\text{B.107})$$

This pairing may be implemented by cyclically rerouting Bob's modes before the interferometers.

For each pair (A_t, B_{t-1}) , with $B_0 := B_n$, we define the commuting EPR quadratures

$$\hat{x}_{-,t} = \frac{\hat{x}_{B_{t-1}} - \hat{x}_{A_t}}{\sqrt{2}}, \quad \hat{p}_{+,t} = \frac{\hat{p}_{A_t} + \hat{p}_{B_{t-1}}}{\sqrt{2}}. \quad (\text{B.108})$$

Since they satisfy

$$[\hat{x}_{-,t}, \hat{p}_{+,t}] = 0, \quad (\text{B.109})$$

their joint measurement can be implemented by a balanced beam splitter followed by homodyne detection on its two output modes. The EPR projector can therefore be written [120]

$$|\Omega\rangle\langle\Omega| = \pi |p_+ = 0, x_- = 0\rangle\langle p_+ = 0, x_- = 0|. \quad (\text{B.110})$$

With the beam-splitter convention adopted in Eq. (6.52), the relevant output quadratures obey

$$\hat{B}_{XY}^\dagger \hat{p}_X \hat{B}_{XY} = \frac{\hat{p}_X + \hat{p}_Y}{\sqrt{2}} = \hat{p}_+, \quad (\text{B.111})$$

and

$$\hat{B}_{XY}^\dagger \hat{x}_Y \hat{B}_{XY} = \frac{\hat{x}_Y - \hat{x}_X}{\sqrt{2}} = \hat{x}_-. \quad (\text{B.112})$$

where $\hat{B}_{XY}$ denotes the balanced beam splitter unitary between modes X and Y . Thus, homodyne detection of $\hat{p}_X$ on the output mode of X and $\hat{x}_Y$ on the output mode of Y provides a simultaneous measurement of $\hat{p}_+$ and $\hat{x}_-$.

The EPR projector can consequently be written formally as

$$|\Omega\rangle\langle\Omega|_{XY} = \pi \hat{B}_{XY}^\dagger [(|p=0\rangle\langle p=0|)_X \otimes (|x=0\rangle\langle x=0|)_Y] \hat{B}_{XY}. \quad (\text{B.113})$$

Combining Eqs. (B.92) and (B.113), we obtain

$$w_n = \pi^n P_{w_n}(0), \quad (\text{B.114})$$

where $P_{w_n}(0)$ is the joint probability density for obtaining

$$x_{-,t} = 0, \quad p_{+,t} = 0, \quad t = 1, \dots, n. \quad (\text{B.115})$$

For example, for low values of n , the interfered pairs on the balanced beam splitters are

n	interfered pairs	
1	(A_1, B_1)	
2	$(A_1, B_2), (A_2, B_1)$	
3	$(A_1, B_3), (A_2, B_1), (A_3, B_2)$.	(B.116)

No genuine n -port interferometer is therefore required. The setup consists only of a cyclic rerouting followed by n independent balanced beam splitters, all of which may operate in parallel, and ends with parallel homodyne measurements.

B.6.3.2 Measurement of r_{2n}

The even moment r_{2n} requires $2n$ copies of the state. According to Eq. (B.101), Alice's copies are interfered in the pairs

$$(A_1, A_2), (A_3, A_4), \dots, (A_{2n-1}, A_{2n}), \quad (\text{B.117})$$

while Bob's copies are interfered according to the staggered cyclic pairing

$$(B_2, B_3), (B_4, B_5), \dots, (B_{2n}, B_1). \quad (\text{B.118})$$

Each swap operator $\hat{S}_{XY}$ can be measured using a balanced beam splitter $\hat{B}_{XY}$ followed by a photon-number-parity measurement on the antisymmetric output mode. With the convention of Eq. (6.52), one has

$$\hat{S}_{XY} = \hat{B}_{XY}^\dagger (-1)^{\hat{n}_Y} \hat{B}_{XY}, \quad (\text{B.119})$$

where $\hat{n}_Y$ denotes the photon-number operator of the antisymmetric (second) output mode.

Consequently,

$$r_{2n} = \left\langle \prod_{t=1}^n (-1)^{\hat{n}_{A,t}^{\text{out}}} \prod_{t=1}^n (-1)^{\hat{n}_{B,t}^{\text{out}}} \right\rangle, \quad (\text{B.120})$$

where $\hat{n}_{A,t}^{\text{out}}$ denotes the photon number in the antisymmetric output of the beam splitter acting on

$$(A_{2t-1}, A_{2t}), \quad (\text{B.121})$$

and $\hat{n}_{B,t}^{\text{out}}$ denotes the photon number in the antisymmetric output of the beam splitter acting on

$$(B_{2t}, B_{2t+1}), \quad B_{2n+1} := B_1. \quad (\text{B.122})$$

Thus, r_{2n} is obtained by multiplying the photon-number parities measured at all antisymmetric output modes and averaging the product over the experimental runs.

For example, the purity r_2 requires two copies and the pairings

$$(A_1, A_2), \quad (B_1, B_2). \quad (\text{B.123})$$

The fourth moment r_4 requires four copies and the pairings

$$(A_1, A_2), \quad (A_3, A_4), \quad (\text{B.124})$$

on Alice's side, and

$$(B_2, B_3), \quad (B_4, B_1), \quad (\text{B.125})$$

on Bob's side.

Appendix to Chapter 9

We give here detailed calculations for Chapter 9, devoted to the study of the majorization lattice from a geometric point of view.

C.1 Majorization minimum of the ball centered on a given vector

We present how to find the majorization minimum of the ball $\mathcal{B}_\varepsilon(x) := \{y \in \mathcal{P}_d^\downarrow : d(x, y) \leq \varepsilon\}$, which we denote as $M_\varepsilon(x)$. It is characterized by $M_\varepsilon(x) < y$ for every $y \in \mathcal{B}_\varepsilon(x)$, and, if $z \in \mathcal{P}_d^\downarrow$ also satisfies $z < y$ for every $y \in \mathcal{B}_\varepsilon(x)$, then $z < M_\varepsilon(x)$. We mainly follow Ref. [131]; see also Ref. [114] for a related construction. The saturation relation $d(x, M_\varepsilon(x)) = \varepsilon$ in the nontrivial regime¹, implying that the infimum is actually a minimum, is discussed in Ref. [114].

Denote by $\nu \in \mathcal{P}_d^\downarrow$ the uniform vector. If $d(\nu, x) \leq \varepsilon$, we are in the trivial regime, and the TV constraint need not be saturated. Therefore, we set $M_\varepsilon(x) = \nu$. Otherwise, if $d(\nu, x) > \varepsilon$, do the following. For given $a, b \in [0, 1]$, we define

$$R(a) = \sum_{i=1}^d (x_i - a)_+, \quad (\text{C.1})$$

$$L(b) = \sum_{i=1}^d (b - x_i)_+. \quad (\text{C.2})$$

Here, $R(a)$ is the mass removed from the components above a , and $L(b)$ is the mass lifted into the components below b . Since $d(\nu, x) > \varepsilon$, there are unique levels a^*, b^* satisfying $R(a^*) = \varepsilon$ and $L(b^*) = \varepsilon$. Moreover, the two clipping regions do not overlap, i.e., $a^* > b^*$. Finally, we define $M_\varepsilon(x)$ using the following clipping formula

$$M_\varepsilon(x)_i = \min\{a^*, \max\{x_i, b^*\}\}. \quad (\text{C.3})$$

Equivalently, entries larger than a^* are lowered to a^* , entries smaller than b^* are raised to b^* , and all other entries are left unchanged. Since x is non-increasingly ordered and the clipping map $t \mapsto \min\{a^*, \max\{t, b^*\}\}$ is monotone non-decreasing, applying it componentwise to a non-increasingly ordered vector preserves the ordering. Hence, $M_\varepsilon(x) \in \mathcal{P}_d^\downarrow$. Figure C.1 illustrates the procedure in the nontrivial regime.

¹We call the regime trivial when the uniform vector belongs to the ball $B_\varepsilon(x)$.

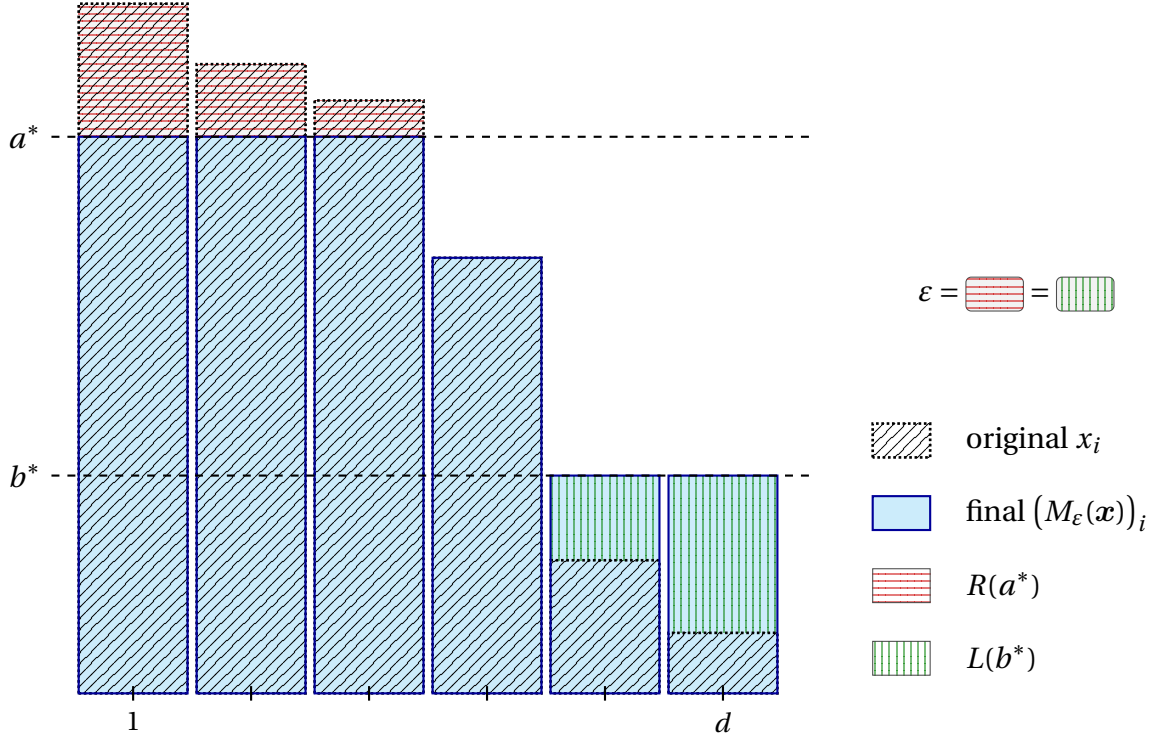


Figure C.1: Schematic depiction of the clipping procedure used to obtain $M_\varepsilon(x)$ from the original vector x . The largest components of x are lowered to the upper level a^* , while the smallest components are raised to the lower level b^* ; intermediate components are left unchanged. The light blue rectangles represent the components of $M_\varepsilon(x)$. The mass removed from the upper components is $R(a^*)$, and the mass lifted into the lower components is $L(b^*)$. In the depicted case, the uniform vector is not contained in $\mathcal{B}_\varepsilon(x)$, and the levels a^* and b^* are chosen so that $R(a^*) = L(b^*) = \varepsilon$. Consequently, $d(x, M_\varepsilon(x)) = \varepsilon$.

C.2 Further tools and results for the proofs

Lemma C.1. *Let $I_1, \dots, I_r$ denote the final blocks produced by the PAVA construction of $u \vee v$ (see Section 7.3.2). If we write $I_\alpha = \{p+1, \dots, q\}$, then, for every $\alpha \in \{1, \dots, r\}$,*

$$\sum_{i=p+1}^q |v_i - \beta(u, v)_i| = \sum_{i=p+1}^q |v_i - (u \vee v)_i|, \quad (\text{C.4})$$

Furthermore, the same holds when replacing v_i with u_i .

Summing Eq. (C.4) over all final PAVA blocks, the above lemma gives

$$d(v, u \vee v) = d(v, \beta(u, v)). \quad (\text{C.5})$$

The analogous identity with u follows in the same way.

Proof of Lemma C.1. Fix a final PAVA block

$$I_\alpha = \{p+1, \dots, q\}. \quad (\text{C.6})$$

On this block, the PAVA procedure replaces the entries $\beta(\mathbf{u}, \mathbf{v})_i$ by their average

$$(\mathbf{u} \vee \mathbf{v})_i = \frac{\sum_{k=p+1}^q \beta(\mathbf{u}, \mathbf{v})_k}{q-p} =: c, \quad i = p+1, \dots, q. \quad (\text{C.7})$$

If $q = p+1$, then $c = \beta(\mathbf{u}, \mathbf{v})_{p+1}$, and the claim is immediate. We therefore assume $q - p \geq 2$.

We claim that, in the final block I_α , either

$$\beta(\mathbf{u}, \mathbf{v})_i \leq v_i \quad \text{and} \quad c \leq v_i \quad (\text{C.8})$$

for all $i = p+1, \dots, q$, or

$$\beta(\mathbf{u}, \mathbf{v})_i \geq v_i \quad \text{and} \quad c \geq v_i \quad (\text{C.9})$$

for all $i = p+1, \dots, q$.

When $q - p \geq 2$, there are two possible cases: Either the affine segment starts on the Lorenz curve of $\mathbf{u}$ and ends on the Lorenz curve of $\mathbf{v}$, or we have the same situation with $\mathbf{u}$ and $\mathbf{v}$ exchanged. Suppose we are in the first case, namely

$$s_p(\beta(\mathbf{u}, \mathbf{v})) = s_p(\mathbf{u}), \quad s_q(\beta(\mathbf{u}, \mathbf{v})) = s_q(\mathbf{v}). \quad (\text{C.10})$$

Since the affine segment of slope c lies above the Lorenz curves of $\mathbf{u}$ and $\mathbf{v}$, we have

$$s_p(\mathbf{u}) + c \geq s_{p+1}(\mathbf{u}). \quad (\text{C.11})$$

Therefore,

$$c \geq u_{p+1}. \quad (\text{C.12})$$

Since $\mathbf{u}$ is non-increasingly ordered, it follows that

$$u_i \leq c, \quad \forall i = p+1, \dots, q. \quad (\text{C.13})$$

Similarly, using the last interior point,

$$s_q(\mathbf{v}) - c \geq s_{q-1}(\mathbf{v}), \quad (\text{C.14})$$

and therefore,

$$c \leq v_q. \quad (\text{C.15})$$

Since v is non-increasingly ordered, this implies

$$c \leq v_i, \quad \forall i = p+1, \dots, q. \quad (\text{C.16})$$

Hence,

$$u_i \leq c \leq v_i, \quad \forall i = p+1, \dots, q. \quad (\text{C.17})$$

In particular, $u_i \leq v_i$ on the block. Thus, the sequence $(s_k(u) - s_k(v))$ is non-increasing on $k = p, \dots, q$. Hence, the curve $s_k(\beta(u, v)) = \max\{s_k(u), s_k(v)\}$ can switch at most once from $s_k(u)$ to $s_k(v)$ on this block. Therefore, for each $i = p+1, \dots, q$, there are only three possibilities. If $s_{i-1}(\beta(u, v)) = s_{i-1}(u)$ and $s_i(\beta(u, v)) = s_i(u)$, then

$$\beta(u, v)_i = u_i \leq v_i. \quad (\text{C.18})$$

If $s_{i-1}(\beta(u, v)) = s_{i-1}(v)$ and $s_i(\beta(u, v)) = s_i(v)$, then

$$\beta(u, v)_i = v_i. \quad (\text{C.19})$$

Finally, if the switch from the curve of u to the curve of v occurs between $i-1$ and i , then

$$\beta(u, v)_i = s_i(v) - s_{i-1}(u) = v_i - (s_{i-1}(u) - s_{i-1}(v)) \leq v_i. \quad (\text{C.20})$$

Thus, in this first case,

$$\beta(u, v)_i \leq v_i \quad \text{and} \quad c \leq v_i \quad (\text{C.21})$$

for all $i = p+1, \dots, q$.

The second case, where the affine segment starts on the Lorenz curve of v and ends on the Lorenz curve of u , is symmetric. It gives

$$\beta(u, v)_i \geq v_i \quad \text{and} \quad c \geq v_i \quad (\text{C.22})$$

for all $i = p+1, \dots, q$.

Consequently, all the terms inside the absolute values on the block have a common sign before and after the replacement of $\beta(u, v)_i$ by c , namely,

$$\sum_{i=p+1}^q |v_i - \beta(u, v)_i| = \left| \sum_{i=p+1}^q (v_i - \beta(u, v)_i) \right|, \quad (\text{C.23})$$

and

$$\sum_{i=p+1}^q |v_i - c| = \left| \sum_{i=p+1}^q (v_i - c) \right|. \quad (\text{C.24})$$

Using the definition of c , we have

$$\sum_{i=p+1}^q \beta(\mathbf{u}, \mathbf{v})_i = \sum_{i=p+1}^q c. \quad (\text{C.25})$$

Therefore,

$$\left| \sum_{i=p+1}^q (v_i - \beta(\mathbf{u}, \mathbf{v})_i) \right| = \left| \sum_{i=p+1}^q (v_i - c) \right|, \quad (\text{C.26})$$

which in turn implies Eq. (C.4). The proof with u_i instead of v_i in Eq. (C.4) is identical, exchanging the roles of $\mathbf{u}$ and $\mathbf{v}$ in the above development.

Note that this proof relies on the fact that we are dealing with sums of absolute values, i.e., with the ℓ_1 -norm. It does not extend in general to ℓ_p -norms with $p > 1$. $\square$

Proposition C.1. Consider $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$ and $\alpha \in [0, 1]$. Then,

$$d(\mathbf{u}, \mathbf{v}) = d(\mathbf{u}, \alpha \mathbf{u} + (1 - \alpha) \mathbf{v}) + d(\alpha \mathbf{u} + (1 - \alpha) \mathbf{v}, \mathbf{v}), \quad (\text{C.27})$$

Proof. A straightforward calculation allows us to conclude.

$$\begin{aligned} d(\mathbf{u}, \alpha \mathbf{u} + (1 - \alpha) \mathbf{v}) + d(\alpha \mathbf{u} + (1 - \alpha) \mathbf{v}, \mathbf{v}) &= \frac{1}{2} \sum_{i=1}^d \left(|u_i - \alpha u_i - (1 - \alpha) v_i| \right. \\ &\quad \left. + |\alpha u_i + (1 - \alpha) v_i - v_i| \right) \\ &= \frac{1}{2} \sum_{i=1}^d \left((1 - \alpha) |u_i - v_i| + \alpha |u_i - v_i| \right) \\ &= \frac{1}{2} \sum_{i=1}^d |u_i - v_i| \\ &= d(\mathbf{u}, \mathbf{v}). \end{aligned} \quad (\text{C.28})$$

$\square$

Lemma C.2. Let $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{P}_d^\downarrow$ be such that

$$d(\mathbf{a}, \mathbf{c}) = d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c}). \quad (\text{C.29})$$

Then, for every $\alpha \in [0, 1]$, if

$$\mathbf{w}_\alpha := \alpha \mathbf{a} + (1 - \alpha) \mathbf{b}, \quad (\text{C.30})$$

we have

$$d(\mathbf{a}, \mathbf{c}) = d(\mathbf{a}, \mathbf{w}_\alpha) + d(\mathbf{w}_\alpha, \mathbf{c}). \quad (\text{C.31})$$

Proof. First, since

$$\mathbf{a} - \mathbf{w}_\alpha = (1 - \alpha)(\mathbf{a} - \mathbf{b}), \quad (\text{C.32})$$

we get

$$d(\mathbf{a}, \mathbf{w}_\alpha) = \frac{1}{2} \sum_i |(1 - \alpha)(a_i - b_i)| = (1 - \alpha)d(\mathbf{a}, \mathbf{b}). \quad (\text{C.33})$$

Secondly, since

$$\mathbf{w}_\alpha - \mathbf{c} = \alpha(\mathbf{a} - \mathbf{c}) + (1 - \alpha)(\mathbf{b} - \mathbf{c}), \quad (\text{C.34})$$

the triangle inequality gives

$$d(\mathbf{w}_\alpha, \mathbf{c}) = \frac{1}{2} \sum_i |\alpha(a_i - c_i) + (1 - \alpha)(b_i - c_i)| \quad (\text{C.35})$$

$$\leq \frac{1}{2} \sum_i (\alpha|a_i - c_i| + (1 - \alpha)|b_i - c_i|) = \alpha d(\mathbf{a}, \mathbf{c}) + (1 - \alpha)d(\mathbf{b}, \mathbf{c}). \quad (\text{C.36})$$

Therefore,

$$\begin{aligned} d(\mathbf{a}, \mathbf{w}_\alpha) + d(\mathbf{w}_\alpha, \mathbf{c}) &\leq (1 - \alpha)d(\mathbf{a}, \mathbf{b}) + \alpha d(\mathbf{a}, \mathbf{c}) + (1 - \alpha)d(\mathbf{b}, \mathbf{c}) \\ &\leq \alpha d(\mathbf{a}, \mathbf{c}) + (1 - \alpha)[d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c})], \end{aligned} \quad (\text{C.37})$$

since

$$(1 - \alpha)d(\mathbf{a}, \mathbf{b}) + \alpha d(\mathbf{a}, \mathbf{c}) + (1 - \alpha)d(\mathbf{b}, \mathbf{c}) = \alpha d(\mathbf{a}, \mathbf{c}) + (1 - \alpha)[d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c})]. \quad (\text{C.38})$$

Using the assumption

$$d(\mathbf{a}, \mathbf{c}) = d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c}), \quad (\text{C.39})$$

we obtain

$$d(\mathbf{a}, \mathbf{w}_\alpha) + d(\mathbf{w}_\alpha, \mathbf{c}) \leq d(\mathbf{a}, \mathbf{c}). \quad (\text{C.40})$$

The reverse inequality follows from the triangle inequality applied to the triple $\mathbf{a}, \mathbf{w}_\alpha, \mathbf{c}$, i.e.,

$$d(\mathbf{a}, \mathbf{c}) \leq d(\mathbf{a}, \mathbf{w}_\alpha) + d(\mathbf{w}_\alpha, \mathbf{c}). \quad (\text{C.41})$$

Hence, equality holds,

$$d(\mathbf{a}, \mathbf{c}) = d(\mathbf{a}, \mathbf{w}_\alpha) + d(\mathbf{w}_\alpha, \mathbf{c}). \quad (\text{C.42})$$

This concludes the proof. $\square$

Lemma C.3. For all $a, b, x, y \geq 0$, if

$$a + b = x + y, \quad (\text{C.43})$$

and

$$a \leq x \leq b, \quad a \leq y \leq b, \quad (\text{C.44})$$

then

$$xy \geq ab. \quad (\text{C.45})$$

Proof. Using $a + b = x + y$, we have

$$xy - ab = (x - a)(y - a). \quad (\text{C.46})$$

Since $x \geq a$ and $y \geq a$, it follows that $xy - ab \geq 0$, which is the desired result. $\square$

Lemma C.4. For any $\mathbf{u}, \mathbf{v} \in \mathcal{P}_d^\downarrow$ such that $\mathbf{u} < \mathbf{v}$, there exists a finite sequence of vectors $\mathbf{u}^{(0)}, \dots, \mathbf{u}^{(N)}$, with

$$\mathbf{u}^{(0)} = \mathbf{u}, \quad \mathbf{u}^{(N)} = \mathbf{v}, \quad (\text{C.47})$$

such that the following is satisfied: At each step $r = 1, \dots, N$, there exist indices $1 \leq i_r < l_r \leq d$ and a positive number δ_r such that

$$\mathbf{u}^{(r)} = \mathbf{u}^{(r-1)} + \delta_r (\mathbf{e}_{i_r} - \mathbf{e}_{l_r}). \quad (\text{C.48})$$

Moreover, $\mathbf{u}^{(r)} \in \mathcal{P}_d^\downarrow$, and

$$\mathbf{u}^{(r-1)} < \mathbf{u}^{(r)} < \mathbf{v}. \quad (\text{C.49})$$

We name these operations *reverse T-transforms* since, rather than obtaining $\mathbf{u}$ from $\mathbf{v}$ using a finite number of T -transforms [161], we obtain $\mathbf{v}$ from $\mathbf{u}$.

Proof of Lemma C.4. Assume that $\mathbf{u}^{(r-1)} < \mathbf{v}$ (this is true for $r = 1$ since $\mathbf{u}^{(0)} = \mathbf{u} < \mathbf{v}$). We define the majorization deficits as

$$D_k := s_k(\mathbf{v}) - s_k(\mathbf{u}^{(r-1)}) \geq 0, \quad k = 0, \dots, d. \quad (\text{C.50})$$

Notice that $D_0 = D_d = 0$. If $D_k = 0$ for all $k = 0, \dots, d$, we stop the procedure, since this means $\mathbf{u}^{(r-1)} = \mathbf{v}$.

Otherwise, choose a maximal interval of strictly positive deficits, i.e., indices $i < l$ such that

$$D_{i-1} = D_l = 0, \quad (\text{C.51})$$

and

$$D_k > 0, \quad \forall k = i, \dots, l-1. \quad (\text{C.52})$$

Let the amount transferred from component l to i be

$$\delta := \min_{i \leq k < l} D_k > 0, \quad (\text{C.53})$$

and define

$$\mathbf{u}^{(r)} = \mathbf{u}^{(r-1)} + \delta(\mathbf{e}_i - \mathbf{e}_l). \quad (\text{C.54})$$

The cumulative sums of $\mathbf{u}^{(r)}$ are increased by δ precisely for $k = i, \dots, l-1$. Therefore, the new deficits are

$$D_k - \delta \geq 0 \quad k = i, \dots, l-1, \quad (\text{C.55})$$

while all other deficits, in particular D_{i-1} and D_l , are unchanged. Moreover, at least one of the deficits in the interval $\{i, \dots, l-1\}$ becomes equal to zero. Hence,

$$\mathbf{u}^{(r-1)} < \mathbf{u}^{(r)} < \mathbf{v}. \quad (\text{C.56})$$

We now check that $\mathbf{u}^{(r)}$ is still non-increasingly ordered. Since only components i and l change, it is enough to check the ordering relations immediately above component i and immediately below component l . If $i > 1$, then

$$\begin{aligned} u_i^{(r)} &\stackrel{(a)}{=} u_i^{(r-1)} + \delta \\ &\stackrel{(b)}{\leq} u_i^{(r-1)} + D_i \\ &\stackrel{(c)}{=} v_i \\ &\stackrel{(d)}{\leq} v_{i-1} \\ &\stackrel{(e)}{\leq} u_{i-1}^{(r-1)} \stackrel{(f)}{=} u_{i-1}^{(r)}, \end{aligned} \quad (\text{C.57})$$

where (a) is by Eq. (C.54), (b) because $\delta \leq D_i$, (c) because $D_{i-1} = 0$, (d) because $\mathbf{v} \in \mathcal{P}_d^{\downarrow}$, (e) because $D_{i-1} - D_{i-2} = v_{i-1} - u_{i-1}^{(r-1)}$, $D_{i-1} = 0$ and $D_k \geq 0$ for all $k = 0, \dots, d$, and (f) because the $(i-1)^{\text{th}}$ component has been untouched during the transfer.

Similarly, if $l < d$, then

$$\begin{aligned}
 u_l^{(r)} &\stackrel{(a)}{=} u_l^{(r-1)} - \delta \\
 &\stackrel{(b)}{\geq} u_l^{(r-1)} - D_{l-1} \\
 &\stackrel{(c)}{=} v_l \\
 &\stackrel{(d)}{\geq} v_{l+1} \\
 &\stackrel{(e)}{\geq} u_{l+1}^{(r-1)} \stackrel{(f)}{=} u_{l+1}^{(r)},
 \end{aligned} \tag{C.58}$$

where (a) is by Eq. (C.54), (b) because $\delta \leq D_{l-1}$, (c) because $D_l = 0$, (d) because $v \in \mathcal{P}_d^\downarrow$, (e) because $D_{l+1} - D_l = v_{l+1} - u_{l+1}^{(r-1)}$, $D_l = 0$ and $D_k \geq 0$ for all $k = 0, \dots, d$, and (f) because the $(l+1)^{\text{th}}$ component has been untouched during the transfer.

Finally, if $l = d$, nonnegativity of the last component follows from

$$u_d^{(r)} = u_d^{(r-1)} - \delta \geq u_d^{(r-1)} - D_{d-1} = v_d \geq 0, \tag{C.59}$$

where we used $D_d = 0$.

Since the other components have been unchanged by the transfer, we get $u^{(r)} \in \mathcal{P}_d^\downarrow$. At each step, at least one strictly positive deficit D_k becomes zero, and no zero deficit becomes strictly positive. Thus, since there are only $d - 1$ intermediate deficits, the procedure terminates after finitely many steps, yielding $u^{(N)} = v$. $\square$

Bibliography

- [1] Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, STOC '11*, pages 333–342, New York, NY, USA, 2011. Association for Computing Machinery.
- [2] Dmitry A. Abanin and Eugene Demler. Measuring Entanglement Entropy of a Generic Many-Body System with a Quantum Switch. *Phys. Rev. Lett.*, 109(2):020504, 2012.
- [3] Itai Afek, Oron Ambar, and Yaron Silberberg. High-NOON States by Mixing Quantum and Classical Light. *Science*, 328(5980):879–881, 2010.
- [4] Girish S. Agarwal. *Quantum Optics*. Cambridge University Press, Cambridge, 2012.
- [5] Mohammad A. Alhejji and Graeme Smith. A Tight Uniform Continuity Bound for Equivocation. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2270–2274, 2020.
- [6] Robert Alicki and Mark Fannes. Continuity of quantum conditional information. *Journal of Physics A: Mathematical and General*, 37(5):L55–L57, 2004.
- [7] Ehud Altman, Kenneth R. Brown, Giuseppe Carleo, Lincoln D. Carr, Eugene Demler, Cheng Chin, Brian DeMarco, Sophia E. Economou, Mark A. Eriksson, Kai-Mei C. Fu, Markus Greiner, Kaden R.A. Hazzard, Randall G. Hulet, Alicia J. Kollár, Benjamin L. Lev, Mikhail D. Lukin, Ruichao Ma, Xiao Mi, Shashank Misra, Christopher Monroe, Kater Murch, Zaira Nazario, Kang-Kuen Ni, Andrew C. Potter, Pedram Roushan, Mark Saffman, Monika Schleier-Smith, Irfan Siddiqi, Raymond Simmonds, Meenakshi Singh, I.B. Spielman, Kristan Temme, David S. Weiss, Jelena Vučković, Vladan Vuletić, Jun Ye, and Martin Zwierlein. Quantum Simulators: Architectures and Opportunities. *PRX Quantum*, 2(1):017003, 2021.
- [8] Carolina Moura Alves, Paweł Horodecki, Daniel K. L. Oi, L. C. Kwek, and Artur K. Ekert. Direct estimation of functionals of density operators by local operations and classical communication. *Phys. Rev. A*, 68(3):032306, 2003.
- [9] Matthieu Arnhem, Célia Griffet, and Nicolas J. Cerf. Multicopy observables for the detection of optically nonclassical states. *Phys. Rev. A*, 106(4):043705, 2022.
- [10] Warit Asavanant, Yu Shiozawa, Shota Yokoyama, Baramée Charoensombutamon, Hiroki Emura, Rafael Alexander, Shuntaro Takeda, Jun-ichi Yoshikawa, Nicolas Menicucci, Hidehiro Yonezawa, and Akira Furusawa. Generation of time-domain-multiplexed two-dimensional cluster state. *Science*, 366:373–376, 2019.

- [11] Alain Aspect, Jean Dalibard, and Gérard Roger. Experimental Test of Bell's Inequalities Using Time-Varying Analyzers. *Phys. Rev. Lett.*, 49(25):1804–1807, 1982.
- [12] Koenraad Audenaert. A sharp continuity estimate for the von Neumann entropy. *Journal of Physics A: Mathematical and Theoretical*, 40(28):8127–8136, 2007.
- [13] Koenraad Audenaert, Bjarne Bergh, Nilanjana Datta, Michael G. Jabbour, Ángela Capel, and Paul Gondolf. Continuity bounds for quantum entropies arising from a fundamental entropic inequality. *IEEE Trans. Inf. Theory*, 71(9):7029–7038, 2025.
- [14] Waseem S. Bakr, Jonathon I. Gillen, Amy Peng, Simon Fölling, and Markus Greiner. A quantum gas microscope for detecting single atoms in a Hubbard-regime optical lattice. *Nature*, 462(7269):74–77, 2009.
- [15] John Bardeen and Walter H. Brattain. The Transistor, A Semi-Conductor Triode. *Phys. Rev.*, 74(2):230–231, 1948.
- [16] Simon Becker, Nilanjana Datta, and Michael G. Jabbour. From Classical to Quantum: Uniform Continuity Bounds on Entropies in Infinite Dimensions. *IEEE Trans. Inf. Theory*, 69(7):4128–4144, 2023.
- [17] Simon Becker, Nilanjana Datta, Ludovico Lami, and Cambyse Rouze. Classical Shadow Tomography for Continuous Variables Quantum Systems. *IEEE Trans. Inf. Theory*, 70(5):3427–3452, 2024.
- [18] John S. Bell. On the Einstein Podolsky Rosen paradox. *Phys. Phys. Fiz.*, 1(3):195–200, 1964.
- [19] Charles H. Bennett, Herbert J. Bernstein, Sandu Popescu, and Benjamin Schumacher. Concentrating partial entanglement by local operations. *Phys. Rev. A*, 53(4):2046–2052, 1996.
- [20] Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing*, pages 175–179, Bangalore, India, 1984. IEEE.
- [21] Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.*, 70(13):1895–1899, 1993.
- [22] Charles H. Bennett and Stephen J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69(20):2881–2884, 1992.

- [23] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.
- [24] Immanuel Bloch, Jean Dalibard, and Sylvain Nascimbène. Quantum simulations with ultracold quantum gases. *Nat. Phys.*, 8(4):267–276, 2012.
- [25] Luca Bombelli, Rabinder K. Koul, Joohan Lee, and Rafael D. Sorkin. Quantum source of entropy for black holes. *Phys. Rev. D*, 34(2):373–383, 1986.
- [26] Gustavo M. Bosyk, Giuseppe Sergioli, Hector Freytes, Federico Holik, and Guido Belomo. Approximate transformations of bipartite pure-state entanglement from the majorization lattice. *Physica A: Statistical Mechanics and its Applications*, 473:403–411, 2017.
- [27] Fernando Brandão, Michał Horodecki, Nelly Huei Ying Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proc. Natl. Acad. Sci.*, 112(11):3275–3279, 2015.
- [28] Samuel L. Braunstein and Peter van Loock. Quantum information with continuous variables. *Rev. Mod. Phys.*, 77(2):513–577, 2005.
- [29] Todd A. Brun. Measuring polynomial functions of states. *Quant. Inf. Comp.*, 4(5):401, 2004.
- [30] Ning Cai, Andreas Winter, and Raymond W. Yeung. Quantum privacy and quantum wiretap channels. *Problems of Information Transmission*, 40(4):318–336, 2004.
- [31] Pasquale Calabrese and John Cardy. Entanglement entropy and quantum field theory. *Journal of Statistical Mechanics: Theory and Experiment*, 2004(06):P06002, 2004.
- [32] Pasquale Calabrese, John Cardy, and Erik Tonni. Entanglement Negativity in Quantum Field Theory. *Phys. Rev. Lett.*, 109(13):130502, 2012.
- [33] Pasquale Calabrese, John Cardy, and Erik Tonni. Entanglement Negativity in Extended Systems: A Field Theoretical Approach. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(02):P02008, 2013.
- [34] Curtis Callan and Frank Wilczek. On geometric entropy. *Phys. Lett. B*, 333(1-2):55–61, 1994.
- [35] Elena Callus, Martin Gärttner, and Tobias Haas. Interferometric detection of continuous-variable entanglement using two states. *Phys. Rev. A*, 112(3):033704, 2025.

- [36] Horacio Casini and Marina Huerta. Entanglement entropy in free quantum field theory. *J. Phys. A Math. Theo.*, 42(50):504007, 2009.
- [37] Nicolas J. Cerf, Chris Adami, and Paul G. Kwiat. Optical simulation of quantum logic. *Phys. Rev. A*, 57(3):R1477–R1480, 1998.
- [38] Nicolas J. Cerf and Tobias Haas. Information and Majorization Theory for Fermionic Phase-Space Distributions. *Phys. Rev. Lett.*, 135(11):110201, 2025.
- [39] Nicolas J. Cerf, Gerd Leuchs, and Eugene S. Polzik. *Quantum Information with Continuous Variables of Atoms and Light*. Imperial College Press, 2007.
- [40] Risheng Cheng, Yiyu Zhou, Sihao Wang, Mohan Shen, Towsif Taher, and Hong X. Tang. A 100-pixel photon-number-resolving detector unveiling photon statistics. *Nat. Photon.*, 17(1):112–119, 2023.
- [41] Eric Chitambar, Runyao Duan, and Yaoyun Shi. Tripartite Entanglement Transformations and Tensor Rank. *Phys. Rev. Lett.*, 101(14):140502, 2008.
- [42] Eric Chitambar and Gilad Gour. Quantum resource theories. *Rev. Mod. Phys.*, 91(2):025001, 2019.
- [43] Eric Chitambar, Debbie Leung, Laura Mančinska, Maris Ozols, and Andreas Winter. Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask). *Communications in Mathematical Physics*, 328:303–326, 2014.
- [44] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285–290, 1975.
- [45] Ferdinando Cicalese, Luisa Gargano, and Ugo Vaccaro. Information theoretic measures of distances and their econometric applications. In *2013 IEEE International Symposium on Information Theory*, pages 409–413, 2013.
- [46] Ferdinando Cicalese and Ugo Vaccaro. Supermodularity and subadditivity properties of the entropy on the majorization lattice. *IEEE Transactions on Information Theory*, 48:933–938, 2002.
- [47] William R. Clements, Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley. Optimal design for universal multiport interferometers. *Optica*, 3(12):1460–1465, 2016.

- [48] Jukka Corander, Ulpu Remes, and Timo Koski. On the Jensen–Shannon divergence and the variation distance for categorical probability distributions. *Kybernetika*, 57(6):879–907, 2021.
- [49] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory, Second Edition*. John Wiley and Sons, 2006.
- [50] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nat. Commun.*, 1(1):149, 2010.
- [51] David R. Curtiss. Recent Extensions of Descartes’ Rule of Signs. *Annals of Mathematics*, 19(4):251–278, 1918.
- [52] Mauro D’Achille, Martin Gärttner, and Tobias Haas. Configurable Photonic Simulator for Quantum Field Dynamics. *PRX Quantum*, 7(2):020370, 2026.
- [53] Andrew J. Daley, Hannes Pichler, Johannes Schachenmayer, and Peter Zoller. Measuring Entanglement Growth in Quench Dynamics of Bosons in an Optical Lattice. *Phys. Rev. Lett.*, 109(2):020505, 2012.
- [54] G. Mauro D’Ariano, Paolo Placido Lo Presti, and Matteo G. A. Paris. Using Entanglement Improves the Precision of Quantum Measurements. *Phys. Rev. Lett.*, 87(27):270404, 2001.
- [55] Nilanjana Datta and Alexander S. Holevo. Complementarity and additivity for depolarizing channels. *arXiv:0510145*, 2005.
- [56] Brian A. Davey and Hilary A. Priestley. *Introduction to Lattices and Order*. Cambridge Mathematical Textbooks. Cambridge University Press, second edition, 2002.
- [57] Carlos de Gois, Kiara Hansenne, and Otfried Gühne. Uncertainty relations from graph theory. *Phys. Rev. A*, 107(6):062211, 2023.
- [58] Jan de Leeuw, Kurt Hornik, and Patrick Mair. Isotone optimization in R: Pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32(5):1–24, 2009.
- [59] A. de Oliveira Junior, Jakub Czartowski, Karol Życzkowski, and Kamil Korzekwa. Geometric structure of thermal cones. *Phys. Rev. E*, 106:064109, 2022.

- [60] Yannick Deller, Martin Gärttner, Tobias Haas, Markus K. Oberthaler, Moritz Reh, and Helmut Strobel. Area laws and thermalization from classical entropies in a Bose-Einstein condensate. *Phys. Rev. A*, 112(1):L011303, 2025.
- [61] Yannick Deller, Martin Gärttner, Tobias Haas, Markus K. Oberthaler, Moritz Reh, and Helmut Strobel. Entropy estimation in a spin-1 Bose–Einstein condensate. *New Journal of Physics*, 27:043004, 2025.
- [62] Serge Deside, Matthieu Arnhem, Célia Griffet, and Nicolas J. Cerf. Probabilistic pure state conversion on the majorization lattice. *Phys. Rev. Res.*, 6(2):023156, 2024.
- [63] Serge Deside, Tobias Haas, and Nicolas J. Cerf. Detecting strongly non-Gaussian entanglement. *Phys. Rev. Res.*, 8(3):033064, 2026.
- [64] Serge Deside, Michael G. Jabbour, and Nicolas J. Cerf. Geometric aspects of majorization: Relating distances and entropies in the majorization lattice. *In advanced preparation*.
- [65] Serge Deside, Vojtěch Kala, and Nicolas J. Cerf. Measurable realignment criteria for entanglement in continuous-variable quantum systems. *In advanced preparation*.
- [66] David Deutsch. Quantum Theory, the Church–Turing Principle and the Universal Quantum Computer. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 400(1818):97–117, 1985.
- [67] David Deutsch and Richard 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.
- [68] Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Trans. Inf. Theory*, 51(1):44–55, 2005.
- [69] Dennis Dieks. Communication by EPR devices. *Physics Letters A*, 92(6):271–272, 1982.
- [70] Ruifang Dong, Mikael Lassen, Joel Heersink, Christoph Marquardt, Radim Filip, Gerd Leuchs, and Ulrik L. Andersen. Experimental entanglement distillation of mesoscopic quantum states. *Nat. Phys.*, 4(12):919–923, 2008.
- [71] Jonathan P. Dowling. Quantum optical metrology - the lowdown on high-N00N states. *Contemp. Phys.*, 49(2):125–143, 2008.
- [72] Lu-Ming Duan, Géza Giedke, J. Ignacio Cirac, and Peter Zoller. Inseparability Criterion for Continuous Variable Systems. *Phys. Rev. Lett.*, 84(12):2722–2725, 2000.

- [73] Benoît Dubus, Tobias Haas, and Nicolas J. Cerf. From bosons and fermions to spins: A multi-mode extension of the Jordan–Schwinger map. *arXiv:2411.04918*, 2024.
- [74] Wolfgang Dür, Guifré Vidal, and J. Ignacio Cirac. Three qubits can be entangled in two inequivalent ways. *Phys. Rev. A*, 62:062314, 2000.
- [75] Albert Einstein. Letter to Max Born, 3 March 1947. In Max Born, editor, *The Born–Einstein Letters: Correspondence between Albert Einstein and Max and Hedwig Born from 1916 to 1955*, pages 157–158. Walker and Company, New York, 1971.
- [76] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Phys. Rev.*, 47(10):777–780, 1935.
- [77] Jens Eisert, Marcus Cramer, and Martin B. Plenio. Colloquium: Area laws for the entanglement entropy. *Rev. Mod. Phys.*, 82(1):277–306, 2010.
- [78] Artur K. Ekert. Quantum cryptography based on Bell’s theorem. *Phys. Rev. Lett.*, 67(6):661–663, 1991.
- [79] Artur K. Ekert, Carolina Moura Alves, Daniel K. L. Oi, Michał Horodecki, Paweł Horodecki, and L. C. Kwek. Direct Estimations of Linear and Nonlinear Functionals of a Quantum State. *Phys. Rev. Lett.*, 88(21):217901, 2002.
- [80] Andreas Elben, Richard Kueng, Hsin-Yuan (Robert) Huang, Rick van Bijnen, Christian Kokail, Marcello Dalmonte, Pasquale Calabrese, Barbara Kraus, John Preskill, Peter Zoller, and Benoît Vermersch. Mixed-State Entanglement from Local Randomized Measurements. *Phys. Rev. Lett.*, 125(20):200501, 2020.
- [81] Matteo Fadel, Tilman Zibold, Boris Décamps, and Philipp Treutlein. Spatial entanglement patterns and Einstein-Podolsky-Rosen steering in Bose-Einstein condensates. *Science*, 360(6387):409–413, 2018.
- [82] Mark Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31(4):291–294, 1973.
- [83] Richard P. Feynman. Simulating Physics with Computers. *International Journal of Theoretical Physics*, 21(6-7):467–488, 1982.
- [84] Stefan Floerchinger, Martin Gärttner, Tobias Haas, and Oliver R. Stockdale. Entropic entanglement criteria in phase space. *Phys. Rev. A*, 105(1):012409, 2022.

- [85] Stefan Floerchinger, Tobias Haas, and Henrik Müller-Groeling. Wehrl entropy, entropic uncertainty relations, and entanglement. *Phys. Rev. A*, 103(6):062222, 2021.
- [86] Peter A. Franken, Alan E. Hill, Charles W. Peters, and Gabriel Weinreich. Generation of Optical Harmonics. *Phys. Rev. Lett.*, 7(4):118–119, 1961.
- [87] Stuart J. Freedman and John F. Clauser. Experimental test of local hidden-variable theories. *Phys. Rev. Lett.*, 28:938–941, 1972.
- [88] Silvio Fuchs, Johann J. Abel, Jan Nathanael, Julius Reinhard, Felix Wiesner, Martin Wünsche, Slawomir Skruszewicz, Christian Rödel, Detlef Born, Heidemarie Schmidt, and Gerhard G. Paulus. Photon counting of extreme ultraviolet high harmonics using a superconducting nanowire single-photon detector. *Appl. Phys. B*, 128(2):26, 2022.
- [89] Srilekha Gandhari, Victor V. Albert, Thomas Gerrits, Jacob M. Taylor, and Michael J. Gullans. Precision Bounds on Continuous-Variable State Tomography Using Classical Shadows. *PRX Quantum*, 5(1):010346, 2024.
- [90] Xiaoting Gao, Mathieu Isoard, Fengxiao Sun, Carlos E. Lopetegui, Yu Xiang, Valentina Parigi, Qiongyi He, and Mattia Walschaers. Correlation-Pattern-Based Continuous Variable Entanglement Detection through Neural Networks. *Phys. Rev. Lett.*, 132(22):220202, 2024.
- [91] Raúl García-Patrón, Carlos Navarrete-Benlloch, Seth Lloyd, Jeffrey H. Shapiro, and Nicolas J. Cerf. Majorization Theory Approach to the Gaussian Channel Minimum Entropy Conjecture. *Phys. Rev. Lett.*, 108(11):110505, 2012.
- [92] Martin Gärttner, Tobias Haas, and Johannes Noll. Detecting continuous-variable entanglement in phase space with the Q distribution. *Phys. Rev. A*, 108(4):042410, 2023.
- [93] Martin Gärttner, Tobias Haas, and Johannes Noll. General Class of Continuous Variable Entanglement Criteria. *Phys. Rev. Lett.*, 131(15):150201, 2023.
- [94] Iulia M. Georgescu, Sahel Ashhab, and Franco Nori. Quantum simulation. *Rev. Mod. Phys.*, 86(1):153–185, 2014.
- [95] Manuel Gessner, Luca Pezzè, and Augusto Smerzi. Efficient entanglement criteria for discrete, continuous, and hybrid variables. *Phys. Rev. A*, 94(2):020101, 2016.
- [96] Sevag Gharibian. Strong NP-Hardness of the Quantum Separability Problem. *Quantum Inf. Comput.*, 10(3):343–360, 2010.

- [97] Vittorio Giovannetti, Saikat Guha, Seth Lloyd, Lorenzo Maccone, and Jeffrey H. Shapiro. Minimum output entropy of bosonic channels: A conjecture. *Phys. Rev. A*, 70(3):032315, 2004.
- [98] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nat. Photonics*, 5(4):222–229, 2011.
- [99] Vittorio Giovannetti, Stefano Mancini, David Vitali, and Paolo Tombesi. Characterizing the entanglement of bipartite quantum systems. *Phys. Rev. A*, 67(2):022320, 2003.
- [100] Marissa Giustina, Marijn A. M. Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K. Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann, and Anton Zeilinger. Significant-loophole-free test of Bell’s theorem with entangled photons. *Phys. Rev. Lett.*, 115(25):250401, 2015.
- [101] Aaron Z. Goldberg, Guillaume S. Thekkadath, and Khabat Heshami. Measuring the quadrature coherence scale on a cloud quantum computer. *Phys. Rev. A*, 107(4):042610, 2023.
- [102] Gilad Gour. Infinite number of conditions for local mixed-state manipulations. *Phys. Rev. A*, 72(2):022323, 2005.
- [103] Johnnie Gray, Leonardo Banchi, Abolfazl Bayat, and Sougato Bose. Machine-Learning-Assisted Many-Body Entanglement Measurement. *Phys. Rev. Lett.*, 121(15):150503, 2018.
- [104] Célia Griffet, Matthieu Arnhem, Stephan De Bièvre, and Nicolas J. Cerf. Interferometric measurement of the quadrature coherence scale using two replicas of a quantum optical state. *Phys. Rev. A*, 108(2):023730, 2023.
- [105] Célia Griffet, Tobias Haas, and Nicolas J. Cerf. Accessing continuous-variable entanglement witnesses with multimode spin observables. *Phys. Rev. A*, 108(2):022421, 2023.
- [106] Christian Gross and Waseem S. Bakr. Quantum gas microscopy for single atom and spin detection. *Nat. Phys.*, 17(12):1316–1323, 2021.
- [107] Christian Gross, Helmut Strobel, Eike Nicklas, Tilman Zibold, Nir Bar-Gill, Gershon Kurizki, and Markus K. Oberthaler. Atomic homodyne detection of continuous-variable entangled twin-atom states. *Nature*, 480(7376):219–223, 2011.

- [108] Otfried Gühne and Géza Tóth. Entanglement detection. *Phys. Rep.*, 474(1):1–75, 2009.
- [109] Antonio Alisson Pessoa Guimarães and Charles Casimiro Cavalcante. Upper bound of ergodic capacity for mimo channels with ricean-fading using majorization theory. *Journal of Communication and Information Systems*, 27(1), 2015.
- [110] Cheng Guo, Eric Chitambar, and Runyao Duan. Common Resource State for Preparing Multipartite Quantum Systems via Local Operations and Classical Communication. *arXiv:1601.06220*, 2016.
- [111] Leonid Gurvits. Classical complexity and quantum entanglement. *Journal of Computer and System Sciences*, 69(3):448–484, 2004.
- [112] Tobias Haas. Area laws from classical entropies. *arXiv:2404.12320*, 2024.
- [113] Eric P. Hanson and Nilanjana Datta. Tight uniform continuity bound for a family of entropies. *arXiv:1707.04249*, 2017.
- [114] Eric P. Hanson and Nilanjana Datta. Maximum and minimum entropy states yielding local continuity bounds. *Journal of Mathematical Physics*, 59(4):042204, 2018.
- [115] Eric P. Hanson and Nilanjana Datta. Entropies, majorization flow, and continuity bounds. In Rupert L. Frank, Ari Laptev, Mathieu Lewin, and Robert Seiringer, editors, *The Physics and Mathematics of Elliott Lieb: The 90th Anniversary Volume I*, pages 473–514. EMS Press, 2022.
- [116] Godfrey H. Hardy, John E. Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, London and New York, second edition, 1952.
- [117] Godfrey H. Hardy, John E. Littlewood, and George Pólya. Some simple inequalities satisfied by convex functions. *Messenger Math.*, 58:145–152, 1929.
- [118] Peter Harremoës. A new look on majorization. In *Proceedings ISITA 2004*, pages 1422–1425, Parma, Italy, 2004.
- [119] Bas Hensen, Hannes Bernien, Anaïs E. Dréau, Andreas Reiserer, Norbert Kalb, Machiel S. Blok, Justus W. Ruitenbergh, Raymond F. L. Vermeulen, Raymond N. Schouten, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Matthew Markham, Daniel J. Twitchen, David Elkouss, Stephanie Wehner, Tim H. Taminiau, and Ronald Hanson. Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526:682–686, 2015.

- [120] Anaëlle Hertz, Matthieu Arnhem, Ali Asadian, and Nicolas J. Cerf. Realignment separability criterion assisted with filtration for detecting continuous-variable entanglement. *Phys. Rev. A*, 104(2):022427, 2021.
- [121] Anaëlle Hertz, Evgueni Karpov, Aikaterini Mandilara, and Nicolas J. Cerf. Detection of non-Gaussian entangled states with an improved continuous-variable separability criterion. *Phys. Rev. A*, 93(3):032330, 2016.
- [122] Anaëlle Hertz, Ognyan Oreshkov, and Nicolas J. Cerf. Multicopy uncertainty observable inducing a symplectic-invariant uncertainty relation in position and momentum phase space. *Phys. Rev. A*, 100(5):052112, 2019.
- [123] Tohya Hiroshima. Majorization Criterion for Distillability of a Bipartite Quantum State. *Phys. Rev. Lett.*, 91(5):057902, 2003.
- [124] Siu-Wai Ho and Sergio Verdú. On the interplay between conditional entropy and error probability. *IEEE Transactions on Information Theory*, 56(12):5930–5942, 2010.
- [125] Alexander S. Holevo. On complementary channels and the additivity problem. *Probability Theory and Applications*, 51:133–143, 2005.
- [126] Chung Ki Hong, Zhe Yu Ou, and Leonard Mandel. Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett.*, 59(18):2044–2046, 1987.
- [127] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-State Entanglement and Distillation: Is there a “Bound” Entanglement in Nature? *Phys. Rev. Lett.*, 80(24):5239–5242, 1998.
- [128] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Limits for entanglement measures. *Phys. Rev. Lett.*, 84(9):2014–2017, 2000.
- [129] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Phys. Lett. A*, 223(1):1–8, 1996.
- [130] Michał Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nat. Commun.*, 4:2059, 2013.
- [131] Michał Horodecki, Jonathan Oppenheim, and Carlo Sparaciari. Extremal distributions under approximate majorization. *Journal of Physics A: Mathematical and Theoretical*, 51(30):305301, 2018.

- [132] Paweł Horodecki, Łukasz Rudnicki, and Karol Życzkowski. Multipartite Entanglement. In Richard Szabo and Martin Bojowald, editors, *Encyclopedia of Mathematical Physics*, volume 2, pages 135–150. Academic Press, Amsterdam, second edition, 2025.
- [133] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81(2):865–942, 2009.
- [134] David B. Hume, Ion Stroescu, Markus Joos, Wolfgang Muesel, Helmut Strobel, and Markus K. Oberthaler. Accurate Atom Counting in Mesoscopic Ensembles. *Phys. Rev. Lett.*, 111(25):253001, 2013.
- [135] Rajibul Islam, Ruichao Ma, Philipp M. Preiss, M. Eric Tai, Alexander Lukin, Matthew Rispoli, and Markus Greiner. Measuring entanglement entropy in a quantum many-body system. *Nature*, 528(7580):77–83, 2015.
- [136] Yonatan Israel, Itai Afek, Shamir Rosen, Oron Ambar, and Yaron Silberberg. Experimental tomography of NOON states with large photon numbers. *Phys. Rev. A*, 85(2):022115, 2012.
- [137] Michael G. Jabbour and Nilanjana Datta. A tight uniform continuity bound for the Arimoto-Rényi conditional entropy and its extension to classical-quantum states. *IEEE Trans. Inf. Theory*, 68(4):2169–2181, 2022.
- [138] Michael G. Jabbour and Nilanjana Datta. Tightening Continuity Bounds for Entropies and Bounds on Quantum Capacities. *IEEE Journal on Selected Areas in Information Theory*, 5:645–658, 2024.
- [139] Ted Jacobson. Entanglement Equilibrium and the Einstein Equation. *Phys. Rev. Lett.*, 116(20):201101, 2016.
- [140] Andrzej Jamiolkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Reports on Mathematical Physics*, 3(4):275–278, 1972.
- [141] Jens G. Jensen and Ruediger Schack. Simple algorithm for local conversion of pure states. *Phys. Rev. A*, 63(6):062303, 2001.
- [142] Daniel Jonathan and Martin B. Plenio. Entanglement-Assisted Local Manipulation of Pure Quantum States. *Phys. Rev. Lett.*, 83(17):3566–3569, 1999.
- [143] Daniel Jonathan and Martin B. Plenio. Minimal conditions for local pure-state entanglement manipulation. *Phys. Rev. Lett.*, 83(7):1455–1458, 1999.

- [144] Richard Jozsa and Noah Linden. On the Role of Entanglement in Quantum-Computational Speed-Up. *Proc. Math. Phys. Eng. Sci.*, 459(2036):2011–2032, 2003.
- [145] Bastian Jungnitsch, Tobias Moroder, and Otfried Gühne. Taming Multiparticle Entanglement. *Phys. Rev. Lett.*, 106(19):190502, 2011.
- [146] Adam M. Kaufman, M. Eric Tai, Alexander Lukin, Matthew Rispoli, Robert Schittko, Philipp M. Preiss, and Markus Greiner. Quantum thermalization through entanglement in an isolated many-body system. *Science*, 353(6301):794–800, 2016.
- [147] Emanuel Knill, Raymond Laflamme, and Gerard J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409(6816):46–52, 2001.
- [148] Pieter Kok, William J. Munro, Kae Nemoto, Timothy C. Ralph, Jonathan P. Dowling, and Gerard J. Milburn. Linear optical quantum computing with photonic qubits. *Rev. Mod. Phys.*, 79(1):135–174, 2007.
- [149] Kamil Korzekwa. Structure of the thermodynamic arrow of time in classical and quantum theories. *Phys. Rev. A*, 95(5):052318, 2017.
- [150] Philipp Kunkel, Maximilian Prüfer, Helmut Strobel, Daniel Linnemann, Anika Frölian, Thomas Gasenzer, Martin Gärttner, and Markus K. Oberthaler. Spatially distributed multipartite entanglement enables EPR steering of atomic clouds. *Science*, 360(6387):413–416, 2018.
- [151] Ludovico Lami, Alessio Serafini, and Gerardo Adesso. Gaussian entanglement revisited. *New J. Phys.*, 20(2):023030, 2018.
- [152] Karsten Lange, Jan Peise, Bernd Lücke, Ilka Kruse, Giuseppe Vitagliano, Iagoba Apellaniz, Matthias Kleinmann, Géza Tóth, and Carsten Klempt. Entanglement between two spatially separated atomic modes. *Science*, 360(6387):416–418, 2018.
- [153] Debbie Leung and Graeme Smith. Continuity of Quantum Channel Capacities. *Communications in Mathematical Physics*, 292(1):201–215, 2009.
- [154] Jianhua Lin. Divergence Measures Based on the Shannon Entropy. *IEEE Trans. Inf. Theory*, 37(1):145–151, 1991.
- [155] Seth Lloyd. Capacity of the noisy quantum channel. *Phys. Rev. A*, 55(3):1613–1622, 1997.
- [156] Seth Lloyd and Samuel L. Braunstein. Quantum Computation over Continuous Variables. *Phys. Rev. Lett.*, 82(8):1784–1787, 1999.

- [157] Max O. Lorenz. Methods of measuring concentration of wealth. *Journal of the American Statistical Association*, 9:209–219, 1905.
- [158] Ian G. Macdonald. *Symmetric Functions and Hall Polynomials*. Oxford University Press, 2 edition, 1995.
- [159] Omar Malik, Xiangyi Meng, Shlomo Havlin, Gyorgy Korniss, Boleslaw K. Szymanski, and Jianxi Gao. Concurrence percolation threshold of large-scale quantum networks. *Commun. Phys.*, 5:193, 2022.
- [160] Stefano Mancini, Vittorio Giovannetti, David Vitali, and Paolo Tombesi. Entangling Macroscopic Oscillators Exploiting Radiation Pressure. *Phys. Rev. Lett.*, 88(12):120401, 2002.
- [161] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer New York, 2011.
- [162] Francesco Marsili, Varun B. Verma, Jeff A. Stern, Scott Harrington, Adriana E. Lita, Thomas Gerrits, Igor Vayshenker, Burm Baek, Matthew D. Shaw, Richard P. Mirin, and Sae Woo Nam. Detecting single infrared photons with 93% system efficiency. *Nat. Photon.*, 7(3):210–214, 2013.
- [163] Adam Miranowicz and Marco Piani. Comment on “Inseparability Criteria for Continuous Bipartite Quantum States”. *Phys. Rev. Lett.*, 97(5):058901, 2006.
- [164] Adam Miranowicz, Marco Piani, Paweł Horodecki, and Ryszard Horodecki. Inseparability criteria based on matrices of moments. *Phys. Rev. A*, 80(5):052303, 2009.
- [165] Morgan W. Mitchell, Jeff S. Lundeen, and Aephraim M. Steinberg. Super-resolving phase measurements with a multiphoton entangled state. *Nature*, 429(6988):161–164, 2004.
- [166] Galan Moody, Volker J. Sorger, Daniel J. Blumenthal, Paul W. Juodawlkis, William Loh, Cheryl Sorace-Agaskar, Alex E. Jones, Krishna C. Balram, Jonathan C. F. Matthews, Anthony Laing, Marcelo Davanco, Lin Chang, John E. Bowers, Niels Quack, Christophe Galland, Igor Aharonovich, Martin A. Wolff, Carsten Schuck, Neil Sinclair, Marko Lončar, Tin Komljenovic, David Weld, Shayan Mookherjea, Sonia Buckley, Marina Radulaski, Stephan Reitzenstein, Benjamin Pingault, Bartholomeus Machielse, Debsuvra Mukhopadhyay, Alexey Akimov, Aleksei Zheltikov, Girish S. Agarwal, Kartik Srinivasan, Juanjuan Lu, Hong X. Tang, Wentao Jiang, Timothy P. McKenna, Amir H. Safavi-Naeini, Stephan Steinhauer, Ali W. Elshaari, Val Zwiller, Paul S. Davids, Nicholas Martinez, Michael Gehl, John Chiaverini, Karan K. Mehta, Jacqueline Romero, Navin B. Lingaraju,

- Andrew M. Weiner, Daniel Peace, Robert Cernansky, Mirko Lobino, Eleni Diamanti, Luis Trigo Vidarte, and Ryan M. Camacho. 2022 Roadmap on integrated quantum photonics. *J. Phys. Photonics*, 4(1):012501, 2022.
- [167] Karl Mosler. Majorization in economic disparity measures. *Linear Algebra and its Applications*, 199(Supplement 1):91–114, 1994.
- [168] Tomohisa Nagata, Ryo Okamoto, Jeremy L. O’Brien, Keiji Sasaki, and Shigeki Takeuchi. Beating the Standard Quantum Limit with Four-Entangled Photons. *Science*, 316(5825):726–729, 2007.
- [169] Antoine Neven, Jose Carrasco, Vittorio Vitale, Christian Kokail, Andreas Elben, Marcello Dalmonte, Pasquale Calabrese, Peter Zoller, Benoît Vermersch, Richard Kueng, and Barbara Kraus. Symmetry-resolved entanglement detection using partial transpose moments. *npj Quantum Inf.*, 7:152, 2021.
- [170] Antoine Neven, David Gunn, Martin Hebenstreit, and Barbara Kraus. Local transformations of multiple multipartite states. *SciPost Phys.*, 11:042, 2021.
- [171] Michael A. Nielsen. Conditions for a Class of Entanglement Transformations. *Phys. Rev. Lett.*, 83(2):436–439, 1999.
- [172] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [173] Michael A. Nielsen and Julia Kempe. Separable States Are More Disordered Globally than Locally. *Phys. Rev. Lett.*, 86(22):5184–5187, 2001.
- [174] Clara I. Osorio Tamayo, Nicolas Bruno Sangouard, and Rob Thew. On the Purity and Indistinguishability of Down-Converted Photons. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 46(5):055501, 2013.
- [175] Matteo G. A. Paris. Quantum estimation for quantum technology. *Int. J. Quantum Inf.*, 07(supp01):125–137, 2009.
- [176] Abdolreza Pasharavesh and Michal Bajcsy. Generation of non-Gaussian states of light using deterministic photon subtraction. *New J. Phys.*, 26:113022, 2024.
- [177] Jan Peise, Ilka Kruse, Karsten Lange, Bernd Lücke, Luca Pezzè, Jan Arlt, Wolfgang Ertmer, Klemens Hammerer, Luis Santos, Augusto Smerzi, and Carsten Klempt. Satisfying the Einstein-Podolsky-Rosen criterion with massive particles. *Nat. Commun.*, 6:8984, 2015.

- [178] Asher Peres. Separability Criterion for Density Matrices. *Phys. Rev. Lett.*, 77(8):1413–1415, 1996.
- [179] Sébastien Perseguers, Maciej Lewenstein, Antonio Acín, and J. Ignacio Cirac. Quantum random networks. *Nat. Phys.*, 6:539–543, 2010.
- [180] Denes Petz. *Quantum information theory and quantum statistics*. Springer, Berlin, Heidelberg, 2008.
- [181] Luca Pezzè, Augusto Smerzi, Markus K. Oberthaler, Roman Schmied, and Philipp Treutlein. Quantum metrology with nonclassical states of atomic ensembles. *Rev. Mod. Phys.*, 90(3):035005, 2018.
- [182] Mark S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, San Francisco, CA and London and Amsterdam, 1964.
- [183] Martin B. Plenio and Shashank Virmani. An Introduction to Entanglement Measures. *Quantum Inf. Comput.*, 7(1):1–51, 2007.
- [184] William N. Plick, Petr M. Anisimov, Jonathan P. Dowling, Hwang Lee, and Girish S. Agarwal. Parity Detection in Quantum Optical Metrology Without Number Resolving Detectors. *New Journal of Physics*, 12:113025, 2010.
- [185] Zbigniew Puchała, Łukasz Rudnicki, and Karol Życzkowski. Majorization entropic uncertainty relations. *Journal of Physics A: Mathematical and Theoretical*, 46(27):272002, 2013.
- [186] Zhongzhong Qin, Manuel Gessner, Zhihong Ren, Xiaowei Deng, Dongmei Han, Weidong Li, Xiaolong Su, Augusto Smerzi, and Kunchi Peng. Characterizing the multipartite continuous-variable entanglement structure from squeezing coefficients and the Fisher information. *npj Quantum Inf.*, 5(1):3, 2019.
- [187] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, 73(1):58–61, 1994.
- [188] Dileep V. Reddy, Robert R. Nerem, Sae Woo Nam, Richard P. Mirin, and Varun B. Verma. Superconducting nanowire single-photon detectors with 98% system detection efficiency at 1550nm. *Optica*, 7(12):1649–1653, 2020.
- [189] Kevin J. Resch, Kenneth L. Pregnell, Robert Prevedel, Alexei Gilchrist, Geoffrey J. Pryde, Jeremy L. O’Brien, and Andrew G. White. Time-Reversal and Super-Resolving Phase Measurements. *Phys. Rev. Lett.*, 98(22):223601, 2007.

- [190] Carles Rodó, Gerardo Adesso, and Anna Sanpera. Operational Quantification of Continuous-Variable Correlations. *Phys. Rev. Lett.*, 100(11):110505, 2008.
- [191] Jacqueline Romero and Gerard J. Milburn. Photonic Quantum Computing. In Brian Foster, editor, *Oxford Research Encyclopedia of Physics*. Oxford University Press, New York, 2025.
- [192] Oliver Rudolph. A separability criterion for density operators. *Journal of Physics A: Mathematical and General*, 33(21):3951, 2000.
- [193] Oliver Rudolph. Some properties of the computable cross-norm criterion for separability. *Phys. Rev. A*, 67(3):032312, 2003.
- [194] Oliver Rudolph. Further Results on the Cross Norm Criterion for Separability. *Quantum Inf. Process.*, 4(3):219–239, 2005.
- [195] Alessandro Saboia, Fabricio Toscano, and Stephen P. Walborn. Family of continuous-variable entanglement criteria using general entropy functions. *Phys. Rev. A*, 83(3):032307, 2011.
- [196] Barry C. Sanders. Quantum dynamics of the nonlinear rotator and the effects of continual spin measurement. *Phys. Rev. A*, 40(5):2417–2427, 1989.
- [197] Igal Sason. Entropy Bounds for Discrete Random Variables via Maximal Coupling. *IEEE Trans. Inf. Theory*, 59(11):7118–7131, 2013.
- [198] Igal Sason. Tight Bounds on the Rényi Entropy via Majorization with Applications to Guessing and Compression. *Entropy*, 20(12):896, 2018.
- [199] David Sauerwein, Katharina Schwaiger, and Barbara Kraus. Discrete And Differentiable Entanglement Transformations. *arXiv:1808.02819*, 2018.
- [200] James Schneeloch and Gregory A. Howland. Quantifying high-dimensional entanglement with Einstein-Podolsky-Rosen correlations. *Phys. Rev. A*, 97(4):042338, 2018.
- [201] James Schneeloch, Christopher C Tison, Michael L Fanto, Paul M Alsing, and Gregory A Howland. Quantifying entanglement in a 68-billion-dimensional quantum state space. *Nat. Commun.*, 10(2785):1–7, 2019.
- [202] Benjamin Schumacher. Quantum coding. *Phys. Rev. A*, 51(4):2738–2747, 1995.

- [203] Issai Schur. Über eine klasse von mittelbildungen mit anwendungen auf die determinanten-theorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22:9–20, 1923.
- [204] Alessio Serafini. *Quantum Continuous Variables*. CRC Press, 2017.
- [205] Alessio Serafini, Gerardo Adesso, and Fabrizio Illuminati. Unitarily localizable entanglement of gaussian states. *Phys. Rev. A*, 71(3):032349, 2005.
- [206] Kaushik P. Seshadreesan, Ludovico Lami, and Mark M. Wilde. Rényi relative entropies of quantum Gaussian states. *J. Math. Phys.*, 59(7):072204, 2018.
- [207] Lynden K. Shalm, Evan Meyer-Scott, Bradley G. Christensen, Peter Bierhorst, Michael A. Wayne, Martin J. Stevens, Thomas Gerrits, Scott Glancy, Deny R. Hamel, Michael S. Allman, Kevin J. Coakley, Shellee D. Dyer, Carson Hodge, Adriana E. Lita, Varun B. Verma, Camilla LAMBROCCO, Edward Tortorici, Alan L. Migdall, Yanbao Zhang, Daniel R. Kumor, William H. Farr, Francesco Marsili, Matthew D. Shaw, Jeffrey A. Stern, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Thomas Jennewein, Morgan W. Mitchell, Paul G. Kwiat, Joshua C. Bienfang, Richard P. Mirin, Emanuel Knill, and Sae Woo Nam. Strong loophole-free test of local realism. *Phys. Rev. Lett.*, 115(25):250402, 2015.
- [208] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.
- [209] Evgeny V. Shchukin and Wolfgang Vogel. Inseparability Criteria for Continuous Bipartite Quantum States. *Phys. Rev. Lett.*, 95(23):230502, 2005.
- [210] Evgeny V. Shchukin and Wolfgang Vogel. Nonclassical moments and their measurement. *Phys. Rev. A*, 72(4):043808, 2005.
- [211] Jacob F. Sherson, Christof Weitenberg, Manuel Endres, Marc Cheneau, Immanuel Bloch, and Stefan Kuhr. Single-atom-resolved fluorescence imaging of an atomic Mott insulator. *Nature*, 467(7311):68–72, 2010.
- [212] Maksim E. Shirokov. Tight uniform continuity bounds for the quantum conditional mutual information, for the Holevo quantity, and for capacities of quantum channels. *Journal of Mathematical Physics*, 58(10):102202, 2017.
- [213] Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, pages 124–134. IEEE, 1994.

- [214] Peter W. Shor. The Quantum Channel Capacity and Coherent Information. Lecture notes, MSRI Workshop on Quantum Computation, San Francisco, 2002.
- [215] Barry Simon. *Trace Ideals and Their Applications*. Mathematical Surveys and Monographs. American Mathematical Society, 2005.
- [216] Rajiah Simon. Peres-Horodecki Separability Criterion for Continuous Variable Systems. *Phys. Rev. Lett.*, 84(12):2726–2729, 2000.
- [217] Uttam Singh, Siddhartha Das, and Nicolas J. Cerf. Partial order on passive states and Hoffman majorization in quantum thermodynamics. *Phys. Rev. Res.*, 3(3):033091, 2021.
- [218] Sergei Slussarenko and Geoff J. Pryde. Photonic quantum information processing: A concise review. *App. Phys. Rev.*, 6(4):041303, 2019.
- [219] Sergei Slussarenko, Morgan M. Weston, Helen M. Chrzanowski, Lynden K. Shalm, Varun B. Verma, Sae Woo Nam, and Geoff J. Pryde. Unconditional violation of the shot-noise limit in photonic quantum metrology. *Nat. Photon.*, 11(11):700–703, 2017.
- [220] Sergey N. Solodukhin. Entanglement Entropy of Black Holes. *Living Rev. Relativ.*, 14(1):8, 2011.
- [221] Mark Srednicki. Entropy and area. *Phys. Rev. Lett.*, 71(5):666–669, 1993.
- [222] Philipp Stammer, Javier Rivera-Dean, Theodoros Lampropoulos, Emilio Pisanty, Marcelo F. Ciappina, Paraskevas Tzallas, and Maciej Lewenstein. High Photon Number Entangled States and Coherent State Superposition from the Extreme Ultraviolet to the Far Infrared. *Phys. Rev. Lett.*, 128(12):123603, 2022.
- [223] Alexander Stévin, Michael G. Jabbour, Serge Deside, and Nicolas J. Cerf. Majorization precursors to supermodularity and subadditivity on the majorization lattice. *arXiv:2605.30331*, 2026.
- [224] Moritz Straeter, Michael Tsesmelis, and Leong-Chuan Kwek. Detecting entanglement of non-Gaussian continuous-variable states from single-copy homodyne measurements. *arXiv:2606.28698*, 2026.
- [225] Helmut Strobel, Wolfgang Muessel, Daniel Linnemann, Tilman Zibold, David B. Hume, Luca Pezzè, Augusto Smerzi, and Markus K. Oberthaler. Fisher information and entanglement of non-Gaussian spin states. *Science*, 345(6195):424–427, 2014.

- [226] Daiqin Su, Casey R. Myers, and Krishna Kumar Sabapathy. Conversion of Gaussian states to non-Gaussian states using photon-number-resolving detectors. *Phys. Rev. A*, 100(5):052301, 2019.
- [227] David Sutter, Volkher B. Scholz, Andreas Winter, and Renato Renner. Approximate Degradable Quantum Channels. *IEEE Trans. Inf. Theory*, 63(12):7832–7844, 2017.
- [228] Daniel S. Tasca, Łukasz Rudnicki, Rafael M. Gomes, Fabricio Toscano, and Stephen P. Walborn. Reliable Entanglement Detection under Coarse-Grained Measurements. *Phys. Rev. Lett.*, 110(21):210502, 2013.
- [229] Mireia Tolosa-Simeón, Álvaro Parra-López, Natalia Sánchez-Kuntz, Tobias Haas, Celia Viermann, Marius Sparn, Nikolas Liebster, Maurus Hans, Elinor Kath, Helmut Strobel, Markus K. Oberthaler, and Stefan Floerchinger. Curved and expanding spacetime geometries in Bose-Einstein condensates. *Phys. Rev. A*, 106(3):033313, 2022.
- [230] Gokhan Torun and Ali Yildiz. Deterministic transformations of bipartite pure states. *Phys. Lett. A*, 379(3):113–118, 2015.
- [231] Zacharie Van Herstraeten. *Majorization theoretical approach to quantum uncertainty: From Wigner entropy to Gaussian bosonic channels*. PhD thesis, Université libre de Bruxelles, 2021.
- [232] Vlatko Vedral, Martin B. Plenio, Michael A. Rippin, and Peter L. Knight. Quantifying Entanglement. *Phys. Rev. Lett.*, 78(12):2275–2279, 1997.
- [233] Guifré Vidal. Entanglement of Pure States for a Single Copy. *Phys. Rev. Lett.*, 83:1046–1049, 1999.
- [234] Guifré Vidal. Entanglement monotones. *Journal of Modern Optics*, 47(2-3):355–376, 2000.
- [235] Guifré Vidal, Daniel Jonathan, and Michael A. Nielsen. Approximate transformations and robust manipulation of bipartite pure-state entanglement. *Phys. Rev. A*, 62(1):012304, 2000.
- [236] Celia Viermann, Marius Sparn, Nikolas Liebster, Maurus Hans, Elinor Kath, Álvaro Parra-López, Mireia Tolosa-Simeón, Natalia Sánchez-Kuntz, Tobias Haas, Helmut Strobel, Stefan Floerchinger, and Markus K. Oberthaler. Quantum field simulator for dynamics in curved spacetime. *Nature*, 611(7935):260–264, 2022.

- [237] Klaus Vogel and Hannes Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A*, 40(5):2847(R)–2849(R), 1989.
- [238] Stephen P. Walborn, Alejo Salles, Rafael M. Gomes, Fabricio Toscano, and Paulo H. Souto Ribeiro. Revealing Hidden Einstein-Podolsky-Rosen Nonlocality. *Phys. Rev. Lett.*, 106(13):130402, 2011.
- [239] Stephen P. Walborn, Bruno G. Taketani, Alejo Salles, Fabricio Toscano, and Ruynet L. de Matos Filho. Entropic Entanglement Criteria for Continuous Variables. *Phys. Rev. Lett.*, 103(16):160505, 2009.
- [240] Daniel F. Walls and Gerard J. Milburn. *Quantum Optics*. Springer-Verlag, Berlin, Heidelberg, second edition, 2008.
- [241] Mattia Walschaers. Non-Gaussian Quantum States and Where to Find Them. *PRX Quantum*, 2(3):030204, 2021.
- [242] Philip Walther, Jian-Wei Pan, Markus Aspelmeyer, Rupert Ursin, Sara Gasparoni, and Anton Zeilinger. De Broglie wavelength of a non-local four-photon state. *Nature*, 429(6988):158–161, 2004.
- [243] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, and Seth Lloyd. Gaussian quantum information. *Rev. Mod. Phys.*, 84(2):621, 2012.
- [244] Reinhard F. Werner and Michael M. Wolf. Bound Entangled Gaussian States. *Phys. Rev. Lett.*, 86(16):3658–3661, 2001.
- [245] Mark M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.
- [246] John Williamson. On the Algebraic Problem Concerning the Normal Forms of Linear Dynamical Systems. *Am. J. Math.*, 58(1):141–163, 1936.
- [247] Andreas Winter. Tight Uniform Continuity Bounds for Quantum Entropies: Conditional Entropy, Relative Entropy Distance and Energy Constraints. *Communications in Mathematical Physics*, 347(1):291–313, 2016.
- [248] William K. Wootters and Wojciech H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.
- [249] Anuj K. Yadav and Yanina Y. Shkel. Geometry of Rényi entropy on the majorization lattice. *arXiv:2605.09655*, 2026.

- [250] Ma-Cheng Yang and Cong-Feng Qiao. Witness high-dimensional quantum steering via majorization lattice. *npj Quantum Inf.*, 12:55, 2026.
- [251] Xiao-Dong Yu and Otfried Gühne. Detecting coherence via spectrum estimation. *Phys. Rev. A*, 99(6):062310, 2019.
- [252] Xiao-Dong Yu, Satoya Imai, and Otfried Gühne. Optimal entanglement certification from moments of the partial transpose. *Phys. Rev. Lett.*, 127(6):060504, 2021.
- [253] Bernard Yurke and David Stoler. Generating quantum mechanical superpositions of macroscopically distinguishable states via amplitude dispersion. *Phys. Rev. Lett.*, 57(1):13–16, 1986.
- [254] Chengjie Zhang, Sophia Denker, Ali Asadian, and Otfried Gühne. Analyzing quantum entanglement with the schmidt decomposition in operator space. *Phys. Rev. Lett.*, 133(4):040203, 2024.
- [255] Marek Żukowski, Anton Zeilinger, Michael A. Horne, and Artur K. Ekert. “Event-ready-detectors” Bell experiment via entanglement swapping. *Phys. Rev. Lett.*, 71(26):4287–4290, 1993.

